# Molecular characterization of metastatic ovarian cancer by MALDI imaging mass spectrometry

A thesis submitted for the degree of

**Doctor of Philosophy**

as a combination of research narrative and portfolio of scientific publications by

**Ove Johan Ragnar Gustafsson**




**Adelaide Proteomics Centre**

**Discipline of Biochemistry**

**School of Molecular and Biomedical Science**

**The University of Adelaide, Australia**

**December 2011**

# Table of contents

**Abstract** ..... **9**

**Declaration** ..... **10**

**Acknowledgement of help** ..... **11**

**Acknowledgements** ..... **12**

**Publications** ..... **13**

**Abbreviations** ..... **14**

**Chapter 1 Manuscript Context** ..... **17**

**Chapter 1 MALDI Imaging Mass Spectrometry (MALDI-IMS) — Application of Spatial Proteomics for Ovarian Cancer Classification and Diagnosis** ..... **19**

**Statement of authorship for chapter 1** ..... **20**

*1.1 Abstract* ..... *23*

*1.2 Epidemiology of Ovarian Cancer* ..... *23*

*1.3 Early Detection of Ovarian Cancer* ..... *24*

*1.4 Molecular Classification of Ovarian Carcinomas* ..... *25*

*1.5 Application of Proteomics to Ovarian Cancer* ..... *27*

*1.6 Tissue Analysis by Mass Spectrometry* ..... *28*

*1.7 Methods for in Situ MALDI-TOF Analysis of Ovarian Cancer Tissue* ..... *32*

*1.8 Profiling Cancer Tissues Using MALDI-TOF MS* ..... *35*

*1.9 Profiling vs. Imaging* ..... *37*

*1.10 Software for Data Analysis* ..... *38*

*1.11 Automated Sample Preparation for Imaging Cancer Tissues ..... 39*

*1.12 Peptide Imaging Provides Data Complementary to Protein Imaging ..... 41*

*1.13 Using Histology to Guide Imaging Mass Spectrometry ..... 42*

*1.14 Ovarian Cancer Biomarker Discovery Using Imaging Mass Spectrometry ..... 43*

*1.15 Application of Tryptic Digestion to Formalin-Fixed Paraffin Embedded Ovarian Tissues ..... 44*

*1.16 Conclusions and Future Prospects ..... 46*

**Thesis synopsis and aims ..... 47**

**Chapter 2 Materials and methods ..... 49**

*2.1 Materials and reagents ..... 49*

2.1.1 Solvents, chemicals and matrixes ..... 49

2.1.2 Consumable materials ..... 50

2.1.3 Proteolytic enzymes ..... 51

2.1.4 Peptide and protein calibration standards ..... 51

2.1.5 Instruments/equipment ..... 52

*2.2 Chapter three methods ..... 53*

2.2.1 Experimental Tissues and Solvents ..... 53

2.2.2 Sample Preparation ..... 53

2.2.3 Matrix Deposition – CHIP-1000 ..... 54

2.2.4 Matrix Deposition – ImagePrep ..... 54

2.2.5 MALDI-TOF Imaging Mass Spectrometry and Data Analysis ..... 55

*2.3 Chapter four methods ..... 55*

2.3.1 Chemicals and reagents ..... 55

2.3.2 Tissue source ..... 55

2.3.3 Tissue sectioning ..... 56

2.3.4 Tryptic digestion and matrix deposition methods (ChIP-1000) ..... 56

2.3.5 Matrix deposition (ImagePrep station) ..... 56

2.3.6 MALDI-TOF/TOF imaging mass spectrometry ..... 57
2.3.7 Data processing ..... 58
2.3.8 Haematoxylin and eosin staining ..... 58
*2.4 Chapter five methods* ..... *58*
2.4.1 Chemicals and Reagents. ..... 58
2.4.2 Tissue Samples. ..... 59
2.4.3 Section Preparation and Antigen Retrieval. ..... 59
2.4.4 Trypsin Deposition. ChIP-1000 Printing. ..... 60
2.4.5 ImagePrep Nebulization. ..... 60
2.4.6 Matrix Deposition. ChIP-1000 Printing. ..... 61
2.4.7 ImagePrep Nebulization. ..... 61
2.4.8 Imaging MALDI-TOF-MS. ..... 61
2.4.9 Imaging Tandem MS. ..... 62
2.4.10 Haematoxylin and Eosin Staining. ..... 62
2.4.11 *In Situ* Digestion and Extraction. ..... 63
2.4.12 HPLC and Fraction Collection. ..... 63
2.4.13 MALDI-TOF MS for HPLC Fractions. ..... 64
2.4.14 Tandem MS for HPLC Fractions. ..... 64
2.4.15 HPLC-ESI-Ion Trap MS and MS/MS. ..... 65
2.4.16 HPLC-ESI-LTQ-Orbitrap MS and MS/MS. ..... 65
*2.5 Chapter six and seven methods* ..... *66*
2.5.1 Chemicals and Reagents ..... 66
2.5.2 Tissue Samples ..... 66
2.5.3 Section Preparation and Antigen Retrieval ..... 67
2.5.4 *In situ* tryptic digestion ..... 67
2.5.5 Deposition of internal calibrants ..... 67
2.5.6 Matrix deposition ..... 68
2.5.7 MALDI-TOF-MS imaging data acquisition ..... 68

2.5.8 Data processing ..... 69
2.5.9 Haematoxylin and Eosin Staining ..... 70
2.5.10 K-means clustering ..... 70
2.5.11 Laser micro-dissection, tryptic digestion and purification ..... 70
2.5.12 HPLC-ESI-LTQ-Orbitrap MS and MS/MS ..... 71

**Chapter 3 Manuscript Context ..... 73**

**Chapter 3 Imaging Mass Spectrometry and Its Methodological Application to Murine Tissue ..... 75**

**Statement of authorship for chapter 3 ..... 76**

*3.1 Abstract ..... 78*

*3.2 Introduction ..... 78*

*3.3 Results and Discussion ..... 80*

3.3.1 Stability of Imaging Mass Spectrometry Instrumentation Allows Successful Application ..... 80
3.3.2 Imaging the Murine Brain at the Protein Level with the CHIP-1000 and ImagePrep ..... 84

*3.4 Conclusion ..... 85*

*3.5 Acknowledgement ..... 85*

**The need for further MALDI-IMS method development ..... 87**

**Chapter 4 Development and application of automated sample preparation methods for MALDI imaging mass spectrometry ..... 89**

*4.1 Introduction ..... 89*

*4.2 Optimizing tissue section fixation to minimize analyte delocalization ..... 90*

*4.3 Application of automated sample preparation instruments ..... 96*

4.3.1 Application of the ImagePrep station for protein MALDI-IMS ..... 96

*4.4 Application of the Chemical Inkjet Printer for protein imaging ..... 102*

*4.5 Application of methods for tryptic peptide imaging ..... 106*

4.5.1 Tryptic peptide MALDI imaging mass spectrometry ..... 106

4.5.2 Tryptic peptide imaging using a ChIP-1000 ..... 107

*4.6 Application of tryptic peptide imaging to ovarian cancer ..... 113*

*4.7 Concluding remarks ..... 115*

*4.8 Supplementary Information ..... 116*

**Chapter 5 Manuscript Context ..... 119**

**Chapter 5 Citric Acid Antigen Retrieval (CAAR) for Tryptic Peptide Imaging Directly on Archived Formalin-Fixed Paraffin-Embedded Tissue ..... 121**

**Statement of authorship for chapter 5 ..... 122**

*5.1 Abstract ..... 124*

*5.2 Introduction ..... 124*

*5.3 Results and Discussion ..... 127*

5.3.1 Citric Acid Antigen Retrieval Provides Access to the Proteome of Archived Ovarian Cancer Tissue. 127

5.3.2 Citric Acid Antigen Retrieval Utilized Alongside Automated Nebulization. ..... 132

5.3.3 Tandem MS Acquisition from Tissue Sections. ..... 141

*5.4 Concluding Remarks. ..... 147*

*5.5 Acknowledgment. ..... 147*

*5.6 Supporting information ..... 148*

**Chapter 6 Evaluation of internal calibrants for tryptic peptide MALDI-IMS data ..... 173**

*6.1 Introduction ..... 173*

*6.2 Results and Discussion ..... 175*

6.2.1 Automated generation of ion maps, mass deviation maps and grouped peak lists ..... 175

6.2.2 Characterizing m/z deviation across tissues using internal calibrants ..... 180

6.2.3 Use of internal calibrant peptides to reduce *in situ m/z* deviation ..... 182

*6.3 Concluding Remarks* ..... *192*

**Chapter 7 – Characterization of peritoneal ovarian cancer metastasis by tryptic peptide imaging mass spectrometry** ..... **193**

*7.1 Introduction* ..... *193*

*7.2 Results and Discussion* ..... *196*

7.2.1 Tryptic peptide MALDI-IMS analysis of ovarian cancer metastases ..... 196

7.2.2 Linking peak groups to histology using IonMapper generated signal to noise maps ..... 197

7.2.3 Peptide identification and the application of k-means clustering ..... 203

7.2.4 Matching clusters to histology ..... 208

7.2.5 Cancer peak group features – Hamming distance clustering results ..... 210

7.2.6 Cancer features – Cosine distance clustering results ..... 216

7.2.7 Combining peak group features from Hamming and cosine distance results ..... 220

7.2.8 Peritoneal stroma features ..... 220

7.2.9 Identification of cancer and stroma-specific peak groups by LC-coupled mass spectrometry ..... 226

7.2.10 Improved workflow for characterization of ovarian peritoneal metastasis ..... 235

*7.3 Supplementary Information* ..... *241*

**Chapter 8 Concluding remarks** ..... **253**

*8.1 Aim one* ..... *253*

*8.2 Aim two* ..... *254*

*8.3 Aim three* ..... *254*

*8.4 Aim four* ..... *255*

*8.5 Conclusions and future prospects* ..... *256*

**References** ..... **259**

# Abstract

Imaging mass spectrometry (IMS) is a novel technology which measures the spatial distribution of drugs, lipids, peptides and proteins across tissue sections by application of mass spectrometry (MS) directly to the section surface. Several hundred analytes can be measured across a tissue in a single IMS experiment, without the need for antibodies and without prior knowledge of tissue composition or structure. In the context of human cancers, the molecular information collected by IMS approaches has been used to grade cancers and predict patient survival. IMS is thus a potentially technology capable of providing valuable complementary information to classical histology and immuno-histochemistry.

Ovarian cancers have the highest mortality of any gynaecological cancer. The high mortality results from late diagnosis due to the asymptomatic nature of ovarian malignancies. Advanced stage ovarian tumours will shed cancer cells into the abdominal cavity, where they subsequently implant into the peritoneum and form metastatic tumour nodules. Despite invasive surgery and adjuvant chemotherapy, there is a large increase in patient morbidity following peritoneal metastasis. Compounding this issue further is the absence of reliable grading systems for ovarian cancer and a subsequent lack of individualized treatments for specific cancer sub-types. As a result of the potential ability to grade tumours and provide patient prognoses based on IMS data, the molecular composition of ovarian metastatic tumours was investigated by IMS.

The novelty of IMS required set up of a robust and reproducible workflow. Methods were thus optimized for IMS analysis of both frozen and formalin-fixed paraffin-embedded (FFPE) ovarian tumour tissue. Subsequently it was shown that optimization of available antigen retrieval and tryptic digest methods for accessing FFPE tissues could achieve higher tryptic peptide signal to noise at a better spatial resolution than methods available in the literature. As such, a complete tryptic peptide IMS workflow was developed alongside liquid chromatography (LC) and MS/MS based peptide identification. In conjunction with this workflow, methods for improving the matching of IMS peptides to LC-MS/MS identified peptides using internal calibrants and development of an in-house software tool were described.

As a result of the work presented in this thesis, a complete tryptic peptide IMS workflow which could be applied to virtually any cancer tissue was developed. The application of this workflow, and exploratory k-means clustering, to ovarian peritoneal metastases showed that key tryptic peptides could be found which distinguish cancer tissue from the surrounding peritoneal stroma. This represented the first step in characterizing these metastatic tumours at the molecular level. The results in this thesis are a precursor to future work which will validate these peptide markers and develop a classification system for metastatic ovarian cancers based on patient survival and response to chemotherapy.

# Declaration

This work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution to Ove Johan Ragnar Gustafsson and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. I give consent to this copy of my thesis when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

The author acknowledges that copyright of published works contained within this thesis (as listed in the publications list below) resides with the copyright holder(s) of those works. I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library catalogue, the Australasian Digital Theses Program (ADTP) and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

Ove Johan Ragnar Gustafsson

.....

# Acknowledgement of help

I would like to acknowledge the following people for their help during my thesis research:

All co-authors for the scientific research articles included in this thesis, in particular for their roles in evaluating manuscripts, which can be an arduous task. I would like to extend a special thanks to my supervisors Dr. Peter Hoffmann and Professor Shaun McColl, who were co-authors on and evaluated all manuscripts and thesis chapters.

A/Prof. Martin Oehler for his constant support of MALDI-IMS as a collaborator, contributor to project design and implementation as well as providing human ovarian tissue samples for the research presented in this thesis.

A/Prof. Inge Koch and Prof. Steve Marron for their optimization and implementation of k-means clustering, the results of which are presented in chapter seven.

James Eddes for his assistance in writing the software required to process and present the data in chapters six and seven.

Carmela Ricciardelli, Miranda Ween and Noor Lokman for their assistance with technical questions and tissue processing after surgeries.

Dr. Fergus Whitehead for his assistance annotating the metastatic tumour sections in chapter seven.

Prof. Mark Baker for arranging the loan of the ChIP-1000 instrument to the Adelaide Proteomics Centre.

# Acknowledgements

Most acknowledgements finish with a profound statement, something along the lines of…..wise men can't jump…….ok so that's not profound at all but you know what I mean. I've never followed trends so I'm going to start mine with a joke. I don't know the reference but obviously the joke is not mine.

A cation runs into a bar screaming "I've lost my electron, I've lost my electron!". The bartender calms the ion down and asks "Are you sure?" The ion replies, "I'm positive."

Now that the giggles have subsided……When I started my Honors year in 2007, one of my supervisors asked me why I chose science. The answer now, as I write the final parts of my doctoral thesis, is thankfully the same as it was then, which I take to mean that I probably made the right choice. The answer I gave was that I will never be happy as part of the status quo. I want to be a force for change in the world, no matter how small my contribution. I have many people to thank for their support, friendship and help over the past three years. It would take many pages to properly thank everyone who has contributed to my life so I will have to settle for those which have had the greatest impact.

The Adelaide Proteomics Centre has been my home away from home for four years. Most important to my PhD experience at the centre have been my supervisors, Peter Hoffmann and Shaun McColl, who have always been supportive, patient and willing to discuss new methods and experiments.

I am forever grateful to the crew of the proteomics centre, including Mark Condina, Megan "Retallicka" Penno, James Eddes, Sandra Hack, Florian "Florider" Weiland, Karina "Kaz" Martin, Tomas "Charlie Brown" Koudelka, Yin Ying Ho and Chris "Vinnie" Cursaro. Without you all this thesis would not be possible and the past three years would have been a lot less fun. I also want to thank my close circle of friends. You know who you are (at least you should) and you also know how much friends mean in my life.

The support of parents is often unseen and even more often taken for granted. Thank you to my parents. I love you both and am forever in your debt for providing everything I ever needed to reach my goals. Finally I want to thank my beautiful fiancé Tanja. She has been the most important person in my life over the past seven years and her unwavering support during my PhD has been a blessing. Convincing such an amazing woman that I'm worth her time remains my greatest achievement.

Ove Johan Ragnar Gustafsson

.....

# Publications

Directly related to thesis:

Gustafsson, J. O. R., M. K. Oehler *et al.* (2011). MALDI Imaging Mass Spectrometry (MALDI IMS) - Application of Spatial Proteomics for Ovarian Cancer Classification and Diagnosis. International Journal of Molecular Sciences *12*(1), 773-794 published online 21 January 2011.

Gustafsson, J. O. R., S. R. McColl, *et al.* (2008). "Imaging mass spectrometry and its methodological application to Murine tissue." Journal of Proteomics and Bioinformatics 1(9): 458-463.

Gustafsson, J. O. R., M. K. Oehler, et al. (2010). "Citric Acid Antigen Retrieval (CAAR) for Tryptic Peptide Imaging Directly on Archived Formalin-Fixed Paraffin-Embedded Tissue." J Proteome Res 9(9): 4315-4328.

Arising from thesis:

Condina, M. R., Gustafsson, J. O. R., *et al.* (2010). EZYprep LC-coupled MALDI-TOF/TOF MS: An improved matrix spray application for phosphopeptide characterisation. Proteomics, 10, 2516-2530.

# Abbreviations

| | |
|---|---|
| ACN | Acetonitrile |
| ANI | Aniline |
| 3-AP | 3-acetyl-pyridine |
| AR | Antigen retrieval |
| AWM | Abundance weighted mean |
| CAAR | Citric acid antigen retrieval |
| CCTV | Close circuit television |
| CHCA | α-cyano-4-hydroxycinnamic acid |
| ChIP-1000 | Chemical inkjet printer 1000 |
| CID | Collision induced dissociation |
| Da | Dalton |
| kDa | Kilo dalton |
| 2-DE | Two dimensional electrophoresis |
| 2,5-DHB | 2,5-dihydroxybenzoic acid |
| EDTA | Ethylene diamine tetra acetic acid |
| ELISA | Enzyme linked immuno-sorbent assay |
| EOC | Epithelial ovarian carcinoma |
| ESI | Electro-spray ionization |
| EtOH | Ethanol |
| FA | Formic acid |
| FFPE | Formalin-fixed paraffin-embedded |
| FIGO | International federation of gynecology and obstetrics |
| FWHM | Full width half maximum |
| H&E | Haematoxylin and eosin |
| GAPDH | Glyeraldehyde-3-phosphate dehydrogenase |
| HFIP | 1,1,1,3,3,3-Hexafluoro-2-propanol |

| | |
|---|---|
| hnRNP | Heterogeneous nuclear ribonucleoprotein |
| HPLC | High performance liquid chromatography |
| IHC | Immuno-histochemistry |
| IMS | Imaging mass spectrometry |
| IPA | Isopropanol |
| IT | Ion trap |
| ITO | Indium tin oxide |
| LC | Liquid chromatography |
| LCM | Laser capture micro-dissection |
| LC-MS | Liquid chromatography – mass spectrometry |
| LTQ | Linear trap with quadrupole |
| MALDI | Matrix assisted laser desorption/ionization |
| MALDI-IMS | Matrix assisted laser desorption/ionization imaging mass spectrometry |
| MeOH | Methanol |
| mL | Milli litre |
| MS | Mass spectrometry |
| m/z | Mass to charge ratio |
| nL | Nano litre |
| NSCLC | Non small cell lung cancer |
| OCT | Optimal cutting temperature polymer |
| PCA | Principal component analysis |
| PEG | Polyethylene glycol |
| PIC | Percentage intensity contribution |
| pmol | Pico mole |
| png | Portable network graphic |
| POS | Percentage of spectra (within a cluster) |
| ppm | Parts per million |
| PTM | Post translational modification |

| | |
|---|---|
| SA | Sinapinic acid |
| SIMS | Secondary ion mass spectrometry |
| SLSC | Standard light scatter curve |
| SNAP | Sophisticated numerical annotation procedure |
| STS | Soft tissue sarcoma |
| TOF | Time-of-flight |
| TFA | Trifluoroacetic acid |
| TFE | 2,2,2-trifluoroethanol |

# Chapter 1 Manuscript Context

The ultimate goal of the research presented in this thesis was the molecular characterizing of ovarian cancer metastasis by mass spectrometry (MS). The technology selected for this undertaking was matrix assisted laser desorption/ionization imaging mass spectrometry (MALDI-IMS), which allows the mapping of peptide and/or protein distributions across sectioned tissue by MS measurements directly from the section surface. This project was sub-divided into four distinct aims designed to achieve molecular characterization of ovarian cancer metastases. These were:

1. Develop MALDI-IMS into a platform for reproducibly tracking protein ion changes across sections derived from frozen tissue.
2. Apply developed MALDI-IMS methods to formalin-fixed paraffin-embedded (FFPE) tissue and optimize to achieve a reproducible platform for imaging FFPE tissues.
3. Apply LC-MS/MS for the identification of protein/peptide ions of interest.
4. Application of developed MALDI-IMS methods to the characterization of peritoneal metastases of serous epithelial ovarian cancer.

The structure of the presented thesis, which details the experiments performed to achieve these aims, is as follows:

*Chapter 1* – Manuscript chapter (Thesis introduction)

*Chapter 2* – Materials and Methods

*Chapter 3* – Manuscript chapter (proof-of-principle IMS application)

*Chapter 4* – Research chapter (completion of thesis aim one)

*Chapter 5* – Manuscript chapter (completion of thesis aim two)

*Chapter 6* – Research chapter (completion of thesis aim three)

*Chapter 7* – Research chapter (completion of thesis aim four)

*Chapter 8* – Concluding remarks

In addition to the presented research chapters (**chapter 4, 6 and 7**), three manuscripts directly related to this thesis work were published.

The first was a proof-of-principle demonstration that the distribution of proteins across tissue sections could be achieved using available MS technologies (see **chapter 3**). The second detailed the

development and application of an antigen retrieval (AR) method to allow analysis of formalin-fixed paraffin-embedded (FFPE) tissue. This method was subsequently used to map the distribution of tryptic peptides by IMS (see **chapter 5**). The third and final was a review manuscript which tied together the state of the art methods, technologies and applications within the IMS field as of early 2011. The review manuscript outlined the classification and diagnostic criteria currently used for ovarian cancer, the IMS technology (along with typical workflows), key applications of IMS as well as the IMS projects which have studied ovarian cancer. This review is presented as the introductory chapter of this thesis. As the review was published last, it contains references to the manuscripts presented in **chapters 3** and **5**.

# Chapter 1 MALDI Imaging Mass Spectrometry (MALDI-IMS) – Application of Spatial Proteomics for Ovarian Cancer Classification and Diagnosis

*Int. J. Mol. Sci.* **2011**, *12*, 773-794; doi:10.3390/ijms12010773

International Journal of Molecular Sciences ISSN 1422-0067 www.mdpi.com/journal/ijms

Review

Johan O. R. Gustafsson[1], Martin K. Oehler[2], Andrew Ruszkiewicz[3], Shaun R. McColl[1] and Peter Hoffmann [1*]

1 Adelaide Proteomics Centre, School of Molecular and Biomedical Science, The University of Adelaide, SA 5005, Adelaide, Australia; E-Mails: ove.gustafsson@adelaide.edu.au (J.O.R.G.); shaun.mccoll@adelaide.edu.au (S.R.M.)

2 Robinson Institute, Research Centre for Reproductive Health, School of Paediatrics and Reproductive Health, The University of Adelaide, SA 5005, Adelaide, Australia; E-Mail: martin.oehler@adelaide.edu.au

3 Surgical Pathology, SA Pathology, Adelaide SA 5000, Australia; E-Mail: andrew.ruszkiewicz@health.sa.gov.au

* Author to whom correspondence should be addressed; E-Mail: peter.hoffmann@adelaide.edu.au; Tel.: +61-(08)-8303-5507; Fax: +61-(08)-0-8303-4362.

Received: 1 December 2010; in revised form: 10 January 2011 / Accepted: 17 January 2011 / Published: 21 January 2011

Keywords: MALDI; imaging; mass spectrometry; ovarian cancer; grading; biomarker

# Statement of authorship for chapter 1

**Ove Johan Ragnar Gustafsson (Candidate)**

Statement of contribution (in terms of the conceptualization of the work, its realization and its documentation)

Performed analysis on all samples, interpreted data and wrote the manuscript.

Certification that the statement of contribution is accurate

Signed.....Date.....

**Martin K. Oehler (co-author)**

Statement of contribution (in terms of the conceptualization of the work, its realization and its documentation)

Manuscript evaluation, provided tissue samples for ovarian cancer research.

Certification that the statement of contribution is accurate and permission is given for the inclusion of the paper in the thesis

Signed.....Date.....

**Andrew Ruszkiewicz (co-author)**

Statement of contribution (in terms of the conceptualization of the work, its realization and its documentation)

Manuscript evaluation.

Certification that the statement of contribution is accurate and permission is given for the inclusion of the paper in the thesis

Signed.....Date.....

**Shaun R. McColl (co-author)**

Statement of contribution (in terms of the conceptualization of the work, its realization and its documentation)

Supervised development of work, manuscript evaluation and provided tissue for murine brain imaging research.

Certification that the statement of contribution is accurate and permission is given for the inclusion of the paper in the thesis

Signed.....Date.....

**Peter Hoffmann (co-author)**

Statement of contribution (in terms of the conceptualization of the work, its realization and its documentation)

Supervised development of work, helped in data interpretation, manuscript evaluation and acted as corresponding author

Certification that the statement of contribution is accurate and permission is given for the inclusion of the paper in the thesis

Signed.....Date.....

## 1.1 Abstract

MALDI imaging mass spectrometry (MALDI-IMS) allows acquisition of mass data for metabolites, lipids, peptides and proteins directly from tissue sections. IMS is typically performed either as a multiple spot profiling experiment to generate tissue specific mass profiles, or a high resolution imaging experiment where relative spatial abundance for potentially hundreds of analytes across virtually any tissue section can be measured. Crucially, imaging can be achieved without prior knowledge of tissue composition and without the use of antibodies. In effect MALDI-IMS allows generation of molecular data which complement and expand upon the information provided by histology including immuno-histochemistry, making its application valuable to both cancer biomarker research and diagnostics. The current state of MALDI-IMS, key biological applications to ovarian cancer research and practical considerations for analysis of peptides and proteins on ovarian tissue are presented in this review.

## 1.2 Epidemiology of Ovarian Cancer

In 2010, an estimated 21 880 new cases of ovarian cancer will be diagnosed in the USA **[1]**. http://www.cancer.org/research/cancerfactsfigures/cancerfactsfigures/cancer-facts-and-figures-2010.

With a projected 13 850 deaths from this disease in 2010, ovarian cancer has the highest mortality rate of all gynaecological malignancies. The high mortality from ovarian cancer is due to the majority of patients (64%, see **Table 1.1**) being diagnosed with advanced (International Federation of Gynecology and Obstetrics (FIGO) stage III + IV) disease, which has a maximum 5-year survival of only 30% **[2]**. In contrast, the 5-year survival for patients with organ-confined FIGO stage I ovarian cancer exceeds 90% and a large number of these patients are cured. Thus, early detection is the key to increased survival in ovarian cancer.

**Table 1.1.** FIGO ovarian cancer stages, prevalences and anatomical features.

| FIGO Stage | Prevalence (%) | Anatomical features |
|---|---|---|
| I | 25 | Limited to ovaries |
| II | 11 | Pelvic extension |
| III | 47 | Abdominal extension and/or positive lymph nodes |
| IV | 17 | Distant metastases |

## 1.3 Early Detection of Ovarian Cancer

While other gynaecological cancers can be diagnosed at an early stage due to effective screening (e.g., PAP smear in the case of cervical cancer) or symptoms (e.g., bleeding in the case of endometrial cancer), neither specific early disease symptoms or an early detection test exist for ovarian cancer. Presently, diagnosis involves a combination of physical examination, followed by trans-vaginal ultrasound, measurement of serum levels of the glycoprotein CA-125 and exploratory surgery if a suspicious ovarian lesion has been identified **[1]**. CA-125, when combined with ultrasound, has a positive predictive value (PPV) of only 35.1% for primary EOCs **[2]**. This low PPV indicates that two out of every three patients will be wrongly diagnosed and undergo unnecessary and potentially harmful invasive procedures. Novel biomarkers are therefore required to improve ovarian cancer detection.

Single markers, such as CA125, are unlikely to provide the sensitivity and specificity required for ovarian cancer screening **[3]**. The focus has thus shifted to panels of biomarkers, which for the moment are additional diagnostic tools, not screening options **[4, 5]**. Further improvements to these panels require not only discovery of new biomarkers, but also validation of existing biomarker candidates. Moreover, the large numbers of newly identified potential biomarkers have to be validated individually in a large cohort of patients, which is currently impractical. The use of serum or plasma for many biomarker discovery projects also complicates the process of biomarker discovery, as serum has a high protein complexity, large dynamic range of protein concentration ($10^{12}$) **[6]** and contains non-specific acute phase proteins. A more promising approach is therefore the direct analysis of the cancer tissue, as it

should have the highest concentration of disease specific markers **[7]** and a smaller dynamic range of protein concentration ($10^6$) **[8]**. Thus, by focusing on tissue identified candidates it should be possible to compile a smaller subset of biomarkers with a higher specificity which can be validated *in situ* by immuno-histochemistry (IHC) and subsequently in large patient cohorts by established methods like enzyme linked immuno-sorbent assays (ELISA). These biomarkers could then be used in novel high specificity panels for early diagnosis of ovarian cancer.

## 1.4 Molecular Classification of Ovarian Carcinomas

The absence of reliable biomarkers is not the only issue with respect to ovarian cancer diagnosis. Following histologic confirmation of ovarian disease, treatment is assigned based upon stage **[1]**. Ovarian cancer staging is currently defined by the FIGO staging system for tumour dissemination into extra-ovarian sites (see **Table 1.1**), which correlates well with patient five year survival (see SEER http://seer.cancer.gov/statfacts/html/ovary.html)**[9, 10]**. However, grade is an additional important prognostic parameter **[11]**. Grade, as determined by light microscopy describes morphological characteristics of tumour tissue including percentage of solid growth, architecture, nuclear features and mitotic activity (see **Table 1.2**) **[12]**. These characteristics are subjective and their reproducibility may be suboptimal **[12]**. Moreover, contention exists as to which grading systems most accurately reflect ovarian tumour differentiation status and therefore optimal treatment **[12, 13]**.

**Table 1.2.** Grading systems for epithelial ovarian carcinoma: FIGO, universal three tier grading and two tier grading.

| Grading system | | Grade | Key features | | | Ref. |
|---|---|---|---|---|---|---|
| FIGO | | 1 | Well differentiated | Grade based on % solid non-squamous growth, grade + 1 if nuclear atypia apparent | <5% solid growth | [10] |
| | | 2 | Moderately differentiated | | 6–50% solid growth | |
| | | 3 | Poorly differentiated | | >50% solid growth | |
| 3-tier universal grading | | 1 | Grade based on sum of individual feature scores (see right)<br>1 = 3–5 points<br>2 = 6–7 points<br>3 = 8–9 points | Architecture based score | Glandular = 1 point<br>Papillary = 2 points<br>Solid = 3 points | [13] |
| | | 2 | | Nuclear pleomorphism score | Slight = 1 point<br>Moderate = 2 points<br>Marked = 3 points | [12] |
| | | 3 | | Mitotic activity score | 0–9 = 1 point<br>10–24 = 2 points<br>≥25 = 3 points | |
| 2-tier grading | Serous tumour | Low grade (type I) | Slow development | Low chromosomal instability | Gene mutation–KRAS, BRAF, ERBB2 | [14] |
| | | High grade (type II) | Rapid development | High chromosomal instability | Gene mutation–P53 | |
| | Endometroid tumour | Low grade | Well differentiated, no necrosis | Solid glandular architecture | Gene mutation–Wnt, PI3K/Akt | [13] |
| | | High grade | Solid growth >50%, necrosis | Diffusely infiltrative or expansive growth, no glandular architecture | Gene mutation–TP53 | [16] |

Based on recent advances in the understanding of the molecular biology of ovarian cancer it is now believed that the major ovarian cancer subtypes can be separated (see **Table 1.2**) into type I (low grade) or type II (high grade) based upon differential gene and/or protein expression **[13-16]**. These two-tiered molecular systems of ovarian cancer grading provide an avenue for defining cancer differentiation state in absolute terms. As such, molecular grading systems need to be developed to a point where they can complement routine histo-pathological examination of ovarian cancer tissue. Importantly, this also needs to be achieved on a similar time scale to histology, in this case one to two hours. Thus, to improve EOC management and outcome for patients, both discovery of novel, effective biomarkers and development of a new molecular grading/classification system are required.

## 1.5 Application of Proteomics to Ovarian Cancer

Although gene expression is useful for distinguishing ovarian tumour subtypes **[14]**, it does not always correlate with protein translation **[17, 18]**, nor can levels of post translational modification (PTM) be directly inferred from genetic analyses **[19]**. However, both protein expression level and PTM state have drastic effects on cellular function/dysfunction and as a result it is more meaningful to analyze the disease-related proteins and peptides. Generating protein profiles with sufficient molecular features is impossible with IHC, as it is limited to a maximum of three to four antibodies at a time and, crucially, depends on antibody quality. Proteomics, however, allows analysis of hundreds to thousands of peptide and protein features in biological samples **[20]**, in many cases without the need for antibodies.

The term "proteomics" was coined to describe the quantitative analysis of the proteome, which represents all proteins expressed in a given cell, tissue (e.g., cancer) or biological fluid (e.g., serum) at a given point in time or under the effects of a defined biological stimulus **[21]**. High analytical sensitivity is achieved in proteomics because complex protein mixtures are fractionated following tissue or cell lysis (disruption), followed by further purification or direct analysis by mass spectrometry (MS) **[22, 23]**. These methods allow for identification of thousands of proteins from a single cell lysate. For example, two separate studies from 2006 **[24]** and 2008 **[19]** demonstrated profiling of ovarian cancer subtypes using liquid chromatography (LC) separation followed by MS (LC-MS). The 2008 study showed that early and late stage endometroid ovarian carcinoma MS profiles can be distinguished using a clustering analysis, which separates profiles based on feature similarity; in this case similar protein masses **[19]**. Importantly, the 2008 publication also combined profiling MS data for serous and clear cell tumours from the 2006 study **[24]** to show that the three subtypes grouped separately in a principal component analysis (PCA). These studies are significant as they indicate that "classical proteomics" can generate molecular fingerprints of disease. However, there are two issues for implementing proteomics in this manner. Firstly, tissue disruption for analysis removes spatial proteome information, which is critical for

clinical application, especially in heterogeneous carcinomas where different structural elements will express a unique proteome with subsequent unique cellular function. A common method for addressing this problem is laser capture micro-dissection (LCM) **[7]**, which can isolate specific cell populations for analysis. However, similar to many proteomics methods, including liquid phase separation, LCM is time consuming. The second issue is thus that a proteomic method is required that can be implemented in the same time frame as classical histology (*i.e.*, one to two hours).

## 1.6 Tissue Analysis by Mass Spectrometry

Direct tissue section analysis utilizing an MS instrument removes the need for disruption and the subsequent loss of spatial proteome information. This approach also provides an avenue for molecular classification/grading because tissue sections can be prepared and analyzed rapidly (1–3 hours) using standardized protocols. Importantly, tissue specific biomarkers can be visualized and subsequently identified using “classical proteomics” methods such as LC-MS. For easy reference, the advantages and disadvantages of methods for analyzing tissues (histology, IHC and proteomics) are summarized in **Table 1.3**.

**Table 1.3.** Comparison of different methods (histology, immuno-histochemistry (IHC) and proteomics (fractionation coupled to mass spectrometry (MS) and direct tissue MS) for peptide/protein analysis in tissue samples.

| | Histology | IHC | Proteomics | |
|---|---|---|---|---|
| | | | Fractionation-MS | Direct tissue MS |
| **Methods** | Cellular staining | Antibody directed staining of specific proteins | Liquid phase separation (*i.e.*, liquid chromatography) | Direct measurement of peptides and proteins from tissue section |
| **Analysis** | Tissue morphology assessment by light microscopy | Protein distribution across tissue sections | MS protein identification | MS profiles of tissue sections |
| | | | Quantitation using protein labelling | Peptide and protein intensity maps showing distribution across tissue sections |
| **Advantages** | Easy staining methods | Highly specific | Highly sensitive | Rapid |
| | Cellular microscopy resolution | Cellular microscopy resolution | Thousands of proteins analysed at a time | Spatial proteome information |
| | Well established | Well established | Heavily automated | Measurement of hundreds of molecular features at a time |
| | Clinical personnel already available | Clinical personnel already available | Highly modular workflows | No antibodies required |
| **Disadvantages** | Reproducibility issues | Time consuming | Time consuming | Expensive equipment |
| | Based on visual assessment of morphology | Labor intensive | Labor intensive | Novel technology |
| | Non-specific | Limited to 3–4 proteins | Removes spatial information | Requires fraction-MS based proteomics to identify peptide and protein features |
| | Analysis is subjective | Dependent on antibody quality | Requires specialist personnel | Analytical resolution limited to a maximum of 20–50 μm |

MS measurement of molecules directly from tissue was first described in 1997 **[25]**. MS instruments measure the mass to charge ratio (*m/z*) of gaseous ions, in this case peptide or protein ions. Mass is of value because it indicates composition, which, for example, can be used to identify proteins of interest by their component peptides. To generate ions directly from tissue, either secondary ion MS (SIMS) or matrix-assisted laser desorption/ionization (MALDI) instruments are utilized.

SIMS utilizes the impact of an ion beam (e.g., Ar+ or Ga+) to induce a localized gain in kinetic energy on the tissue surface. Once a sufficient energy level is reached secondary ions (e.g., peptides) are ejected from the tissue for mass measurement **[26]**. In practice, SIMS causes surface fragmentation and as a result limits measurement to metabolites, lipids and small peptides (<1000 Da) **[26]**.

MALDI represents a more suitable ionization method for direct application to tissue. Preparation for MALDI requires a tissue section to be coated with a low molecular weight organic molecule, called the "matrix". The most common matrix compounds, 2,5-dihydroxybenoic acid (DHB), α-cyano-4-hydroxycinnamic acid (CHCA), and sinapinic acid (SA) are listed in **Table 1.4** along with example modifications made to the matrix composition, their full chemical names and bio-molecule specificity. Several of these matrix combinations have been applied to ovarian tissue including DHB/3-AP **[27]**, CHCA **[28]**, CHCA/ANI **[28]**, SA **[28, 29]**, SA/3-AP **[30]** and SA/HFIP **[30, 31]**. Most matrixes are dissolved in a 50–60% acidified organic solvent solution, which extracts lipids, peptides and proteins from the tissue prior to evaporation, allowing the matrix to crystallize. The end result is a field of sample-matrix co-crystals on the tissue surface.

**Table 1.4.** List of the three most common matrix types—2,5-dihydroxybenzoic acid (DHB), α-cyano-4-hydroxycinnamic acid (CHCA) and 3,5-dimethoxy-4-hydroxycinnamic acid (sinapinic acid, SA) as well as their documented modifications - for MALDI mass spectrometry. Suitability for measurement of bio-molecules is specified **[26]**.

| Matrix | Chemical name | Biomolecule specificity |
|---|---|---|
| DHB | 2,5-dihydroxybenzoic acid | Lipids, peptides, <10 kDa proteins |
| DHB/aniline | DHB + aniline | Lipids, peptides, <10 kDa proteins |
| DHB/3-AP | DHB + 3-acetyl pyridine | Lipids, peptides, <10 kDa proteins |
| CHCA | α-cyano-4-hydroxycinnamic acid | Peptides, small proteins (<10 kDa) |
| CHCA/aniline | CHCA + aniline | Peptides, <10 kDa proteins |
| SA | 3,5-dimethoxy-4-hydroxycinnamic acid | Proteins (>10 kDa) |
| SA/aniline | SA + aniline | Proteins (>10 kDa) |
| SA/3-AP | SA + 3-acetyl pyridine | Proteins (>10 kDa) |
| SA/HFIP | SA + 1,1,1,3,3,3-hexafluoro-2-propanol | Proteins (>30 kDa) |
| SA/TFE | SA + 2,2,2-trifluoroethanol | Proteins (>30 kDa) |

MALDI is achieved by directing a laser beam at the co-crystals. The matrix absorbs the bulk of incident laser energy, causing an explosive transition from solid crystal to a gaseous plume, during which ionization of the sample occurs (see **Figure 1.1**) **[32-35]**. MALDI is suited to bio-molecule analysis because it is a *"soft"* ionization process, in that the matrix is the energy absorber, minimizing protein/peptide fragmentation. MALDI ion sources are typically coupled to time-of-flight (TOF) mass

analyzers. Ions from the MALDI process are accelerated into the TOF tube, which is an electric field free flight region. The kinetic energy gained during acceleration decreases with increasing mass and as such heavier ions will fly slower and therefore have a longer time-of-flight. This is the basis of TOF mass analysis. When an ion hits an attached detector, the time from laser ionization to detection is used to derive *m/z* (see **Figure 1.1**). The end result is a plot of *m/z* against intensity (ion counts); commonly referred to as an MS spectrum. The preference of most groups for MALDI-TOF instruments is due to their sensitivity (femto to attomolar range under ideal conditions), ease of use and achievable mass range of MALDI-TOF, which reaches from small molecules (100 Da) to large proteins (>300 kDa), allowing measurement of metabolites, lipids, peptides and proteins on the same instrument.

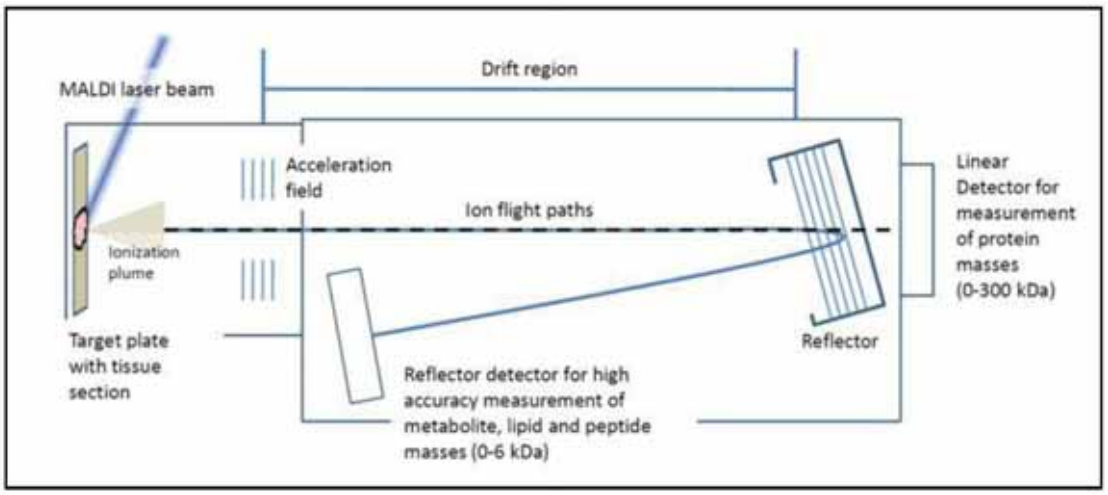



**Figure 1.1.** Basic principles of matrix-assisted laser desorption/ionization (MALDI) time-of-flight (TOF) mass spectrometry. Following ionization, sample ions are accelerated into an electric field free drift region. The larger the ion the less energy it will gain during acceleration and as a result it will travel slower than smaller ions. This is the basis of time of flight separation. Time from laser ionization to detection at the opposite end of the drift region is used to determine mass to charge ratio (*m/z*) for masses between 0–300 kDa. High mass accuracy is achieved using a reflector field that focuses ions from 0–6 kDa onto a secondary detector.

The key advantages of MALDI-TOF MS application directly to tissue are thus that:

(i) Several bio-molecule classes from different mass ranges can be measured, including drugs **[36]**, lipids **[27]**, peptides **[37, 38]** and proteins **[31, 39, 40]**.

(ii) Several hundred molecular features can be measured in a single experiment (see **Figure 1.3a-c**)

(iii) No preliminary knowledge about tissue composition is required.

(iv) No antibodies are required.

## 1.7 Methods for *in Situ* MALDI-TOF Analysis of Ovarian Cancer Tissue

An outline of the methodology for *in situ* MS analysis of ovarian tissue is shown in **Figure 1.2**. Sectioned tissue (2–10 μm thick) is mounted directly onto chilled conductive glass slides (indium tin oxide coated) or metallic targets (e.g., gold-coated target). The mounted sections are dried (15–45 min) before fixation with a graded alcohol series (70% and 100% v/v ethanol/isopropanol) or organic solvents such as chloroform or acetone **[41-44]**. This fixation has been shown to increase the quality of MS signals, most likely as a result of physiological salt and lipid removal, both of which interfere with matrix crystallization and subsequent quality of MS data **[42, 43]**. For ovarian tissue, washes with 70% and 100% isopropanol are sufficient to generate rich small protein (**Figure 1.3a**) and endogenous peptide (**Figure 1.3b**) MS signals directly from carcinoma sections. However, chloroform can also be used quite successfully as a stand-alone wash prior to analysis **[30]**. As described previously, to allow MALDI measurement, the tissue is coated with a matrix solution. The method of matrix coating is important as it affects the type (single spectra or multiple) and quality (MS sensitivity) of data obtained. These methods are discussed further as they pertain to the two types of *in situ* MALDI-TOF MS experiment, profiling and imaging MS (IMS).

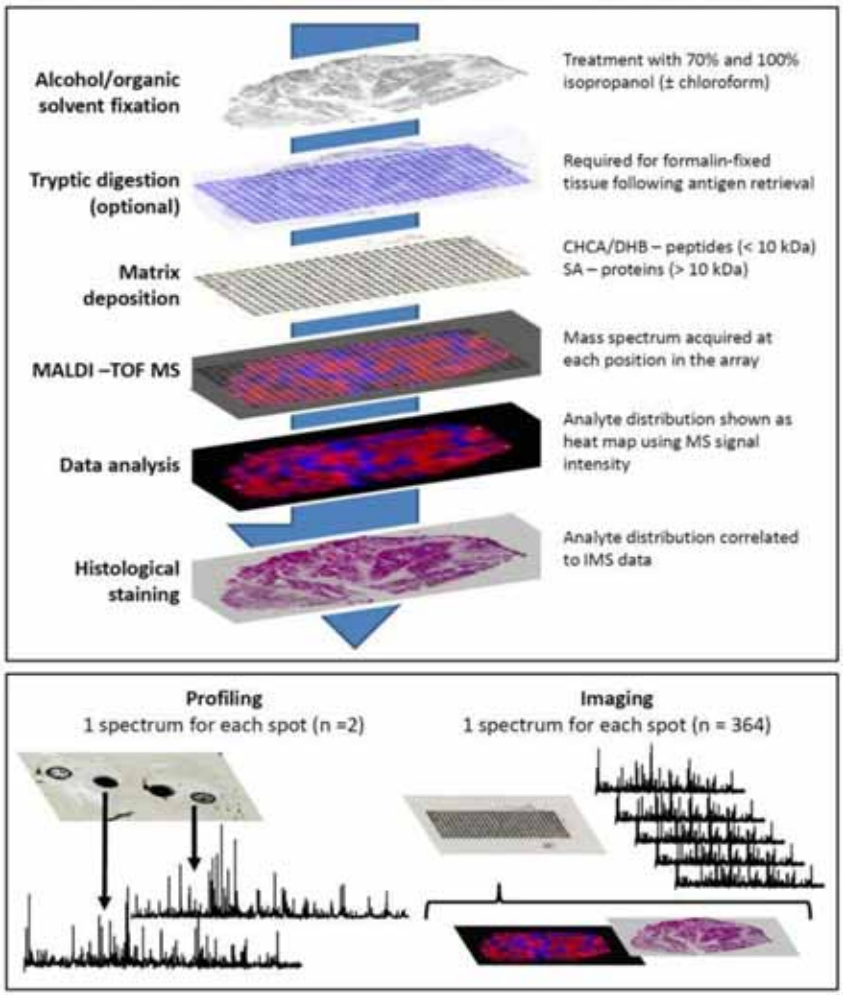



**Figure 1.2.** The top panel shows a typical workflow for IMS on ovarian tissue. Note the optional tryptic digest (absolute requirement for formalin-fixed paraffin-embedded (FFPE) tissue). Antigen retrieval can also be used to partially hydrolyze formalin-induced protein cross-links. The bottom panel shows the two analysis workflows possible for an IMS experiment, profiling and imaging.

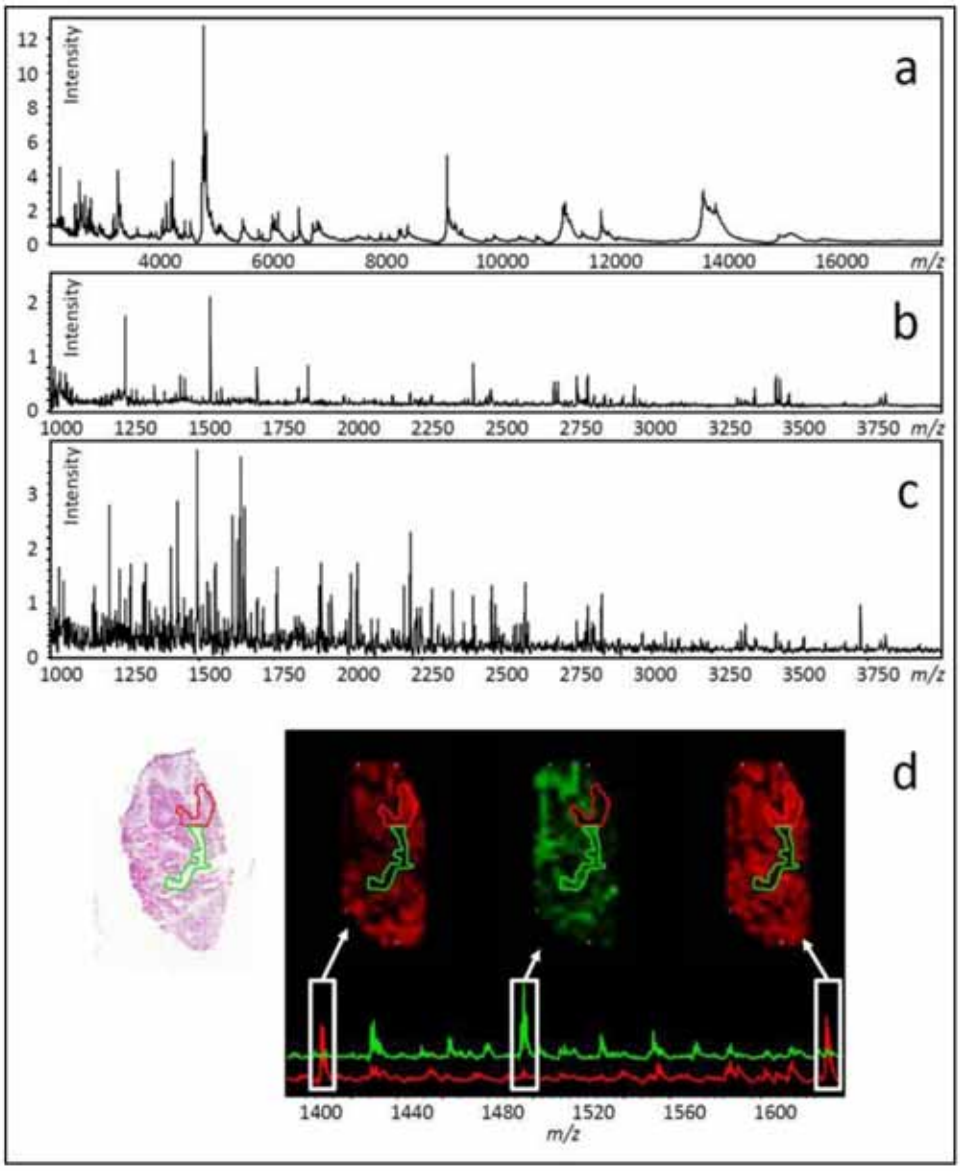



**Figure 1.3.** Data from printed arrays on stage IIIC ovarian epithelial carcinoma. The spectra in a-d represent the sum of all spectra for small protein (**a**–matrix only + peptide/ small protein mass range), endogenous peptide (**b**–matrix only + peptide mass range) and tryptic peptide (**c**–trypsin digestion + matrix with peptide mass range) analysis using a MALDI-TOF/TOF MS instrument. Twenty mg/mL DHB in 50% methanol and 0.2% trifluoroacetic acid was used as a matrix. Trypsin was used at 40 ng/µL in a 5 mM $NH_4HCO_3$ and 12% acetonitrile buffer at pH ~8.5. Panel **d** shows two morphologically different areas on a H&E stained section (green/red outlines), previously analyzed by tryptic peptide IMS. Three ion intensity maps with associated spectra for the green and red areas show differential peptide distribution.

## 1.8 Profiling Cancer Tissues Using MALDI-TOF MS

Typically, *in situ* MS methods are split into two types of workflows, these being profiling MS or IMS (see **Figure 1.2**, bottom panel). The profiling MS approach uses manual or automated deposition of matrix at discrete locations on a tissue section. MS spectra are then acquired from these positions and compared. If tissue MS profiles are known from previous analyses, the same tissue type can be identified in future studies based on this profile, a process similar to the MALDI Biotyper platform used for identification of micro-organisms **[45]**. Because of the novelty of IMS research, there are few publications dealing exclusively with ovarian cancer. However, successful examples of molecular classification/grading have been published for non-small cell lung cancer (NSCLC) **[46]** and soft tissue sarcomas (STS) **[47]**. In the lung cancer study, 100 nL of sinapinic acid (SA, see **Table 1.4**) matrix was manually spotted onto 42 NSCLC tumours (>70% tumour cellularity) and 8 normal lung sections. MS data was acquired in a *m/z* range of 2–25 kDa on a MALDI-TOF MS system and 82 peaks specific to the cancerous tissue were selected for development of a class prediction model. The training model generated was applied to a validation set of 32 tumour and 5 normal lung sections. Based on the MS data, all 32 tumours of the validation set were classified correctly as tumour or normal. Moreover, segregation of these mass signals based on tumour subtype, in this case 14 adenocarcinoma, 15 squamous cell carcinoma and 5 large cell carcinoma, allowed for 100% separation of adenocarcinomas and squamous cell as well as squamous and large cell tumours. Only one large cell tumour was mis-classified as adenocarcinoma in the study **[46]**. The STS study was able to distinguish low and high grade STS using MALDI IMS profiling. Drops of SA matrix (200 nL) were applied directly to areas showing cellular proliferation following MALDI-compatible cresyl violet staining **[47, 48]**. It was determined that calgizzarin (S100 A11), calcyclin, macrophage inhibitory factor and calgranulin were potentially diagnostic for high grade STS, with key extracellular proteins such as myosin being down-regulated in both low and high grade tumours as compared to control muscle tissue. These findings were confirmed using IHC against sections of tumour and control tissue **[47]**.

In addition, pilot experiments have been published for grading follicular lymphoma **[49]**, detection of pre-invasive bronchial lesions **[50]** and prostate cancer **[51, 52]**, classification of meningiomas **[53]**, and generation of prognostic information for gliomas **[54]**. These pilot studies have shown the utility of characterizing disease via direct MS tissue section analysis to gain diagnostic **[47]** and prognostic data **[54]**.

The most comprehensive study to date analyzing EOC **[30]** profiled tissue from 19 ovarian tumours (10 benign, 6 carcinoma and 3 borderline). Three mass ranges were examined, combining data from typical IMS peptide (CHCA matrix, see **Table 1.4**) and small protein (SA matrix) analysis with a novel method for extracting high molecular weight proteins using SA dissolved in 1,1,1,3,3,3-Hexafluoro-2-propanol (HFIP). Using stage III and IV tumours, as compared to benign tissue, it was possible to profile masses matching those of cancerous marker proteins previously identified in EOC, including tetranectin (17.7 kDa) and urokinase plasminogen activator (36.9 kDa) **[30]**. **Figure 1.4** highlights the differences between tissues as a plot of *m/z* against sample number as well as a PCA loadings plot showing separation of the benign, carcinoma and borderline tissues.

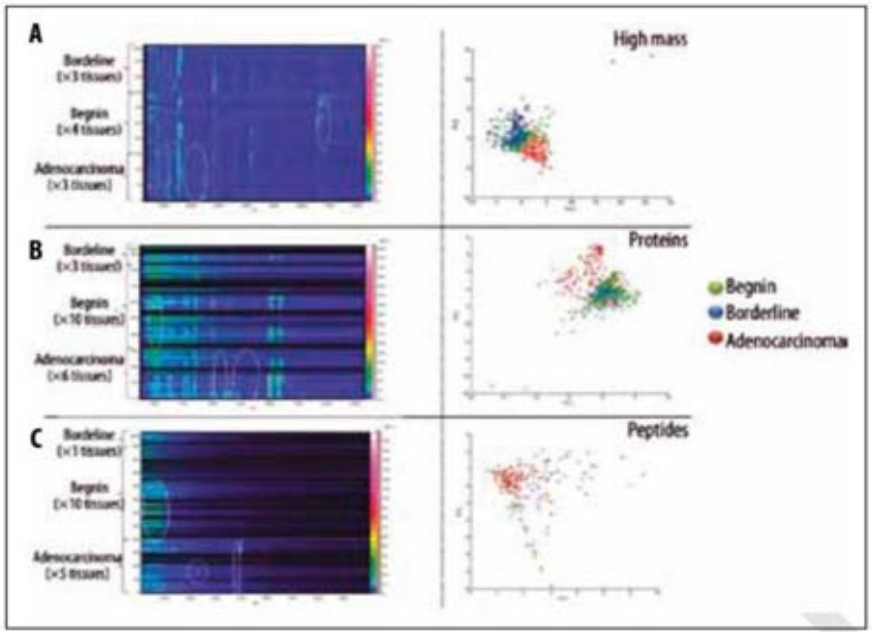



**Figure 1.4.** Figure from El Ayed *et al.* 2010 showing MALDI-TOF MS profiling on three classes of ovarian tissue (adenocarcinoma, borderline and benign). Plots of m/z against spectral source as well as loadings plots from principal component analysis are included for high mass proteins (**A**), small proteins (**B**) and peptides (**C**). Figure reprinted, with permission, from El Ayed *et al.* 2010 **[30]**.

## 1.9 Profiling *vs.* Imaging

Sample preparation for profiling involves deposition of larger (100–500 nL) drops of matrix onto discrete positions of the tissue section. In contrast, IMS methods require nebulization of a homogeneous layer of matrix or deposition of a rectangular array of smaller droplets (0.1–0.2 nL) onto tissue sections (**Figure 1.2**, bottom panel). The benefit of IMS is that instead of documenting profiles for singular locations, the relative abundance (based on MS signal intensity) of hundreds of protein or peptide ions is mapped across an entire tissue section at a centre to centre acquisition distance of 250 µm or smaller. This is achieved by combining all spectra, acquired from a matrix array coating a single tissue section, into a sum spectrum. Mass filters are applied to the sum spectrum, which subsequently mines data from the individual spectra in the data set, presenting the normalized intensity of individual mass ranges as a 2-D heat map (see **Figure 1.3d**). It is this heat map, otherwise known as an ion intensity map, which allows

visualization of peptide and protein distribution across a tissue section. It is thus possible to document changing molecular profiles as tissue composition changes, a process which can be likened to molecular histology **[55]**.

## 1.10 Software for Data Analysis

Several software platforms are currently available, which generate ion intensity maps from spatially referenced IMS data. A selection of IMS software platforms that were available as of 2008 were listed in Jardin-Mathe *et al.* (2008) **[56]** along with important features. Most vendors offer IMS software packages for their MS instruments, including Shimadzu Biotechnology (Intensity Mapping software), AB-SCIEX (TissueView software), Bruker Daltonics (flexImaging software), Waters (conversion tool to use BioMAP) and Thermo Fisher Scientific (ImageQuest software).

Additional freeware programs available include Novartis BioMap, data cube explorer (AMOLF, Amsterdam, Netherlands), fxSpectViewer (CEA, Saclay, France), Mirion (Justus Liebig University, Giessen, Germany) and the MALDI imaging team imaging computing system (MITICS, Lille, France) **[56]**. Of the mentioned software, only BioMap and data cube explorer are readily available by download. Data cube explorer, for example, uses a universal IMS file format imzML to avoid problems with file compatibility for newly developed software. ImzML, is based on the proteomics standard mzML **[57]** and is currently being proposed as a global IMS standard because it maintains the spatial coordinate system of IMS data in a universally recognizable format – in this case a smaller file for meta data and a larger binary file for the MS data (see http://www.maldi-msi.org/ and Römpp *et al.* (2011) **[58]**). The widespread use of imzML would allow IMS researchers to directly access publicly available datasets, compare data sets to their own, and compile analysis scripts to accompany published data. Both BioMap and data cube explorer are available for download from http://www.maldi-msi.org/, along with tutorials on usage.

## 1.11 Automated Sample Preparation for Imaging Cancer Tissues

IMS matrix deposition can be achieved using manual deposition of dry matrix powder via a sieve or sublimation, nebulizing instruments such as handheld air brushes **[55]** and the Bruker Daltonics ImagePrep station **[39]**, or printers such as the Labcyte Portrait **[59]** and Shimadzu Chemical Inkjet Printer (ChIP-1000) **[60]**. **Table 1.5** summarizes important features, advantages and disadvantages of the four most common matrix deposition methods.

**Table 1.5.** Summary of reproducibility, acquisition resolution, the advantages and the disadvantages of four different matrix deposition methods are listed here for air brushes and the ImagePrep station (matrix nebulizing/spray instruments) as well as the ChIP-1000 and Labcyte Portrait (matrix printing instruments).

| | Nebulising instruments | | Printers | |
|---|---|---|---|---|
| | **Air brush** | **ImagePrep station** | **ChIP-1000** | **Labcyte Portrait** |
| **Reproducibility** | Poor | Good | Excellent | Excellent |
| **Acquisition resolution** | ≥5 µm | ≥20 µm | ≥150 µm | ≥150 µm |
| **Advantages** | Cheap | Automated | Automated | Automated |
| | High resolution MS acquisition | High resolution MS acquisition | Control over reagent volume deposited | Control over reagent volume deposited |
| | Good for start up imaging MS laboratories | Default methods available but methods can be modified by user | High MS sensitivity | High MS sensitivity |
| **Disadvantages** | Lower peptide/protein incorporation into matrix | Lower peptide/protein incorporation into matrix | Expensive | Most expensive |
| | Requires experienced user | Requires experienced user | Time consuming preparation | Time consuming preparation |
| | Manual preparation | Expensive | Lower data acquisition resolution than nebulised preparations | Lower data acquisition resolution than nebulised preparations |

Similar to matrix choice, deposition method can also be bio-molecule specific. For example, dry deposition or sublimation of matrix leads to poor incorporation of larger molecules such as peptides and

proteins into matrix crystals because there is no extracting solvent. As a result this type of deposition is typically employed for IMS of metabolites and lipids, which have a higher ionization efficiency. The air brush and ImagePrep station are more efficient in terms of sample incorporation into the matrix crystals and are suitable for all bio-molecule types (optimization of methods may be necessary). However, only experienced users should attempt air brush deposition of matrix, as volume, flow and subsequently reproducibility are difficult to control (see **Table 1.5**). Greater control is possible for nebulization using the ImagePrep, where matrix solution is gravity fed onto a porous metal film, which is vibrated by current flow through an attached piezoelectric sheet. As a result, the matrix is vaporized (nebulized) and settles as a dense mist onto the tissue section. Matrix deposition in the ImagePrep is controlled by measurement of light scatter, which increases with greater crystal density. These nebulized preparations generate a homogeneous matrix field where the spatial acquisition resolution is usually limited to 20–50 μm for the homogeneous matrixes (CHCA and SA, see **Table 1.4**), although higher resolution work has been reported for manual spray preparations **[55]**. A steady loss of MS sensitivity (*i.e.*, ion count intensity) is experienced as resolution is increased, as a result of the smaller area and therefore smaller amount of sample being analyzed. Importantly, the push button functionality of the ImagePrep and its standardized methods make it a viable candidate for clinical application.

Printed IMS arrays are, in effect, whole tissue profiling experiments generated by repeated deposition of picolitre volumes of matrix in a rectangular grid (see **Figure 1.2**). Deposition of matrix in this manner limits users to a maximum acquisition of resolution determined by the droplet size on the tissue which can vary from 150-250 μm, centre to centre, depending on the quality of the preparation.

The ChIP-1000, for example, uses a pressure manifold to maintain solution in a reservoir mounted on top of a 55 μm printing nozzle. Droplets ranging from 100–200 pL are ejected using force generated by current flow through a piezo electric material. The principal down side to the ChIP-1000 is the nozzle itself, which can clog with crystallized matrix. In terms of printing, DHB is the most stable ChIP-1000 matrix. Because DHB is water soluble, and water is not as volatile as organic solvents, printing can be performed for hours without direct supervision. However, several solid ionic matrixes (e.g., CHCA with

molar excess of aniline, see **Table 1.4**) have been developed that increase print stability by reducing the rate of CHCA and SA crystallization **[61]**. The gold standard for printed arrays is a nozzle free system such as the Labcyte Portrait printer, which focuses sound waves at the surface of a matrix solution. Turbulence at the surface ejects droplets (of similar size to the ChIP-1000) vertically onto the tissue section which is suspended, face down, above the solution tray **[59]**. The high cost of this instrument and methods to overcome matrix clogging issues on the ChIP-1000 have unfortunately prevented widespread application of the Portrait.

As already discussed, there is a balance between sensitivity and spatial resolution. Because the volume of matrix deposited is greater for printers than nebulizing instruments, sample extraction efficiency is also greater, leading to improved MS sensitivity. However, for the purposes of a grading approach there is typically no reason to implement the highest resolution nebulized IMS methods, with most studies settling for a 100–200 μm spatial resolution **[39]**. Moreover, deposition of printed arrays or singular spots onto a tumour section, whether guided by histology or independent of it, is more than sufficient to generate MS profiles for grading and biomarker detection **[28, 30, 62]**.

## 1.12 Peptide Imaging Provides Data Complementary to Protein Imaging

Despite advances such as HFIP solvent for improved protein extraction, for the moment, IMS is limited to masses below 70 kDa **[31]**, preventing ready detection of higher molecular weight proteins such as cell surface receptors. Moreover, MS sensitivity decreases dramatically as protein mass increases. Consequently, only the very highest abundance high molecular weight proteins will be observed. To circumvent these issues, it is possible to perform *in situ* proteolytic digests by deposition of enzymes such as trypsin. The digested tissue is coated with matrix (homogeneous layer or printed array) and MALDI-TOF MS acquisition is performed in the peptide mass range (0–6 kDa, see **Figure 1.1**). The resulting peptide MS spectra (**Figure 1.3c**) are vastly more complex than the protein level (**Figure 1.3a**).

However, digest methods allow (i) higher molecular weight proteins to be analyzed via their component peptides, (ii) fragmentation of highly abundant peptides directly from tissue to gain sequence information **[63]** and (iii) direct extraction from the tissue and identification using LC-MS methods, which are well established in most proteomics facilities **[64]**.

## 1.13 Using Histology to Guide Imaging Mass Spectrometry

After MS acquisition is complete, the matrix crystals can be removed using ethanol to allow histological staining and assessment by a pathologist. MS compatible stains such as cresyl violet can also be used prior to IMS to guide analyses **[47, 48]**. Importantly, good correlation between MS ion intensity maps and anatomical structures has been demonstrated previously for various tissues including neuroendocrine **[65]**, breast **[39]** and ovarian cancer (see **Figure 1.3d**) **[64]**. This correlation shows the value of IMS as a complement to histology.

The ability to correlate histology and IMS data was exploited recently for investigation of the changing molecular profile of tumour interfaces. Upon analysis of the tumour boundaries of renal cell carcinoma, the definition of "normal" surrounding tissue has been called into question, with demonstration of potential tumour associated protein changes appearing well past the histological tumour margin **[66]**. In a separate study on serous ovarian carcinomas, IMS was used to show that the tumour interface zone contains a unique set of MS detectable masses as compared to tumour and surrounding normal tissue **[29]**. IMS can thus generate molecular data which is unique and novel to that provided by morphology alone.

## 1.14 Ovarian Cancer Biomarker Discovery Using Imaging Mass Spectrometry

In developing the IMS technology, preliminary IMS biomarker discovery projects for ovarian cancer have been reported by the same group that presented profiling work on ovarian carcinomas (see **section 1.8**). Putative biomarkers of ovarian cancer were detected using printed IMS arrays and subsequently identified using LC-MS/MS of digested cancerous tissue **[30]**. These included 11S proteasome activator complex Reg Alpha fragment **[28]**, oviductin (mucin-9) **[30]** and orosomucoid **[30]**, the roles of which are described briefly here. Reg-Alpha, or PA28, is an antigen processing protein, an increased expression of which may allow presentation of self peptides on tumour cells, and subsequently immune evasion **[28, 30]**. Oviductin is a marker of oviductal epithelium and tubal differentiation marker **[30, 67]**, and finally orosomucoid is an acute phase protein previously evaluated as a marker of ovarian cancer and possible immune suppressor through action on T lymphocytes **[30, 68]**. It is clear that relevant markers of disease can be identified. However, more work is required to determine how effectively MS profiles of such markers can distinguish the subtypes of ovarian cancer and how well these markers translate to cancer detection and screening.

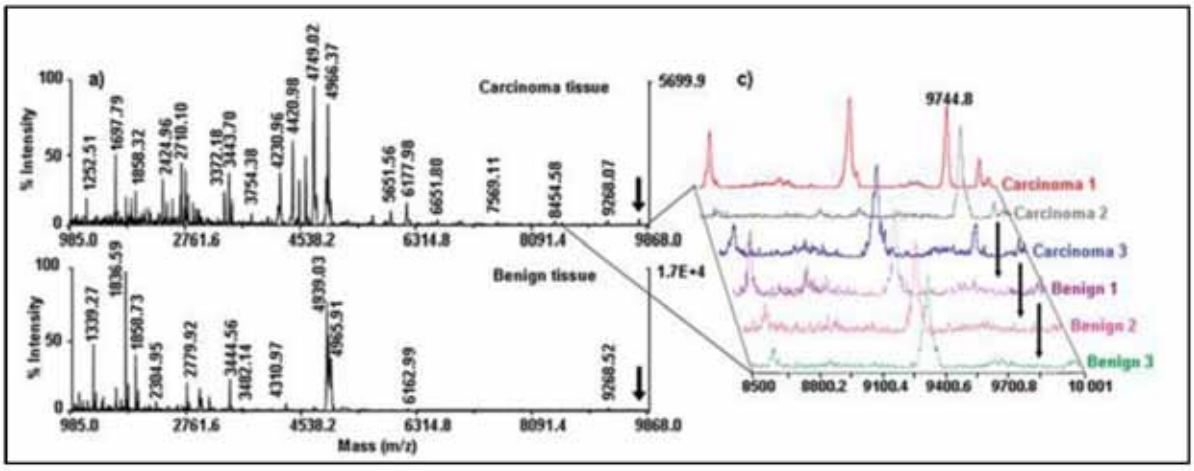



**Figure 1.4.** MALDI-TOF mass spectrometry profiles of ovarian carcinoma and benign tissue sections (a). The individual profiles of three carcinomas and three benign tumour sections are shown in (c), with the mass at 9744.8 m/z highlighted as the potential biomarker 11S proteasome activator complex Reg Alpha fragment. Figure reprinted, with permission, from Lemaire et al. (2007) **[28]**.

## 1.15 Application of Tryptic Digestion to Formalin-Fixed Paraffin Embedded Ovarian Tissues

Frozen tissue represents the current gold standard for IMS, given that a freshly preserved tissue will harbor a freshly preserved proteome, which is easy to access using standard methods for both protein and tryptic peptide IMS (see **Figure 1.3**). However, the limited archival life of frozen tissue (maximum two years for proteomics applications) has forced researchers to adopt methods for accessing peptide and protein mass data in formalin-fixed paraffin embedded (FFPE) tissue; the current global standard for long term tissue preservation in medical centres and research laboratories worldwide **[69-71]**. However, formalin fixation induces cross-linking between multiple amino acid side chains, creating a linked protein network **[72]**. To access these tissues by MS, antigen retrieval (AR) **[62, 64]** and/or *in situ* tryptic digestion **[73, 74]** are required. While AR is not completely understood, its most likely effects are to partially hydrolyze cross-links and denature linked proteins. Typically, this is insufficient for

subsequent MS acquisition of the same quality as frozen tissue, because the cross-linking is not reversed completely. Thus, AR is usually followed by tryptic digestion **[62]**.

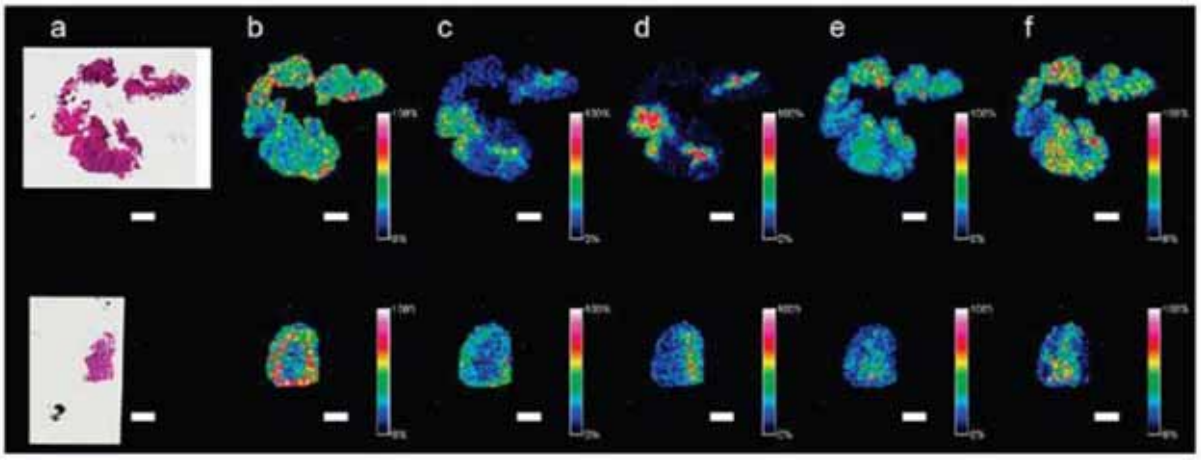



**Figure 1.5.** MALDI-TOF IMS of formalin-fixed paraffin-embedded (FFPE, top row) and frozen (bottom row) ovarian carcinoma. FFPE sections were treated with antigen retrieval. Frozen sections were washed using a previously described protocol. Both sections were digested with trypsin and coated with CHCA matrix using an ImagePrep station. Figure reprinted, with permission, from Gustafsson *et al.* 2010 **[64]**. Scale bars = 2 mm.

Several publications have so far demonstrated successful application of AR methods for IMS on various tissues **[62, 74, 75]**, including FFPE ovarian cancer **[64, 73]**. A 2007 study applying tryptic digestion alone to de-paraffinized and rehydrated sections of ovarian cancer showed that many high abundance proteins could be identified directly from tissue **[73]**. However, this study was a proof of principle application and as such did not demonstrate disease specific distribution of peptides generated from the tissue sections. Our own group has successfully applied citric acid antigen retrieval to ovarian cancer for tryptic peptide IMS (see **Figure 1.5**). Peptides were also extracted from *in situ* proteolytic digests and identified using liquid phase peptide separation and MS. Using this method it was possible to assign tentative identities to 48 individual peptides **[64]**. Because of the ability to rapidly extract and identify peptides directly from tissue using “classical” fractionation-based proteomics, the translation from peptide IMS data to peptide identification and subsequent *in situ* validation by IHC becomes less labour

intensive. Furthermore, the large existing archives of FFPE ovarian tissue will allow any acquired peptide IMS data to be matched to patient history and clinical outcome.

## 1.16 Conclusions and Future Prospects

The successful application of MS profiling and IMS for tissue classification and grading has been demonstrated for different types of cancer. In the case of ovarian cancer, preliminary studies have isolated and identified potential tissue specific peptide and proteins masses using IMS. The current aim is continued application of *in situ* MS methods to demonstrate acquisition of ovarian cancer grade and/or subtype specific protein/peptide profiles from both frozen and archived FFPE tissues. Importantly, from further investigation, IMS derived markers could be used to track molecular changes across ovarian tumours as well as their interfaces with normal tissue to determine the importance of subsequent protein and peptide masses as tissue markers. Following selection of these specific markers, identification, as already demonstrated in several publications, can be achieved using classical proteomics at either the protein or peptide level (fractionation/identification methods such as LC-MS). Subsequent validation of these masses using IHC will ultimately indicate the suitability of masses for further development as diagnostic markers for ovarian cancer sub-type or grade and for validation in large patient cohorts as biomarkers.

# Thesis synopsis and aims

As outlined at the conclusion of the introductory chapter (**see 1.16 above**), MALDI-IMS has provided a means to investigate the molecular composition of cancerous tissue relative to surrounding biological structures. Most importantly the molecular composition of the tissue can yield important clinical information including the grade of cancers and patient prognosis. In order to confidently apply MALDI-IMS to characterizing the molecular MS profile of human ovarian cancer, it was necessary to achieve four aims.

1. Develop MALDI-IMS into a platform for reproducible tracking protein ion changes across sections derived from frozen tissue (**chapters 3-4**).

2. Apply developed MALDI-IMS methods to formalin-fixed paraffin-embedded (FFPE) tissue and optimize to achieve a reproducible platform for imaging FFPE tissues (**chapter 5**).

3. Apply LC-MS/MS for the identification of protein/peptide ions of interest (**chapters 5-7**).

4. Application of developed MALDI-IMS methods to the characterization of peritoneal metastases of serous epithelial ovarian cancer (**chapter 7**).

# Chapter 2 Materials and methods

## 2.1 Materials and reagents

### 2.1.1 Solvents, chemicals and matrixes

| Solvent/chemical | Company/source | Product number |
|---|---|---|
| 2,5-dihydroxybenzoic acid (2,5-DHB) | Bruker Daltonics, Bremen, Germany | 201346 |
| 3,5-dimethoxy-4-hydroxybenzoic acid (SA) | Bruker Daltonics, Bremen, Germany | 201345 |
| α-cyano-4-hydroycinnamic acid (CHCA) | Bruker Daltonics, Bremen, Germany | 201344 |
| Acetonitrile (ACN) | Merck, Darmstadt, Germany | 1.00030.2500 |
| Ammonium bicarbonate ($NH_4HCO_3$) | Merck, Poole, England | 103025E – 1kg |
| Aniline ACS reagent | Sigma-Aldrich, Steinheim, Germany | 242284-5g |
| Citric acid monohydrate | Sigma-Aldrich, Japan | C0706 – 500g |
| Chloroform | Merck, Darmstadt, Germany | 1.02445.2500 |
| Depex mounting medium | BDH, Poole, UK | 361254D |
| EDTA | Sigma, Steinheim, Germany | 431788 – 100g |
| Eosin (chapter 4/5) | Sigma-Aldrich, MO | E4382 – 25g |
| Eosin (chapter 6/7) | Sigma-Aldrich, Steinheim, Germany | HT110332 – 1L |
| Ethanol (EtOH) | Merck, Victoria, Australia | 4.10230.2511 |
| Formic acid (FA) | Sigma-Aldrich, Steinheim, Germany | 21909098 |
| Haematoxylin (chapter 4/5) | BDH, Poole, UK | 340374T |
| Haematoxylin (chapter 6/7) | Sigma-Aldrich, Steinheim, Germany | GH5232 – 1L |
| Isopropanol (IPA) | BDH, Poole, England | 1525265 |
| Methanol (MeOH) | Merck, Darmstadt, Germany | 1.06018.2500 |
| Pertex mounting medium | Medite Midizintechnik, Burgdorf, Germany | PER40000 |

| | | |
|---|---|---|
| Safsolvent | Ajax Finechem, Raven Hills, Australia | N/A |
| Sodium hydroxide (NaOH) | Merck, Darmstadt, Germany | 1.06498.0500 |
| Trifluoroacetic acid (TFA) | Merck, Darmstadt, Germany | 1.08262.0100 |
| Tris | Merck, Darmstadt, Germany | N/A |
| Ultrapure water ≥18.2 MΩ | BarnsteadInternational, Dubuque, IA | D11951 |
| Xylene | Chem-Supply, Gillman, South Australia | 9326410029266 |

### 2.1.2 Consumable materials

| Consumable | Company/source | Product number |
|---|---|---|
| μ-C18 Zip Tips | Millipore, Billerica, MA | ZTC18M096 |
| C18 spin columns | Thermo-Fisher Scientific, Rockford, IL | 89870 |
| Stainless steel cryotome blades (35 x 80 mm) | Feather Safety Razor Co., Japan | S35 |
| Filter paper sheets (ChIP-1000) | Shimadzu Biotechnology, Japan | 241-94075-01 |
| ImagePrep slide cover slips | Bruker Daltonics, Bremen, Germany | 267942 |
| ImagePrep spray generators | Bruker Daltonics, Bremen, Germany | 261614 |
| Indium Tin Oxide coated glass slides | Bruker Daltonics, Bremen, Germany | 237001 |
| Narrow range pH paper (8.2-10) | Merck, Darmstadt, Germany | 1.09558.0003 |
| OSV solution cups (ChIP-1000) | Shimadzu Biotechnology, Japan | 241-08726-92 |
| PEN membrane slides (micro-dissection) | MicroDissect, Herborn, Germany | 11505158 |
| Safe-Lock polypropylene tubes (0.5 mL) | Eppendorf, Hamburg, Germany | 0030 121.023 |
| Safe-Lock polypropylene tubes (1.5 mL) | Eppendorf, Hamburg, Germany | 0030 120.086 |
| ImagePrep solution vials | Bruker Daltonics, Bremen, Germany | N/A |
| Syringe plunger unit (ChIP-1000) | Shimadzu Biotechnology, Japan | 241-02735-91 |

### 2.1.3 Proteolytic enzymes

| Enzyme | Company/source | Product number |
|---|---|---|
| Modified porcine Trypsin | Promega, Madison, WI | V5111 |
| Trypsin Gold | Promega, Madison, WI | V5280 |

### 2.1.4 Peptide and protein calibration standards

| Standard | Sequence | Company | Product number |
|---|---|---|---|
| ACTH (1-24) | SYSMEHFRWGKPVGKKRRPVKVYP | Sigma-Aldrich, St. Louis, MO | A0298 |
| ACTH (18-39) | RPVKVYPNGAEDESAEAFPLEF | Bruker Daltonics, Bremen, Germany | 225593 |
| Angiotensin I (34-43) | DRVYIHPFHL | BioRad | C10-00002 |
| ClinProT standards | Proprietary peptide/protein mixture | Bruker Daltonics, Bremen, Germany | Protein–206355<br>Peptide–222570 |
| Dynorphin A (1-17) | YGGFLRRIRPKLKWDNQ | Auspep, Victoria, Australia | 2032 |
| [Glu$^1$]-Fibrinopeptide B | EGVNDNEEGFFSAR | Sigma-Aldrich, St. Louis, MO | F3261 |
| Peptide calibration standard II | Proprietary peptide mixture | Bruker Daltonics, Bremen, Germany | 222570 |
| Renin substrate | DRVYIHPFHLVIHN | Bruker Daltonics, Bremen, Germany | 225593 |

### 2.1.5 Instruments/equipment

| Instrument | Company/source |
|---|---|
| Agilent 1100 LC system | Agilent Technologies, Japan |
| Centrifuge 5810R | Eppendorf, Hamburg, Germany |
| Chemical Inkjet Printer (ChIP-1000) | Shimadzu Biotechnology, Japan |
| ImagePrep station | Bruker Daltonics, Bremen, Germany |
| Leica AS LCM microscope | Leica Instruments, Germany |
| Leica CM1850 cryotome | Leica Instruments, Germany |
| Microm HM325 microtome | Zeiss, Germany |
| MTP 384-800 µm AnchorChip target | Bruker Daltonics, Bremen, Germany |
| MTP slide adaptor II | Bruker Daltonics, Bremen, Germany |
| NanoZoomer automated microscope | Hamamatsu, Beijing, China |
| nLC-HCT-Quadrupole Ion Trap | Bruker Daltonics, Bremen, Germany |
| nLC-LTQ-Orbitrap | Thermo-Fisher Scientific, Bremen, Germany |
| Powerbook III Prepress digital scanner | UMAX, Taiwan |
| Proteineer Fraction Collector | Bruker Daltonics, Bremen, Germany |
| ultraflex III MALDI-TOF/TOF | Bruker Daltonics, Bremen, Germany |
| ultrafleXtreme MALDI-TOF/TOF | Bruker Daltonics, Bremen, Germany |

## 2.2 Chapter three methods

NB: Methods shown here are presented as they appeared in the manuscript.

### 2.2.1 Experimental Tissues and Solvents

Murine brain tissue was scavenged from on going experiments in the Chemokine laboratory at the University of Adelaide. All source experiments had full ethics approval from the University of Adelaide Ethics Committee. Solvents, acetonitrile (ACN) and methanol (MeOH) were HPLC grade (Merck, Darmstadt, Germany). Ethanol (EtOH) and isopropanol (IPA) were analytical grade or higher (Merck). Trifluoroacetic acid (TFA) was purchased from Applied Biosystems (Foster City, CA). Sinapinic acid (SA) and peptide/protein standards (ClinProT standards, Bruker Daltonics) used to calibrate the Ultra Flex III MALDI-TOF/TOF mass spectrometer were purchased from Bruker Daltonics (Bremen, Germany).

### 2.2.2 Sample Preparation

Mice were sacrificed by carbon dioxide asphyxiation and tissues of interest were excised within ten minutes of sacrifice. Tissues were slow frozen in foil floated on liquid nitrogen ($N2_{(l)}$) and transferred on dry ice to a -80°C freezer. Tissues were equilibrated to -20°C, followed by sectioning at -20°C (brain) or -18°C (other organs). Tissue was sectioned at a thickness of 10 µm with 35 x 80 mm blades in a Shandon cryotome (Thermo Electron, Pittsburgh, PA). Sections were mounted onto pre-chilled Indium Tin Oxide (ITO) slides (Bruker Daltonics) using heat from the operators hand applied underneath the slide. Mounted sections were desiccated for 45 minutes prior to washes for 60 seconds in 100ml of 70% EtOH or IPA followed by 60 seconds submerged in 100ml of 100% EtOH or IPA. Following washes

sections were desiccated for 15 minutes. Use of IPA in washes was an alteration to protocols due to published data (Seeley, Oppenheimer *et al.* 2008)**[42]**.

### 2.2.3 Matrix Deposition – CHIP-1000

The CHIP-1000 piezo electric printer (Shimadzu Biotechnology, Kyoto, Japan) was operated as per manufacturer's instructions to deposit 250 µm centre to centre arrays of SA matrix directly onto tissue sections. Dwell voltage and dwell time were modulated as necessary to maintain the quality of droplets for deposition. Prior to printing the piezo printer head was flushed with several volumes of 50% IPA. Following printing the piezo printer head was flushed with 200 µl of 50% IPA and the outside rinsed with 50% IPA (3x), 100% MeOH (3x) and 100% ultra pure water (Resistivity 18.2 MΩ/cm, TOC < 1ppb) (3x). SA matrix at 10 mg/ml concentration in 50% ACN, 44.8% ultra pure water, 5% IPA and 0.2% TFA was used. Number of droplets and layers was variable.

### 2.2.4 Matrix Deposition – ImagePrep

The ImagePrep station (Bruker Daltonics) was operated as per manufacturer's instructions to deposit homogeneous matrix layers onto tissue sections. SA at concentrations of 6, 8 and 10 mg/ml in 50% ACN, 5% IPA and 0.2% TFA were deposited using the default SA ImagePrep method (Bruker Daltonics). Microscopic slide images were captured using a Nikon pathology microscope connected to a digital camera.

### 2.2.5 MALDI-TOF Imaging Mass Spectrometry and Data Analysis

Slides for automated MALDI-TOF analysis were scanned at 1200 dpi using a Powerbook III PrePress Digital colour scanner (Amersham Biosciences, Soeborg, Denmark) controlled by Magic Scan software (version 4.6, UMAX). Slides were fitted into a Slide Adapter II MALDI target (Bruker Daltonics). An Ultra Flex III MALDITOF/ TOF mass spectrometer (Bruker Daltonics) operating in linear mode was used for IMS acquisition. Scanned slide images were loaded into Flex Imaging software (version 2.0, Bruker Daltonics), which was used to generate an auto execute sequence and set teach points for each individual IMS experiment. Homogeneous matrix layers generated by the ImagePrep were analyzed at a raster suited to the quality of the preparation. Droplet arrays generated by the CHIP-1000 were analyzed at a raster of the same size as the centre to centre droplet distance (250 μm). Auto execute parameters were set by Flex Control software (version 3.0, Bruker Daltonics) and a fixed laser power was selected by the operator. Results from IMS acquisition were observed in Flex Imaging and Flex Analysis software (version 3.0, Bruker Daltonics).

## 2.3 Chapter four methods

### 2.3.1 Chemicals and reagents

Comprehensive list of chemicals, reagents and instruments appears in methods **section 2.1**.

### 2.3.2 Tissue source

Murine brain was dissected from mice sacrificed by carbon dioxide asphyxiation. All mice were sacrificed for projects with full ethics approval from the University of Adelaide Ethics Committee. Tissue

was immediately frozen in a foil boat floated on liquid nitrogen. Human ovarian cancer tissue was obtained from a patient who underwent surgery at the Department of Gynaecological Oncology, Royal Adelaide Hospital, Adelaide, Australia. Tissue was used for MALDI-IMS with full ethics approval.

### 2.3.3 Tissue sectioning

Prior to sectioning, all tissues were allowed to equilibrate to -20ºC for one hour. Tissues were sectioned at a thickness ranging from 6-10 µm using a Leica CM1850 cryotome (Leica Instruments, Germany). Tissue specific temperatures were -18 to -20ºC for murine brain and -25 to -35ºC for human ovarian cancer tissue. All sectioned tissue was thaw mounted onto indium tin oxide (ITO) slides until the tissue sections appeared macroscopically dry (matte appearance). Sections were desiccated for 15-30 minutes prior to chemical section treatment as outlined in **chapter 4.1**.

### 2.3.4 Tryptic digestion and matrix deposition methods (ChIP-1000)

A ChIP-1000 instrument (Shimadzu Biotechnology, Japan) was used to deposit solutions, as described in detail in **chapter 4.2**. Trypsin solutions were typically deposited in 25 layers using two 100 pL drops per layer. The ChIP-1000 was also used to deposit matrix solutions. See **chapter 4.4** and **4.5** for details of matrix deposition using the ChIP-1000.

### 2.3.5 Matrix deposition (ImagePrep station)

An ImagePrep station (Bruker Daltonics, Bremen, Germany) was used to deposit 10 mg/mL SA in 60% ACN and 0.2% TFA (**Table 2.1**). Preparation stability was determined using a light scatter curve generated in MS Excel (2007) from extracted ImagePrep logs (ASCII text file). Formatted light scatter curves were generated for this thesis using R **[76]**.

**Table 2.1.** ImagePrep phase settings for deposition of 10 mg/mL SA in 60% ACN and 0.2% TFA.

| Phase | Sensor | Nebulization | Incubation | Drying |
|---|---|---|---|---|
| 1 | 0.5V within 8-20 cycles | 25% ± 25% power with fixed spray time of 2.5 s | 5 s | 120 s |
| 2 | 30 s drying | | | |
| 3 | 0.2V within 6-22 cycles | 20% ± 20% power with 0.10V sensor controlled spray time | 30 s ± 30 s | Grade 20% ± 60% complete dry every 2 cycles, safe dry 10 s |
| 4 | 0.3V within 12-32 cycles | 20% ± 25% power with 0.2V sensor controlled spray time | | Grade 20% ± 60% complete dry every 4 cycles, safe dry 40 s |
| 5 | 0.25V ± 0.20V within 4-64 cycles | 25% ± 25% power with 0.3V sensor controlled spray time | | Grade 20% ± 60% complete dry every 4 cycles, safe dry 40 s |

### 2.3.6 MALDI-TOF/TOF imaging mass spectrometry

Mass spectra for all experiments on murine brain were acquired on an ultraflex III MALDI-TOF/TOF MS instrument (Bruker Daltonics, Bremen, Germany) operating in either linear or reflectron positive ion mode. Linear positive mode instrument specific settings were as follows: laser repetition rate – 200 Hz; centre to centre acquisition distance – 100 µm (ImagePrep), 250 µm (ChIP-1000); Mass range (Da) – variable; no. laser shots per position – variable. Reflectron positive mode instrument specific settings were as follows: laser repetition rate – 200 Hz; centre to centre acquisition distance – 100 µm (ImagePrep), 250 µm (ChIP-1000); mass range (Da) – 800-5000; no. laser shots per position – 500. Mass spectra for experiments on ovarian tumour tissue sections were acquired on an ultrafleXtreme MALDI-TOF/TOF MS instrument (Bruker Daltonics, Bremen, Germany) operating in reflectron positive ion mode. Reflectron positive instrument specific settings were as follows: laser repetition rate – 1000 Hz; centre to centre acquisition distance – 250 µm; Mass range (Da) – 800-5000; no. laser shots per position – 500.

### 2.3.7 Data processing

MS spectra were smoothed (chemical noise filter with width of *m/z* 5) and baseline subtracted (TopHat) in flexAnalysis (V3.3, Bruker Daltonics, Bremen, Germany). Sum spectra and ion intensity maps were generated in flexImaging (V2.1, Bruker Daltonics, Bremen, Germany).

### 2.3.8 Haematoxylin and eosin staining

Matrix was eluted from ITO slides following MALDI-TOF MS data acquisition using 100% EtOH (5 min wash). The sections were dried to ensure complete removal of matrix and immersed in haematoxylin (BDH, Pool, UK) for 30-45 s, rinsed in distilled water and washed for one min in Scotts tap water substitute (23.8 mM NaHCO3 with 166.1 mM MgSO4). Following another 60 s wash in distilled water, sections were stained with eosin (Sigma-Aldrich, MO) for 30 s. Excess stain was rinsed off with distilled water and the slides were washed in 70%, 95%, 100% and fresh 100% EtOH for two min each. After two final washes in safsolvent (Ajax Finechem, Raven Hills, Australia) sections were allowed to dry for two min before cover slips were mounted using Depex mounting medium (Merck, Poole, UK). Microscopy was performed using a Nanozoomer automated slide scanner (Hamamatsu, Beijing, China).

## 2.4 Chapter five methods

NB: Methods shown here are presented as they appeared in the manuscript.

### 2.4.1 Chemicals and Reagents.

Trifluoroacetic acid (TFA) was obtained from Applied Biosystems (Warrington, U.K.), while HPLC grades acetonitrile (ACN) and ammonium bicarbonate ($NH_4HCO_3$) were obtained from BDH (Poole,

U.K.). Citric acid and EDTA were purchased from Sigma (Steinheim, Germany) and Tris was from Merck (Darmstadt, Germany). Modified porcine Trypsin and Trypsin Gold were from Promega (Madison, WI). α-Cyano-4-hydroxycinnamic acid (CHCA) and 2,5-dihydroxybenzoic acid (DHB) matrices were from Bruker Daltonics (Bremen, Germany). All aqueous solutions were prepared using ultra pure (UP) de-ionized water having a resistance of ≥18.2 MΩ/cm. All other reagents/solvents were analytical grade or higher and were used without further purification.

### 2.4.2 Tissue Samples.

FFPE human ovarian cancer tissue was obtained from two patients who underwent surgery at the Department of Gynaecological Oncology, Royal Adelaide Hospital, Adelaide, Australia in 1992 (Carcinosarcoma stage 3C) and 2008 (Serous carcinoma stage 3C) with ethics approval. Tissues were (1) embedded in optimal cutting temperature (OCT) polymer to be frozen with liquid nitrogen or (2) fixed in either 10% formalin (1992 sample) or 4% formaldehyde solution (2008 sample) and embedded in paraffin. The difference between formalin and formaldehyde solutions lies in the concentration of formaldehyde (10% versus 4%, respectively) as well as the presence of methanol for stabilization in the formalin. For FFPE, fresh tissue was incubated in formalin or formaldehyde (Sigma, Germany) for 24 h and washed in water. Tissue was stored for 1-4 weeks in 70% ethanol (EtOH) before batch transfer to a Leica automated processor (Leica Instruments, Germany) for incubation in EtOH/xylene gradients. Processed tissues were embedded in liquid paraffin.

### 2.4.3 Section Preparation and Antigen Retrieval.

Long-term (>week) storage was at -80 °C for all frozen sections. Fresh-frozen OCT embedded tissue was sectioned using a Leica CM1850 cryotome (Leica Instruments, Germany) operating at -21 °C.

Sections were thaw mounted onto pre-chilled indium tin oxide (ITO) slides (Bruker Daltonics, Bremen, Germany), desiccated for 30 min and washed using a previously described OCT removal protocol.**[44]** Slides were dried and stored in a desiccator. FFPE blocks were sectioned using a Microm HM325 microtome (Zeiss, Germany), water bath mounted onto ITO slides and dried at room temperature. CAAR was adapted from a previously described AR method.**[77, 78]** Slides were heated at 60 °C for 1 h and washed in 100% xylene for 5 min, 100% EtOH for 2 min and 10 mM $NH_4HCO_3$ for 5 min (washes performed twice). Slides were immersed in 10 mM citric acid monohydrate at pH 6 and pulse boiled for 10 min in a microwave (solution maintained near boiling point). While still in citric acid, slides were heated at 98 °C for 30 min. Solution was allowed to cool and slides were washed in 10 mM $NH_4HCO_3$ twice (5 min). Control slides were not exposed to 60 °C heating or citric acid steps. For comparison to previously described methods of AR for IMS, a Tris/EDTA method was applied to sections from the same FFPE block.**[62]**

#### 2.4.4 Trypsin Deposition. ChIP-1000 Printing.

For Chemical Inkjet Printer (ChIP-1000, Shimadzu Biotechnology) deposition a previously described protocol was used with modifications.**[62]** Trypsin Gold (0.5 $\mu$g/$\mu$L) in 50 mM acetic acid was diluted to 40 ng/$\mu$L in 100 mM $NH_4HCO_3$. A final concentration of 12% v/v ACN was added to improve printing performance and cause slight protein denaturation. Five nanolitres of this Trypsin solution was printed in 25 layers, with two drops (~200 pL) per layer and a 5 min incubation between each layer.

#### 2.4.5 ImagePrep Nebulization.

A 0.66 ng/$\mu$L solution of modified porcine trypsin in 10 mM $NH_4HCO_3$ and 5% v/v ACN was deposited in 25, 2.1 s cycles using an ImagePrep station (Bruker Daltonics). Tissue was incubated for 4 min and

dried for 60 s between cycles. Sections were scanned using a Powerbook III Prepress Digital scanner (UMAX, Taiwan) and slides were weighed prior to and following matrix deposition.

### 2.4.6 Matrix Deposition. ChIP-1000 Printing.

DHB (20 mg/mL) in 50% methanol (MeOH) and 0.2% TFA was deposited onto all AR sections using the ChIP-1000. Printer operation deposited 12 nL of matrix totaling 40 layers (3 drops/layer) with 100 pL droplet volumes.

### 2.4.7 ImagePrep Nebulization.

CHCA (5 mg/mL) in 50% ACN and 0.2% TFA was deposited onto digested sections using an ImagePrep station and an operator modified Bruker Daltonics CHCA default method, optimized for sensor controlled nebulization of matrix.

### 2.4.8 Imaging MALDI-TOF-MS.

flexImaging (V2.0, Bruker Daltonics) was employed for IMS, using flexControl (V3.0, Bruker Daltonics) to acquire data on an ultraflex III MALDI-TOF/TOF system (Bruker Daltonics). Protein IMS used linear positive ion mode and peptide IMS used reflectron positive ion mode. Laser power and instrument specific settings were operator determined. External calibration was performed using either peptide calibration II (PC II, Bruker Daltonics) diluted 1:100 with 0.1% TFA and co-crystallized with matrix or ClinProT standards (CPS, Bruker Daltonics) diluted 1:20 in matrix. One thousand (ChIP-1000 protein and peptide IMS) or 300 (ImagePrep peptide IMS) MS shots were collected at each position, depending on experiment type. ChIP-1000 MS was performed at a raster determined by the center to center distance of the printed arrays. ImagePrep MS was performed at a raster of 200 $\mu$m. MS processing was

performed with flexAnalysis (V3.1, Bruker Daltonics). Linear MS smoothing was Gaussian, 4 cycles, width, *m/z* 2. Reflectron MS smoothing was Chemical noise, width, *m/z* 5. All MS spectra were baseline subtracted (TopHat).

### 2.4.9 Imaging Tandem MS.

Manually selected precursor ions of interest (normalized against total ion current using flexImaging V2.0) were chosen for MS/MS analysis. MS/MS was performed with an optimized method referred to as LIFT (Bruker Daltonics) mode using acquisition areas determined by flexImaging. MS/MS spectra were sent to BioTools (V3.1 build 2.22, Bruker Daltonics) and exported to an in-house MASCOT database-search engine (Matrix Science). Search parameters were as follows: Taxonomy, *Homo sapiens*; MS tolerance, 100 ppm; MS/MS tolerance, 0.8 Da; database, Swiss-Prot (V57.7) or NCBI (20090915); enzyme, trypsin/P; optional modifications, methionine oxidation; missed cleavages, 2. Matches were assigned with at least $p < 0.05$ significance threshold.

### 2.4.10 Haematoxylin and Eosin Staining.

Following IMS analysis, matrix was eluted off slides using 100% ethanol for 5 min. Once dry, slides were immersed in hematoxylin (BDH, Poole, UK) for 30 s, washed using distilled water and immersed for 60 s in 23.8 mM $NaHCO_3$ with 166.1 mM $MgSO_4$ (Scotts Tapwater substitute). Slides were then washed in distilled water for 60 s, stained with eosin (Sigma, MO) for 30 s and rinsed using distilled water. Slides were then moved through 2 min EtOH gradients. In order these were 70%, 95%, 100% and fresh 100%. Final washes were performed twice in safsolvent (Ajax Finechem, Raven Hills, Australia) and slides were air-dried for 2 min. Cover-slips were mounted using Depex medium (Merck, Poole, UK) and dried overnight. Slides were scanned using a Nanozoomer automated slide scanner (Hamamatsu, Beijing, China).

### 2.4.11 *In Situ* Digestion and Extraction.

A slide containing archived FFPE sections of interest was placed on a 37 °C heating block and 20 μL of 10 ng/μL modified porcine trypsin in 10 mM $NH_4HCO_3$ was added to the tissue and incubated (30 min). Digest was repeated. Twenty microliters 0.1% formic acid/10% ACN, 2 × 20 μL 50% ACN and 1 × 20 μL 100% ACN were used to sequentially extract peptides by pipetting up and down on the tissue. Extracts were combined, reduced to ~5 μL in a vacuum centrifuge and made up to ~15 μL with 0.1% TFA. Peptides were purified using C-18 Zip Tips (Millipore, MA). Zip Tips were equilibrated with 100% ACN, 50% ACN/0.1% TFA and 0.1% TFA. Sample was loaded onto the Tip and washed with 0.1% TFA. Peptides were eluted in 4 μL 50% ACN, made up to 25 μL using 0.1% TFA and reduced again to ~10 μL.

### 2.4.12 HPLC and Fraction Collection.

An Agilent 1100 HPLC system was operated using binary gradients of mobile phase A (0.1% TFA in 95% UP H20 and 5% ACN) and B (0.1% TFA in 95% ACN and 5% UP H2O). The Agilent 1100 system was controlled using the Hystar software platform (V3.2, Bruker Daltonics). Two microliters of zip-tipped tryptic digest sample was combined with 4 μL of 0.1% TFA. A micro-WPS auto sampler (G1377A) injected a 5 μL sample onto a 0.18 × 150 mm Acclaim Pepmap100 reverse phase HPLC column packed with 3 μm, porous (100 Å) octadecylsilane C-18 coated silica beads (LC Packings, Dionex). Sample loading was under aqueous conditions (0% B) for 5 min with a flow rate of 1 μL/min. Gradients: 5 min 0-12% B gradient, followed by main 36 min, 12-48% B gradient and final 5 min 48-60% B gradient. Fifteen second fractions were collected onto a MTP 384 MALDI 600 μm AnchorChip target (Bruker Daltonics) using a Proteineer Fraction Collector (Bruker Daltonics). Fifty percent ACN with 0.1% TFA (0.5 μL) was deposited in a dis-continuous flow with fractions. One microliter of Bruker Daltonics

PC II (1:100 with 0.1% TFA) was deposited onto calibration spots. Dry sample and calibration spots were crystallized with 0.5 mg/mL CHCA as described previously.**[79]**

#### 2.4.13 MALDI-TOF MS for HPLC Fractions.

MS analysis was performed on an ultraflex III MALDI-TOF/TOF system. MS was collected in reflectron positive ion mode, using WARP-LC (V1.1 Bruker Daltonics) interfaced with flexControl. Laser power was operator determined to provide optimal MS intensity and resolution. Four hundred and 300 MS shots were collected for sample and calibration spots respectively. Acquisition settings: *m/z* range of 440-5000, 12.8 × reflector gain and 1.00 GS/s detector acquisition rate. Following MS collection, MS spectra were smoothed (Gaussian, 2 cycles, width - *m/z* 0.02) and baseline subtracted (TopHat). Peak masses and intensities were detected with flexAnalysis using the SNAP algorithm.

#### 2.4.14 Tandem MS for HPLC Fractions.

WARP-LC was used to calculate a compound list for automated MS/MS based on a preset LC-MALDI method. MS/MS instrument settings were defined by the LIFT method. Following automated MS/MS acquisition, manual MS/MS spectra were collected for masses which match to IMS peptide masses, but were not selected for automated acquisition. Manual and automated MS/MS spectra were exported to the Mascot database (V2.2) using the following search parameters: Taxonomy, all entries or *Homo sapiens*; database, Swiss-Prot (V57.7) or NCBI (20090915); MS tolerance, 50-100 ppm; MS/MS tolerance, 0.5-0.8 Da; Enzyme, Trypsin/P; Variable modifications, Oxidation (M); Missed cleavages, 2. Matches were assigned with at least $p < 0.05$ significance threshold. The MOWSE and probability scores were used as the basis for ion annotation in the MS/MS spectra.

### 2.4.15 HPLC-ESI-Ion Trap MS and MS/MS.

Vacuum concentrated samples were re-suspended with 0.1% FA to total volume of ~5 μL. LC-ESI-IT MS/MS was performed using an online 1100 series HPLC system (Agilent Technologies) and HCT Ultra 3D-Ion-Trap mass spectrometer (Bruker Daltonics). The LC system was interfaced to the MS using an Agilent Technologies Chip Cube operating with a ProtID-Chip, which integrates the enrichment column (Zorbax 300SB-C18, 4 mm, 40 nL), analytical column (Zorbax 300 SB-C18, 150 mm × 75 μm), and nano-spray emitter. Five microlitres of sample was loaded on the enrichment column at a flow rate of 4 μL/min in Mobile phase A (0.1% FA in 2% v/v ACN) and resolved with 1-30% gradient of Mobile phase B (0.1% FA in 98% w/v ACN) over 32 min at 300 nL/min. Ionizable species ($300 < m/z < 1200$) were trapped and the two most intense ions eluting at the time were fragmented by collision-induced dissociation. Active exclusion was used to exclude a precursor ion for 30 s following the acquisition of two spectra. MS and MS/MS spectra were subjected to peak detection and de-convolution using DataAnalysis (V3.4, Bruker Daltonics). Compound lists were exported into BioTools (V3.1, Bruker Daltonics) then submitted to Mascot (V2.2). The search parameters were as follows: Taxonomy, *Homo sapiens*; Database, Swiss-Prot (V57.7); Enzyme, Trypsin/P; Variable modifications, Oxidation (M); MS tolerance, 0.6 Da; MS/MS tolerance, 0.8 Da; Peptide charge, 1+, 2+ and 3+; Missed cleavages, 2.

### 2.4.16 HPLC-ESI-LTQ-Orbitrap MS and MS/MS.

Vacuum concentrate sample used for LC-ESI-IT was diluted 1:5 with 0.1% FA to a total volume of ~10 μL. HPLC-nESI-LTQ-Orbitrap XL analysis was performed using an online Ultimate 3000 nHPLC system (Dionex) and LTQ-Orbitrap XL instrument (Thermo Scientific). One μL of sample was loaded (using a $H_2O$ buffer, 15 min) onto an Acclaim Pepmap100 C-18 μ-precolumn cartridge (5 μm, 100 Å, 300 μm I.D. × 5 mm) at 5 μL/min flow rate. A micropump was brought inline with the precolumn and analytical separation was achieved using an Acclaim Pepmap100 C18 column (3 μm, 100 Å, 75 μm I.D. × 15

cm) running at 0.3 μL/min over a 60 min 0-25% B (0.1% FA in 100% ACN) gradient. Buffer A was 0.1% FA in $H_2O$. Eluted peptides were introduced by ESI into the LTQ-Orbitrap XL and the top six peptides in each scan were fragmented by CID. Parameters were; minimum signal -1000; isolation width -3.00; normalized collision energy -35.0; charge state - ≥+2. Raw data files were exported to Proteome Discoverer software (V1.1, Thermo Scientific) and submitted to Mascot (V2.2). Search parameters were as follows: Taxonomy, *Homo sapiens*; Database, Swiss-Prot (V57.7); Enzyme, trypsin/P; Variable modifications, Oxidation (M); MS tolerance, 5 ppm; MS/MS tolerance, 0.8 Da; Missed cleavages, 2. Results filters used a general Mascot significance threshold of 0.05 and a peptide confidence value set at high (Mascot significance threshold of 0.01).

## 2.5 Chapter six and seven methods

### 2.5.1 Chemicals and Reagents

Comprehensive list of chemicals, reagents and instruments appears in methods **section 2.1**.

### 2.5.2 Tissue Samples

All ovarian cancer tissue samples were obtained during surgery at the Department of Gynaecological Oncology, Royal Adelaide Hospital, Adelaide, Australia with ethics approval. Tissue was fixed in 4% (w/v) formaldehyde (Sigma-Aldrich, Germany) for 24 h. Tissue blocks were then washed with water and stored in 70% EtOH for 1-4 weeks. A Leica automated processor (Leica Instruments, Germany) was used for incubation in a series of EtOH/xylene gradients, prior to embedding of tissue blocks in liquid paraffin.

### 2.5.3 Section Preparation and Antigen Retrieval

Ovarian cancer FFPE blocks were sectioned at a thickness of 6 µm using a Microm HM325 microtome (Zeiss, Germany). Sections were water bath mounted at a water temperature of 39°C onto indium tine oxide (ITO) coated slides and dried at room temperature. Tissue sections were retrieved using a modified published citric acid antigen retrieval (CAAR) protocol **[64]**. Briefly, sections were stripped of paraffin using 100% xylene (2x 5 min) and washed twice in 100% ethanol for 2 min. Clean sections were washed twice in 10 mM $NH_4HCO_3$ (5 min each) and then pulse boiled in 10 mM citric acid at pH 6 for 10 minutes. The slides were incubated on a heating block at 98°C for 30 minutes, removed from the solution and cooled at room temperature. The sections were then rinsed twice in 10 mM $NH_4HCO_3$ for one min each. The slides were dried in a desiccator prior to tryptic digestion.

### 2.5.4 *In situ* tryptic digestion

An ImagePrep station (Bruker Daltonics, Bremen, Germany) running a standard trypsin digest method. A total of 30 cycles ± 5 cycles were deposited at 38% spray power. Spray time was set at 1.25 s with a 45 second dry time. A total of 200 µL 100 ng/µL trypsin gold (re-suspended in 25 mM $NH_4HCO_3$) were onto tissue sections. Once complete, the slide was incubated for 2 hours at 37°C in a humid box.

### 2.5.5 Deposition of internal calibrants

One hundred µL of internal calibrants were deposited using an ImagePrep station running the same method of deposition as trypsin (see **section 2.5.4**). The full list of calibrants used in this chapter appears in **Table 6.1** (see **chapter 6.2.2**). Those used in the optimized calibration mixture were Angiotensin I (0.4 pmol/µL), [$Glu^1$]-Fibrinopeptide B (0.4 pmol/µL), Dynorphin A (2 pmol/µL) and ACTH

fragment (1-24) [2 pmol/µL]. Prior to matrix deposition a scan was taken of the prepared sections using a Powerbook III Prepress Digital scanner (UMAX, Taiwan).

### 2.5.6 Matrix deposition

Seven mg/mL CHCA (in 50% ACN and 0.2% TFA) was deposited using an ImagePrep station (settings appear in **Table 2.2**). Modifications were; phase one minimum cycle number was increased to at least eight and the minimum spray power was set to 25%. Preparation stability was determined using a light scatter curve generated in MS Excel (2007) from extracted ImagePrep data.

**Table 2.2.** ImagePrep settings for deposition of 7 mg/mL CHCA in 50% ACN and 0.2% TFA.

| Phase | Sensor | Nebulization | Incubation | Drying |
|---|---|---|---|---|
| 1 | 0.65V within 8-20 cycles | 20% ± 35% power with fixed spray time of 2.5 s | 10 s | 90 s |
| 2 | 30 s drying | | | |
| 3 | 0.1V within 4-10 cycles | 25% ± 35% power with 0.05V sensor controlled spray time | 30 s ± 30 s | Complete dry every cycle, safe dry 10 s |
| 4 | 0.1V within 8-12 cycles | 25% ± 35% power with 0.1V sensor controlled spray time | | Grade 20% ± 40% complete dry every 2 cycles, safe dry 20 s |
| 5 | 0.3V within 12-30 cycles | 25% ± 35% power with 0.2V sensor controlled spray time | | Grade 30% ± 40% complete dry every 3 cycles, safe dry 40 s |
| 6 | 0.6V ± 0.5V within 20-64 cycles | 25% ± 35% power with 0.3V sensor controlled spray time | | Grade 40% ± 40% complete dry every 4 cycles, safe dry 40 s |

### 2.5.7 MALDI-TOF-MS imaging data acquisition

Mass spectra were acquired on an ultrafleXtreme MALDI-TOF/TOF MS instrument (Bruker Daltonics, Bremen, Germany) operating in reflectron positive ion mode. Relevant instrument specific settings were

as follows: Laser repetition rate – 1000 Hz; centre to centre acquisition distance – 100 µm; m/z range – 1000-4500; deflection mass – 900 Da; detector gain – 2.573 kV; laser shots per position – 500; laser beam focus – 70; ion source voltage 1 – 25 kV; ion source voltage 2 – 22.15 kV; lens voltage – 7.6 kV; reflector voltage 1 – 26.5 kV; reflector voltage 2 – 13.45 kV. MS instrument was controlled by flexImaging software (V2.1, Bruker Daltonics, Bremen, Germany).

### 2.5.8 Data processing

Visual basic (flexAnalysis scripts) and java (IonMapper) code was written with the assistance of James Eddes (Adelaide Proteomics Centre). Acquired profile mass spectra were processed by Gaussian smoothing (two cycles with a *m/z* width of 0.02) and baseline subtracted (TopHat) in flexAnalysis (V3.3, Bruker Daltonics, Bremen, Germany). Peaks were detected using the SNAP II algorithm and exported to ASCII text files using a custom flexAnalysis script. Each text file was saved with the name of its spectral coordinate to allow for spatially based data processing and image generation. A Java based script (NetBeans IDE6.9.1) was used to automate processing of the ASCII peak list files. A size ordered array of all the peaks in every peak list was generated first. The list was then grouped such that if the second peak in the combined peak list was within a set *m/z* tolerance (E.g. 0.2 Da) from the first peak, the two values were grouped together. If the difference was greater than the tolerance a new peak group was created. This process was completed for the entire combined peak list. The *m/z* value for the peak groups was calculated as an abundance weighted mean (AWM) using peak intensity. Ion intensity maps, based on signal to noise (S/N) values of peak group members, were then plotted if they existed in a minimum number of spectra (E.g. >50 spectra). Similarly for each of these masses the mass deviation of all peaks within those peak groups, from their AWM *m/z*, at each spatial coordinate, was plotted using a red (-0.1 Da), white (0 Da), green (+0.1 Da) color gradient, also known as a mass deviation map. The peak groups were also exported as ASCII text files containing x and y locations, *m/z* value, S/N, resolution, peak area and intensity of all peaks in that group.

### 2.5.9 Haematoxylin and Eosin Staining

Following MS data acquisition matrix was eluted from tissue sections using a five minute wash in 100% EtOH. Sections were stained with haematoxylin (Sigma-Aldrich, Steinheim, Germany) for five minutes, rinsed twice in tap water and dipped in 1% v/v HCl. Sections were rinsed once in tap water prior to staining with fresh eosin (Sigma-Aldrich, Steinheim, Germany) using a single dip. Excess stain was washed off using two rinses in tap water, one dip in 100% EtOH and two dips in 100% xylene. Cover slips were mounted using Pertex mounting medium (Medite Midizintechnick, Burgdorf, Germany) and left to set overnight. Microscopy scans were performed on a Nanozoomer automated microscope (Hamamatsu, Beijing, China). Tissue sections were annotated and graded by Dr. Fergus Whitehead (Adelaide Pathology, Mile End, South Australia).

### 2.5.10 K-means clustering

K-means clustering were performed by A/Prof. Inge Koch (School of Mathematics, The University of Adelaide, Australia) in collaboration with Prof. Steve Marron (Statistics and Operations Research, The University of North Carolina – Chapel Hill). A detailed description of the methods applied is provided elsewhere **[80]**. Subsequent processing and interpretation of clustering results is described in **chapter 7**.

### 2.5.11 Laser micro-dissection, tryptic digestion and purification

Tissue sections for laser capture micro-dissection (LCM) were water bath mounted at 39ºC onto PEN membrane slides (MicroDissect, Germany). Slides were heated at 60ºC for five min to fuse tissue to the PEN membrane, washed with 100% xylene for 90 s and 100% EtOH for 60 seconds. Tissue was extracted by LCM using a Leica AS LCM microscope (Leica Instruments, Germany) and the following

microscope settings. Aperture – 15; intensity – 46; speed – 5; bridge – large (6); offset – 20. Laser cut tissue was dropped into 30 µL of 10 mM citric acid (pH 6) and centrifuged briefly to move tissue to the bottom of the micro-centrifuge tube. An extra 170 µL citric acid was added and the tissue was heated at 95ºC for 45 min to mimic the CAAR step used for MALDI-IMS. Following completion, the tube was centrifuged again, excess solution was carefully removed using a pipette and 10 µL 25 mM $NH_4HCO_3$ was added to neutralize the citric acid. The $NH_4HCO_3$ was removed and a fresh 10 µL added followed by 10 µL 10 ng/µL modified porcine trypsin. The digest was incubated at 37ºC for two hours and stopped with one µL 10% TFA. Tryptic peptide digests were cleaned using C-18 centrifuge columns (Thermo-Fisher Scientific). All centrifugations were at 1500 x g for one minute each. For each spin column (one sample per column), the column was rinsed twice by centrifugation using 200 µL 100% ACN, twice with 200 µL 50% ACN + 0.5% TFA and three times with 200 µL 2% ACN + 0.5% TFA. The column was then rinsed by centrifugation using first 40 µL and then 100 µL 2% TFA. Peptide samples from the tryptic digest were loaded onto the C-18 column and centrifuged into a fresh 1.5 mL polypropylene micro-vial (Eppendorf, Hamburg, Germany). Flow through was centrifuged through the column another two times to ensure complete binding of peptides. The peptides were washed three times by centrifugation using 200 µL 2% ACN + 0.5% TFA. Peptides were eluted by centrifugation into a fresh 1.5 mL micro-vial using three volumes of 20 µL 70% ACN + 1% TFA. Each sample was reduced to 3-5 µL using a vacuum centrifuge and made up to ~10 µL using 2% ACN with 0.1% FA.

#### 2.5.12 HPLC-ESI-LTQ-Orbitrap MS and MS/MS

The nHPLC-ESI-LTQ-Orbitrap was maintained by Yin Ying Ho (Adelaide Proteomics Centre, The University of Adelaide, Australia). Analysis of all C-18 purified tryptic digests was performed using an online Ultimate 3000 nHPLC system (Dionex) and LTQ-Orbitrap XL instrument (Thermo Scientific) running binary gradients of 98% $H_2O$ with 2% ACN and 0.05% FA (A) and 80% ACN with 20% $H_2O$ and

0.1% FA (B). One µL of concentrated sample was injected onto a C-18 µ-pre-column (Acclaim Pepmap100) and loaded for 15 min at five µL/min flow rate using a water buffer. Analytical separation was achieved by bringing an Acclaim Pepmap100 C-18 column (2 $\mu$m, 100 Å, 75 $\mu$m I.D. $\times$ 15 cm, Dionex) into line with the pre-column at a flow rate of 0.3 µL/min and running a 140 minute 4-55% B gradient. The eluted peptides were ionized by ESI and introduced into the LTQ-Orbitrap XL where the top ten peptides in each scan were fragmented by CID. MS parameters were; minimum signal -500; isolation width – 3.00; normalized collision energy – 35.0; charge state – ≥+2. Data files were exported to Proteome Discoverer software (V1.1, Thermo Scientific). Data was searched using Mascot (V2.2). Search parameters were: Taxonomy – *Homo sapiens*; database – Swiss-Prot (2011_08); enzyme – Trypsin; variable modifications – oxidation (M); MS tolerance – 0.5 Da; MS/MS tolerance – 0.8 Da; missed cleavages – 3. CSV files were exported from Mascot using a significance threshold of 0.05. All hits in these CSV files with an ion score below the identity threshold were removed prior to matching of these hits to peak group lists generated from MALDI-IMS experiments. Matches were made within a ± 0.2 Da.

# Chapter 3 Manuscript Context

MALDI-IMS has been applied successfully by numerous groups since its novel application in 1997 **[25]**. The driver behind the increasing interest in MALDI-IMS has been the potential for molecular annotation of tissue, the best example of which comes from MALDI tissue profiling (see **chapter 1.8**). Using profiling based approaches molecular features detectable by MALDI-TOF MS have been used to distinguish sub-types of cancer **[62]**, grade cancers **[47]** and generate prognostic information for cancer patients **[54]**. Not only can a MALDI-IMS approach be used to generate molecular profiles, the technology can also provide molecular distributions of key features contributing to these profiles. In achieving this, tissue specific peptides or proteins can be used to map tissue architecture based on measurable MS features rather than morphology.

Unfortunately, the novelty of MALDI-IMS as a technology platform meant that no expertise was available locally. This prompted preliminary demonstration that imaging of proteins using MALDI-TOF MS was feasible with the instruments available for sample preparation and data acquisition. It was shown in that protein distributions could be measured across tissues at a centre to centre resolution of 70 μm using automated sample preparation. This proof-of-principle investigation was published in the Journal of Proteomics and Bioinformatics **[40]** and is presented in this thesis as **chapter 3**.

# Chapter 3 Imaging Mass Spectrometry and Its Methodological Application to Murine Tissue

Journal of Proteomics & Bioinformatics - Open Access

www.omicsonline.com Research Article JPB/Vol.1/December 2008

J.O.R. Gustafsson[1], S. R. McColl[2], P. Hoffmann[1]

1 Adelaide Proteomics Centre, School of Molecular and Biomedical Science, University of Adelaide, Adelaide, South Australia 5005.

2 Chemokine Laboratory, School of Molecular and Biomedical Science, University of Adelaide, Adelaide, South Australia 5005.

Corresponding author: Peter Hoffmann, Adelaide Proteomics Centre, Gate 8 Victoria Drive,

Molecular Life Sciences Building Rm 1.49, University of Adelaide, Adelaide, South Australia 5005. Ph: (08) 8303 7521.

Received September 09, 2008; Accepted November 02, 2008; Published December 05, 2008

Citation: Gustafsson JOR, McColl SR, Hoffmann P (2008) Imaging Mass Spectrometry and Its Methodological Application to Murine Tissue. J Proteomics Bioinform 1: 458-463.

Copyright: © 2008 Gustafsson JOR, etal. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Key words: MALDI; mass spectrometry; imaging; tissue; methodology

Abbreviations: IMS, Imaging Mass Spectrometry

# Statement of authorship for chapter 3

**Ove Johan Ragnar Gustafsson (Candidate)**

Statement of contribution (in terms of the conceptualization of the work, its realization and its documentation)

Performed analysis on all samples presented, interpreted data and wrote the manuscript

Certification that the statement of contribution is accurate

Signed.....Date.....

**Shaun R. McColl (co-author)**

Statement of contribution (in terms of the conceptualization of the work, its realization and its documentation)

Manuscript evaluation

Certification that the statement of contribution is accurate and permission is given for the inclusion of the paper in the thesis

Signed.....Date.....

**Peter Hoffmann (co-author)**

Statement of contribution (in terms of the conceptualization of the work, its realization and its documentation)

Supervised development of work, helped in data interpretation and manuscript evaluation and acted as corresponding author

Certification that the statement of contribution is accurate and permission is given for the inclusion of the paper in the thesis

Signed.....Date.. .....

## 3.1 Abstract

Imaging mass spectrometry (IMS) is an emerging technology, pioneered by Prof. Richard Caprioli's group starting more than a decade ago. In this study we have demonstrated the simplicity of initial technological set up for IMS experiments with commercially available automated matrix deposition, MALDI-TOF mass spectrometry instrumentation and data handling software for image generation. We have applied two different concepts of automated matrix deposition on murine brain sections and discussed their different features and capabilities in IMS.

## 3.2 Introduction

Proteomics has seen the development of evermore complex and labour intensive methodologies and technologies, with increased qualitative and quantitative analytical power **[22, 81, 82]**. As the workhorses of proteomics, two-dimensional gels and high performance liquid chromatography (HPLC) allow exquisite protein/peptide separation and thus in depth proteome analysis in high throughput formats **[82]**. Such classical proteomics, however, require protein/peptide sample solubilization and treatment (E.g. precipitation, purification, labeling) prior to separation and downstream analysis by mass spectrometry [MS] **[81, 82]**. Furthermore, studies seeking diagnostic markers typically use bodily fluids as the basis for investigation, raising the possibility of multiple high abundance species complicating analysis. Spatial information is also lost when solubilization and protein separation are applied to tissue samples. In this light, it is not surprising that almost a decade ago direct MS analysis on tissue sections was pioneered by Caprioli *et al.* **[25, 44]**. MS is now regularly being applied to in situ tissue analysis **[42, 43, 75]**. Typically, tissue sections are washed in alcohol, dried and either a homogeneous layer or

droplet array of matrix is applied **[43]**. Direct MS desorption from the matrix co-crystallized tissue surface is then performed: predominantly with Matrix Assisted Laser Desorption/Ionization (MALDI) sources **[42, 75]**. Prior knowledge of tissue composition is not necessary, hundreds if not thousands of masses can be measured rapidly and spatial distribution information is maintained **[44]**. To date, IMS has been applied in several biological systems including rodent **[43, 63, 83]** and crustacean brain **[84]**, rodent spinal cord **[85]**, liver **[42]**, testes, kidney **[86]** and skin **[87]**. Due to its relative infancy, IMS still requires optimization of many of its component protocols, including those for section washing, effective and reproducible matrix deposition and processing the large volumes of data generated. Thus, although great progress has been made by several groups no standard IMS methodology has been defined.

With this is mind it was the objective to develop an IMS platform in our laboratory, focused on the imaging aspects of an IMS experiment prior to profiling applications for statistical model generation and potential diagnostic applications **[28, 75]**. One of the most challenging aspects of IMS is the reproducible application of matrix on tissue sections. Three different commercial systems are currently available for automated matrix deposition, including the ImagePrep station (Bruker Daltonics), the ChIP-1000 printer (Shimadzu Biotechnology) and the Portrait 630 printer (Labcyte). The ImagePrep station uses vibrational vaporization to generate a matrix aerosol, which gently settles onto tissue samples. The ChIP-1000 and Portrait 630 are both picolitre range droplet printers capable of generating discrete matrix arrays on tissue sections. However, while the ChIP-1000 uses piezoelectric technology to forcefully eject droplets from a printer head, the Portrait 630 uses acoustic ejection of droplets from a liquid surface. We were able to successfully apply IMS to murine brain sections using both the ChIP-1000 and the ImagePrep in conjunction with a MALDI-TOF-MS instrument and image generation software. Under these conditions a comparison between the two instruments using similar sagittal murine brain sections is provided in this communication. Continued optimization of key methodological steps is still needed, including maintenance of tissue integrity, section washes, matrix solvent composition and matrix deposition strategy. However, we have shown that the technology can be

implemented with commercially available instrumentation in a relatively short timeframe. The potential of this technology for future application in biomarker discovery and diagnostics looks extremely promising.

## 3.3 Results and Discussion

### 3.3.1 Stability of Imaging Mass Spectrometry Instrumentation Allows Successful Application

The IMS methodology is rather straightforward in that it involves use of tissue sections mounted on a conductive surface, which are washed in alcohol to fix protein and wash away contaminants, coated in matrix and analyzed by MALDI-TOF-MS. Established tissue washes were used for our applications, including simple washes in 70% EtOH or IPA and 100% EtOH or IPA **[42, 43]**. The difference in terms of EtOH and IPA is marginal, but as demonstrated previously, IPA generates slightly better spectral data **[42]**. IMS requires automated matrix deposition: stability at this stage is critical to ensure reproducibility and prevent protein/peptide delocalization due to matrix pooling on the section. The two automated systems, the Shimadzu ChIP-1000 Piezoelectric printer and the Bruker Daltonics ImagePrep station, were chosen for their availability and demonstration of applicability, both in the literature **[63]** and in practical terms. The ChIP-1000 has been utilized for experiments on Murine brain, as shown in **Figure 3.1**. The SA matrix crystals in **Figure 3.1A** were arranged in a 250 µm centre to centre array coating the coronal Murine brain section and these array foci appeared as discrete, dark spots on the section surface. Printer stability was paramount, as instability leads to dispensing of droplets in multiple directions, interrupting the grid pattern, but more importantly, potentially causing random delocalization of proteins and peptides. However, with fine modulation of piezoelectric voltage during printing, large droplet arrays, as shown in **Figure 3.1A**, were consistently deposited. Proper printer function was also found to be dependent on the stability of pressure lines feeding into the piezo unit and the

condition/cleanliness of the external portions of the print head. These operations were also considered during fine tuning of the printing process and add to the complexity of the instrument. Furthermore, the matrix itself can generate problems in that SA, for example, can cause significant blockages of the piezo printing orifice at concentrations above 10 mg/ml. The ImagePrep station, which generates homogeneous matrix layers, maintains stability dependent on instrument specific methods, the condition of the piezo controlled nebulizer membrane as well as matrix concentration. In a trial of the instrument it was found that while 6 mg/ml SA can generate a homogeneous field of discrete matrix crystals, 10 mg/ml SA in an identical solvent mixture generates larger matrix foci, decreasing the possible resolution of the homogeneous spray preparation (**Figure 3.1B** and **3.1C** respectively). Matrix blockages caused by crystallization of SA on the nebulizer membrane do not have dramatic effects on spray quality until a significant portion (>50%) of the membrane surface is covered. Resolution, while limited in these examples to arrays of 250 µm or more using a ChIP-1000 printer (**Figure 3.1A, 3.2A-D**), was reduced to 70 µm with the ImagePrep automated spray system (**Figure 3.2E**). Pending further experiments, however, it is unclear how comparable the two systems are in terms of reproducibility.

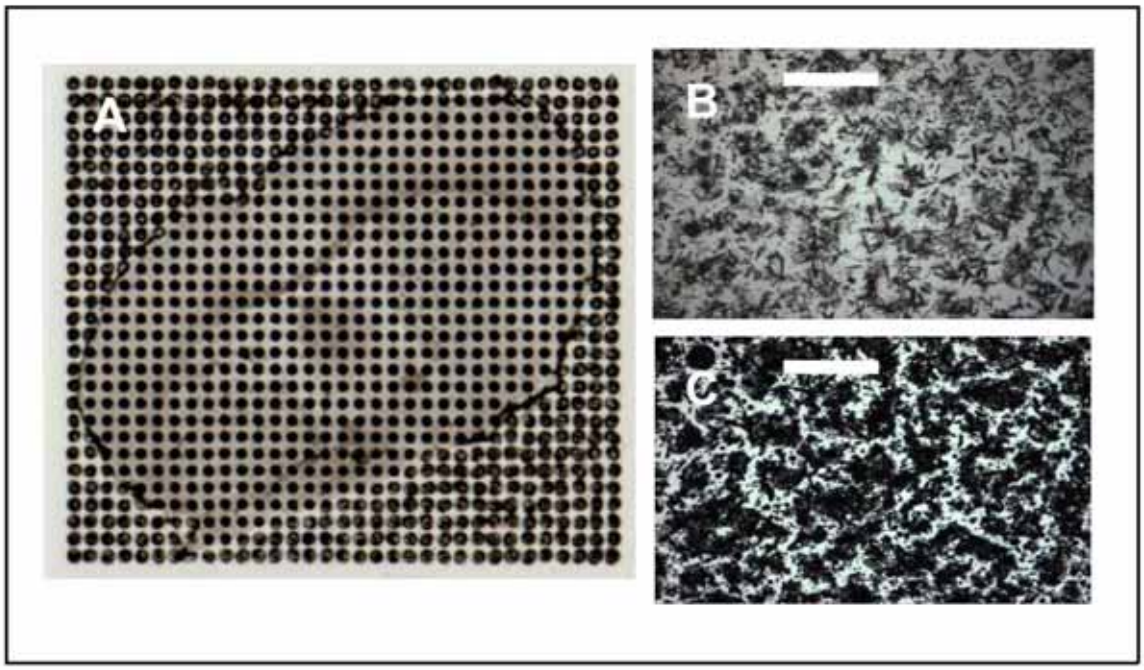



**Figure 3.1.** Matrix deposition strategies and resultant crystal formations The two dominant methods of matrix deposition are droplet deposition in a fixed array (**A**) and spray deposition to generate a homogeneous crystal coating (**B** and **C**). The matrix array shown in **A** was generated by a CHIP-1000 piezo electric printer (Shimadzu) with a centre to centre distance of 250 µm using 10 mg/mL sinapinic acid (SA) in 50% acetonitrile (ACN), 44.8% ultra pure water, 5% isopropanol (IPA) and 0.2% trifluoroacetic acid (TFA). One droplet was deposited at each position for each of 25 layers. Droplet arrayed tissue section was washed in 70% and 100% IPA prior to matrix deposition. The homogeneous matrix layers generated in **B** and **C** were deposited by an ImagePrep station (Bruker Daltonics) using 6 mg/mL (**B**) or 10 mg/mL (**C**) SA with identical solvent compositions to those used for the CHIP-1000. The scale bars in **B** and **C** are 30 µm. ImagePrep tissues were washed in 70% and 100% ethanol prior to matrix deposition.

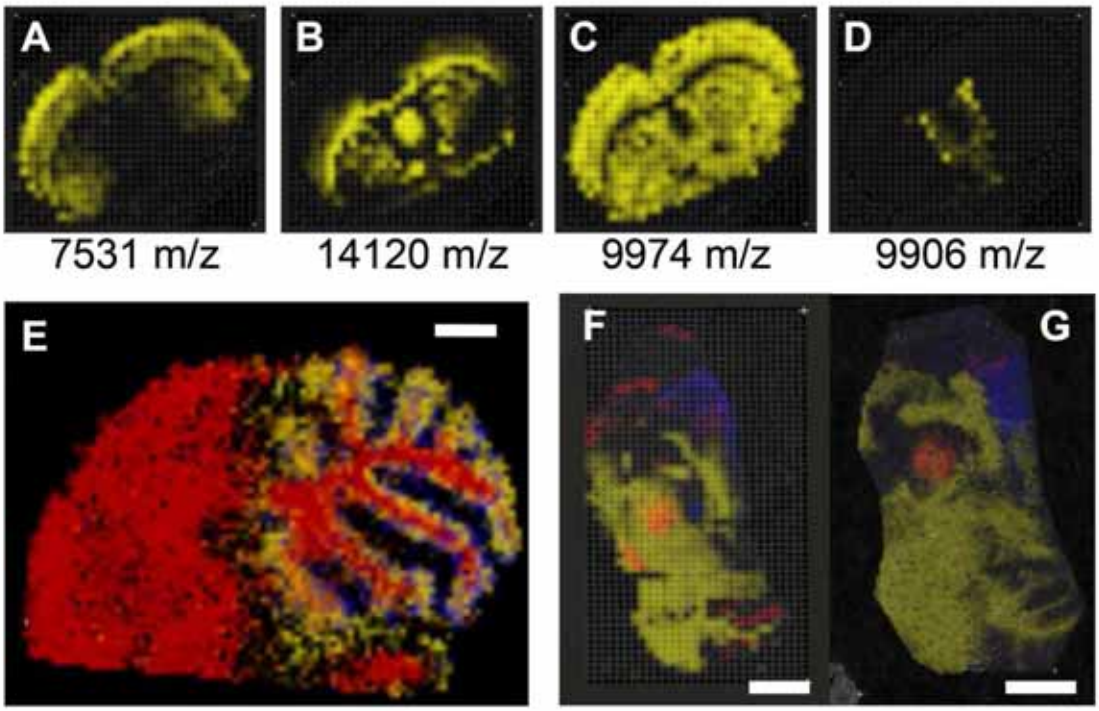



**Figure 3.2.** Imaging the Murine brain utilizing automated matrix deposition Figures **A-D** show single ion intensity map examples from an imaging mass spectrometry experiment where 10 mg/mL sinapinic acid (SA) in 50% acetonitrile (ACN), 44.8% ultra pure water, 5% isopropanol (IPA) and 0.2% trifluoroacetic acid (TFA) was deposited in 1 droplet iterations per layer in 25 layers onto a 10 µm thick section of coronal murine brain tissue using the CHIP-1000 (all experiments below used identical matrix solvent compositions). Array was 250 µm centre to centre and can be seen overlaid onto figures **A-D**. Tissues in **A-D** were washed in 70% and 100% IPA prior to matrix deposition. Figure **E** shows a combination of three separate ion intensity maps where intensity of the individual colour correlates to relative ion intensity at that location. A 10 µm thick sagittal Murine brain section was coated with 10 mg/mL SA using an ImagePrep station. The instrument was operated with default SA settings and manual determination of spray power. Scale bar in **E** is 1 mm. Figure **F** shows three combined ion intensity maps from an experiment where 10 mg/mL SA was deposited in 24 droplet layers onto a 10 µm thick sagittal Murine brain section. Figure **G** shows a combination of the same three separate ion intensity maps at a resolution of 100 µm. This separate 10 µm thick sagittal Murine brain section was coated with 10 mg/mL SA in identical solvents to those above using an ImagePrep station. The instrument was operated with default SA settings and manual determination of spray power. Scale bars for **F** and **G** are 2 mm. An ultraflex III MALDI-TOF/TOF mass spectrometer (Bruker Daltonics) was used for acquisition and analysis was performed in flexImaging software (version 2.0, Bruker Daltonics).

### 3.3.2 Imaging the Murine Brain at the Protein Level with the CHIP-1000 and ImagePrep

Rodent and in particular the Murine brain, was a logical start point for IMS experiments due to its well characterized symmetrical structure and ease of handling once frozen. From a single acquisition point in an IMS experiment anywhere from 100-400 masses can be observed with typical array spectra showing high signal to noise (S/N) values for numerous masses in the range of 5-30 kDa **[44]**. And as shown in **Figures 3.2A-D**, masses analyzed correlate to internal section structure. Similar results are generated for homogeneous matrix layers, but these are generally more difficult to reproduce because control over droplet deposition is lost (data not shown). With proper modulation of instrument settings for both the CHIP-1000 and ImagePrep station, ion intensity maps resulting from multiple successful MALDI-TOF tissue acquisitions were generated as shown in **Figures 3.2F** and **3.2G** respectively. These results are key proofs of principle: currently these experiments show potential biological significance, which can be verified with future experimental determination of mass/protein identity by LC-MS/MS. Importantly, it is worth considering that the CHIP-1000 and ImagePrep, as a result of generating different crystal structures, become well suited to different modes of IMS. Droplet arrays are suited to low resolution protein profiling: where different tissue areas are rapidly scanned for identification of a tissue signature based upon molecular ion determinants **[46]**. Homogeneous matrix layers on the other hand are suited to pure imaging experiments where high resolution tracking of a defined ion or initial tissue overview is needed.

**Figure 3.2F** shows the ion maps corresponding to three different m/z-s and their spatial distribution across the tissue following acquisition from a printed droplet array. Compared to the ImagePrep data in **Figure 3.2G**, the low resolution seems to be a drawback of the CHIP-1000 matrix array. While high resolution is obtainable with the ImagePrep with a hands free automated system, the CHIP-1000 allows rapid deposition of multiple droplets in an array or single point format. Combined with histology, the printer systems are extremely well suited to rapid tissue profiling to provide molecular identification which complements classical pathology. Furthermore, the MS acquisition time for droplet arrays is far

less than the high resolution spray preparations. A typical array coating a coronal Murine brain section at a raster of 250 µm (**Figure 3.1A**) takes approximately one hour to analyze by MALDI-TOFMS, where as a similar experiment using a 70 µm raster or lower can take in excess of four hours. This limits high resolution IMS to the research laboratory until such a time when MALDI laser frequencies have increased by at least two fold.

## 3.4 Conclusion

IMS technology, although in its infancy compared to established proteomic methods, including HPLC and two dimensional gels, has shown and continues to show its extreme potential through numerous applications in biological systems. Our laboratory has set up and begun the process of evaluating the methodology for performing high quality IMS experiments using automated systems such as the CHIP-1000 printer and ImagePrep station for matrix deposition. As a result we have observed that both systems can be confidently applied to Murine brain tissue for the purpose of imaging, at both rapid medium resolution (CHIP-1000) and high resolution (ImagePrep).

## 3.5 Acknowledgement

The authors would like to acknowledge Prof. Deon Venter at the Mater Hospital in Brisbane and Bruker Daltonics for use of the ImagePrep station. We would also like to thank Prof. Mark Baker at the Australian Proteome Analysis Facility (APAF) for arranging a loan of their CHIP-1000 instrument and Shimadzu Biotechnology for their application support with the instrument.

## The need for further MALDI-IMS method development

**Chapter 3** outlined the proof-or-principle application of MALDI-IMS for our laboratory. However, aim one of this thesis project was to develop MALDI-IMS into a platform for reproducibly tracking protein ion changes across sections derived from frozen tissue. Thus, the key to completing this aim was further development of IMS methods to ensure that the quality of sample preparation, the spectra acquired (i.e. S/N), and the ion intensity maps produced downstream, were reproducible.

In order to ensure reproducibility, a better understanding of the individual MALDI-IMS sample preparation steps was necessary. The most important of these steps are chemical section treatment and automated sample preparation (e.g. matrix deposition). **Chapter 4** therefore presents an in-depth discussion of the optimization required to perform reproducible MALDI-IMS on sections obtained from frozen tissue.

# Chapter 4 Development and application of automated sample preparation methods for MALDI imaging mass spectrometry

## 4.1 Introduction

The majority of MALDI-IMS projects use sections from frozen tissue as there are no impediments to direct extraction and analysis of endogenous peptide and/or protein analytes, either by MALDI-IMS or classical proteomics technologies **[51, 52, 86-91]**.

**Chapter 3** outlined the preliminary application of MALDI-IMS to murine brain, demonstrating the ease of use of frozen tissue and the potential of IMS technology for molecular annotation of tissue sections. In order to apply MALDI-IMS to clinically relevant research, the methods presented in chapter three required further development to ensure that any acquired data was;

1. High quality (in terms of signal to noise ratio (S/N, biological relevance) and,
2. Reproducible (for MS data and ion intensity maps)

Of the MALDI-IMS sample preparation steps described in chapter one (see **Figure 1.2**) and three, tissue fixation and automated sample preparation (trypsin and/or matrix deposition) have the greatest effect on the reproducibility of spectral signal to noise (S/N) and the biological relevance of acquired data. As a result, the experiments described in this chapter were used to evaluate and optimize these two sample preparation steps to meet the requirements outlined above.

## 4.2 Optimizing tissue section fixation to minimize analyte delocalization

It has been demonstrated in the literature **[41-43]** and independently by our group (**Figure 4.1**) that the use of organic fixatives to treat tissue sections prior to sample preparation improves the S/N and spectral complexity of *in situ* MS data acquisition.

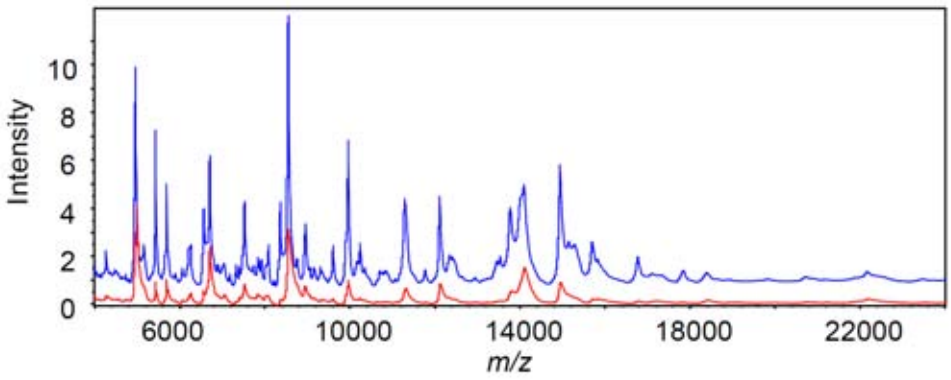



**Figure 4.1.** Sum spectra across two 8 µm coronal murine brain sections, thaw mounted onto indium tin oxide (ITO) slides, desiccated for 30 min and either left untreated (red spectrum) or washed with 70% and 100% ethanol for 60 s each (blue spectrum). One nanolitre total of 10 mg/mL sinapinic acid (in 50% acetonitrile and 0.2% trifluoroacetic acid) was deposited using a ChIP-1000 printer (Shimadzu Biotechnology, Japan) in 250 µm arrays onto both the sections. Data was acquired on an ultraflex III MALDI-TOF/TOF system (Bruker Daltonics) operating in linear positive ion mode. Sum spectra were generated using flexImaging (V2.1, Bruker Daltonics).

As described previously, the positive effect of organic solvent treatment is a likely result of the removal of physiological salts and lipids from the tissue, allowing improved matrix crystallization and therefore increased MS sensitivity **[42, 43]**. However, peptide and protein analytes are also extracted, particularly as the typical section thicknesses used (6-10 µm) will slice a proportion of cells in half, releasing proteins from intra-cellular compartments. Depending on the analyte of interest, duration and number of washes are thus varied to account for analyte extraction. For example, a study analyzing endogenous peptides should minimize washes to one or two short (5-10 second) rinses, while a protein study might wash several times in several solvent types of increasing hydrophobicity **[92]**.

Extraction of analytes becomes a significant problem if the delocalized analytes remain on the tissue or the ITO slide because the biological relevance of any MALDI-IMS data is either reduced or lost depending on the extent of the effect. Because of its potential impact on MALDI-IMS experiments, delocalization was characterized further using isopropanol (IPA) treatment of consecutive sections of a murine brain. IPA was chosen as it has been shown previously to be the best solvent for protein level MALDI-IMS, when compared to other solvents **[42]**.

In total, four conditions (in duplicate), using 50 mL volumes of 70% IPA and 100% IPA, were selected for eight 10 μm thick consecutive coronal brain sections (**i-viii**). The conditions were (as indicated in **Figure 4.2**);

1. Horizontal in a large container (14.5 x 8.5 cm) for 60 seconds each (**i-ii**),
2. Vertical in a polypropylene tube for 60 seconds each (**iii-iv**)
3. Horizontal in a small petri dish (9 cm diameter) for 60 seconds (**v-vi**),
4. Horizontal in a petri dish (9 cm diameter) for 10 seconds (**vii-viii**)

In each case the slides were washed and immediately tipped vertically to rapidly drain solvent. Once the sections were dry, 12 nL 20 mg/mL 2,5-DHB (in 50% MeOH and 0.2% TFA) was printed in an ordered array onto each section using a ChIP-1000 (Shimadzu Biotechnology). Printing removed the possibility of delocalization due to pooling of matrix on the slide, as can occur with spray methods. In addition, using 2,5-DHB (peptide matrix) rather than SA (protein matrix) ensured experimental reproducibility given that 2,5-DHB prints remain stable for hours at a time on a ChIP-1000 instrument where as SA needs to be constantly monitored **[93]**. MS acquisition was performed in linear positive ion mode and the resultant data combined into ion intensity maps for seven example *m/z* values using flexImaging (V2.1, Bruker Daltonics, Bremen, Germany). The maps are shown as a grid in **Figure 4.2** for each of the four conditions. These data indicated that the horizontal wash in the larger container (**i-ii**) caused extensive delocalization for masses below *m/z* 5000. In fact, the major detectable ion in these experiments, *m/z* 4964, consistently delocalized, the extent of which was variable. Conversely *m/z* 6723

and 14124 consistently maintained their position regardless of wash type. Importantly, the majority of measured ions, including *m/z* 4964, remained on the sections washed vertically in a polypropylene tube (**iii-iv**) and horizontally in the petri dish (**v-viii**). It was hypothesized that because the solvent layer was shallower in the large container (**Figure 4.2 i-ii**) complete removal of extracted peptides/proteins was not possible. The analytes could therefore resettle on the slide as it dried. The vertical tube (**Figure 4.2 iii-iv**) and small petri dish (**Figure 4.2 v-viii**) therefore minimized delocalization by providing a sufficient volume of solvent to remove most extracted material. Recently published wash protocols supported this contention as they included six wash steps in solvents of variable composition **[92]**.

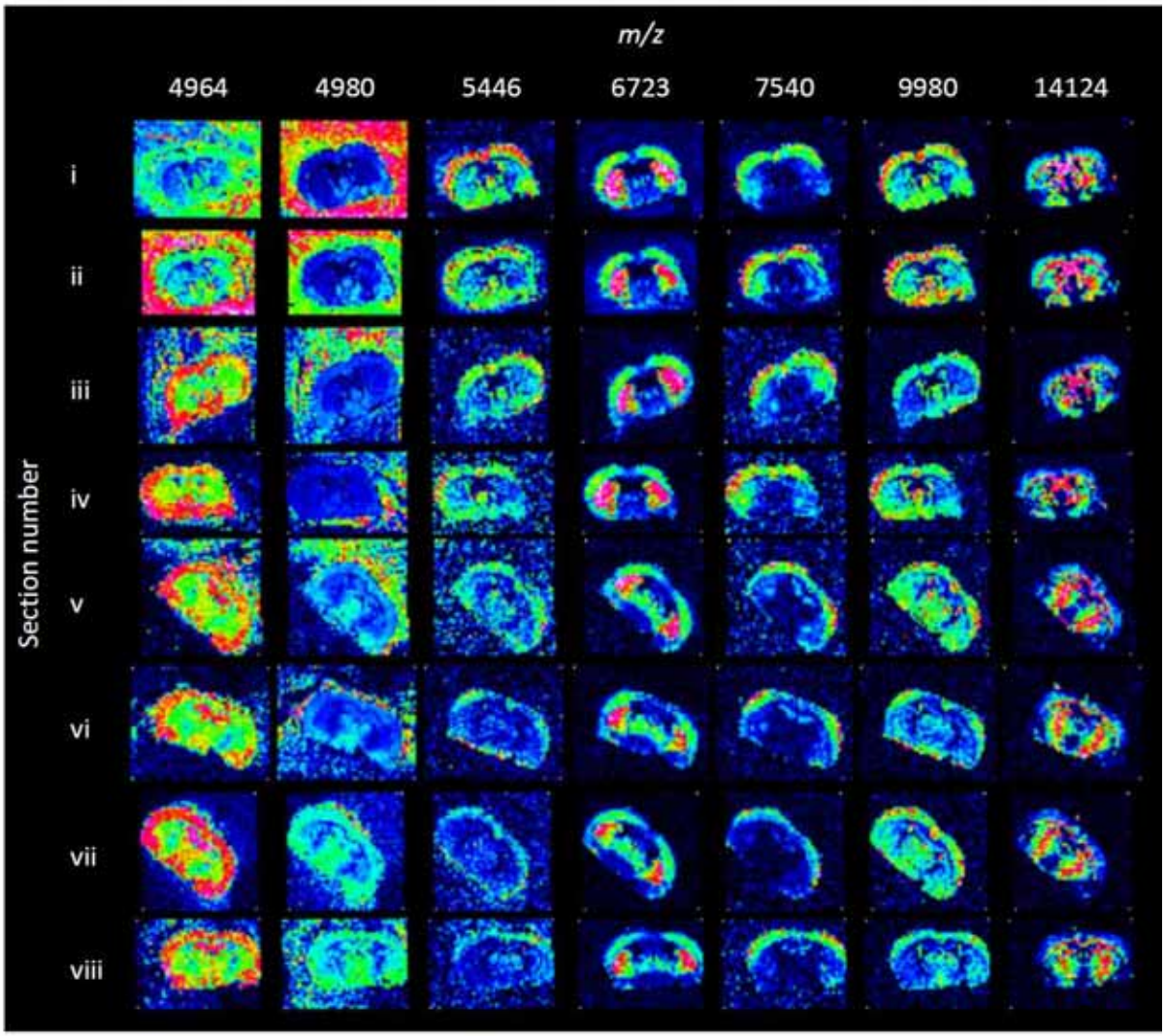



**Figure 4.2.** Eight consecutive 10 µm coronal murine brain sections were washed with 70% and 100% IPA. Wash conditions: (**i/ii**) horizontally in large container (60 s each), (**iii/iv**) vertically in a 50 mL tube (60 s each), horizontally in a petri dish for 60 s (**v/vi**) and 10 s (**vii/viii**). Solvent was tipped off and sections dried vertically. Twelve nL of 20 mg/mL 2,5-DHB (in 50% MeOH and 0.2% TFA) was printed using a ChIP-1000 (250 µm centre to centre array). MS data was acquired on an ultraflex III MALDI-TOF/TOF MS instrument operating in linear positive ion mode. Ion maps were generated in flexImaging (V2.1) from normalized data.

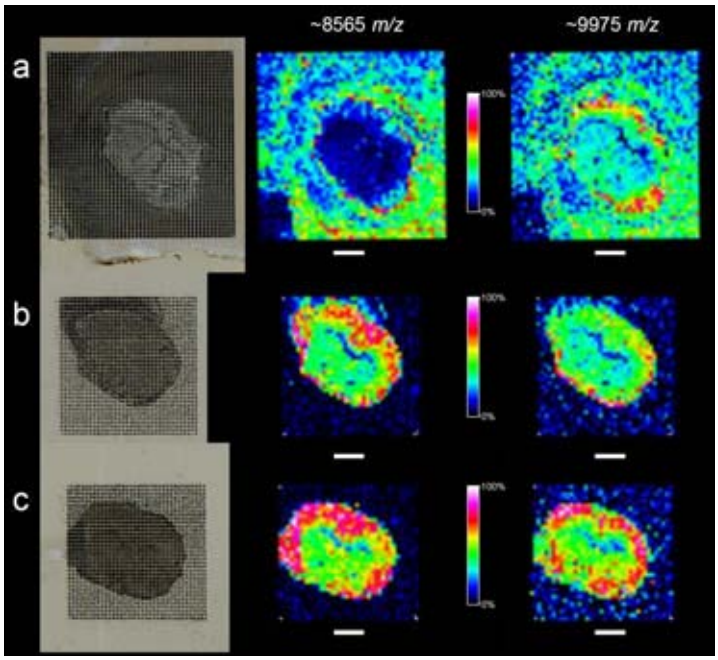



**Figure 4.3.** Consecutive 10 µm thick coronal murine brain sections were either press mounted (**a** and **c**) or lift mounted (**b**) onto indium tin oxide (ITO) slides. All sections were washed horizontally in a small petri dish (9 cm diameter) with 50 mL 70% isopropanol (with 0.1% trifluoroacetic acid) and 50 mL 100% isopropanol for 10 s each. The section in **a** was dried horizontally following wash, **b** and **c** were tipped vertically following wash. Twelve nanolitres 20 mg/mL 2,5-DHB (in 50% methanol and 0.2% trifluoroacetic acid) was printed onto the sections using a ChIP-1000 (Shimadzu Biotechnology). MS was performed on an ultraflex III MALDI-TOF/TOF instrument in linear positive ion mode. Ion intensity maps show m/z 8565 (left) and 9975 (right). Intensity scales are included. Scale bars are 2 mm.

To confirm re-settling of analytes on the slide, horizontal washes (70% IPA + 0.1% TFA and 100% IPA for 10 s each) in the small glass dish were repeated for another set of three slides containing coronal murine brain sections (10 µm thick sections). Following washes, the first slide was lifted horizontally out of the petri dish, preventing the solvent from draining, and allowed to dry at room temperature. **Figure 4.3a** shows a scan of the washed section once dried and prepared for MS acquisition by printing of 2,5-DHB matrix (identical volume and composition to that used above). A large amount of visible tissue-derived material surrounded all three sections on this slide (see bottom edge of scan in **4.3a**). MALDI-IMS analysis indicated that extensive analyte delocalization occurred, as shown by two example ion intensity maps for *m/z* 8565 and 9975 in **Figure 4.3a**.

It was also considered that if sections were dragged across the slide during mounting, tissue material could be smeared across the slide surface. This hypothesis was tested using the remaining two slides by either press mounting (pressing slide to the tissue section as it lies flat on the cryotome blade) or lift mounting (section was lifted onto slide and moved into position using a brush prior to thaw mounting) the brain sections. Matrix deposition and MS data acquisition was identical to the section in **Figure 4.3a**. Despite the use of different section mounting methods, there was no observable mass delocalization for the sections washed for 10 s, tipped vertically and dried immediately (**Figure 4.3b** and **c**), suggesting that diffusion was independent of mounting method. Analyte delocalization was therefore most likely due to resettling of peptides and proteins following a wash. The easiest way to reproducibly remove as much delocalized material as possible was by using vertical washing, where slide removal lifts the sections up and away from the wash solution.

These data led to the implementation of a default vertical wash orientation in 50 mL of each solvent with 60 s wash duration. Typically, sections were treated in this manner using two volumes of 70% IPA and a single volume of 100% IPA to ensure removal of the extracted material. The second 70% IPA wash noticeably reduced the occurrence of tissue shock, where the section dried excessively and detached from the slide. These shock effects could also be reduced by manual poly-L lysine-coating prior to mounting. Furthermore, in those cases where sections were lipid rich (e.g. adipose tissue), a second 100% IPA wash was required to remove macroscopically visible lipid contamination from the slides. While chloroform could also be used it was typically avoided because sections regularly experienced tissue shock (fragmented or generated poor MS data). Furthermore, chloroform can not remove even slight contamination with optimal cutting temperature (OCT) compound, which was used to mount the base of tissue blocks in the cryotome. As a result, the polyethylene glycol (PEG) component of the OCT remained on the slide and introduced highly intensive polymer peak trains with a monomer mass difference of 44 Da (see Supplementary Figure 4.1).

The data presented in this section was acquired following automated matrix deposition using a ChIP-1000 (Shimadzu Biotechnology, Japan). The application of automated sample preparation prior to MS

increased the reproducibility and quality of solution deposition (volume and position) as well as the quality of obtained data, when compared to manual thin layer chromatography sprayers used previously **[94]**. As such, the following section describes the application and method development for the instruments available for this thesis work; the ImagePrep station (Bruker Daltonics) and ChIP-1000.

## 4.3 Application of automated sample preparation instruments

### 4.3.1 Application of the ImagePrep station for protein MALDI-IMS

The ImagePrep station aerosolizes solutions (e.g. trypsin/matrix) by piezo induced vibration of the spray generator; a porous metal membrane, which blocks the open end of an inverted solution reservoir. The vibration vaporizes the reservoir solution coating the spray generator surface. As a result, a fine mist is generated which settles onto an ITO microscopy slide placed on a raised platform in the middle of the instrument. The duration and strength of the spray is controlled by a light sensor in the middle of the raised platform, which measures matrix coverage by the level of apparent light scatter (i.e. voltage output of sensor) induced by the formation of matrix crystals on the slide. Importantly, this light sensor takes measurements from an empty portion of the slide to ensure measurement of light scatter through a homogeneous matrix layer **[95]**.

During a single light sensor controlled spray cycle, the instrument first deposits a layer of matrix solution, which causes the matrix crystals already on the slide to partially re-dissolve and reduce light scatter from an initial plateau level. As the new layer crystallizes the greater density of matrix crystals causes more light scatter and as a result the sensor voltage plateaus at a higher level than the initial plateau. This is referred to as a standard light scatter curve (SLSC).

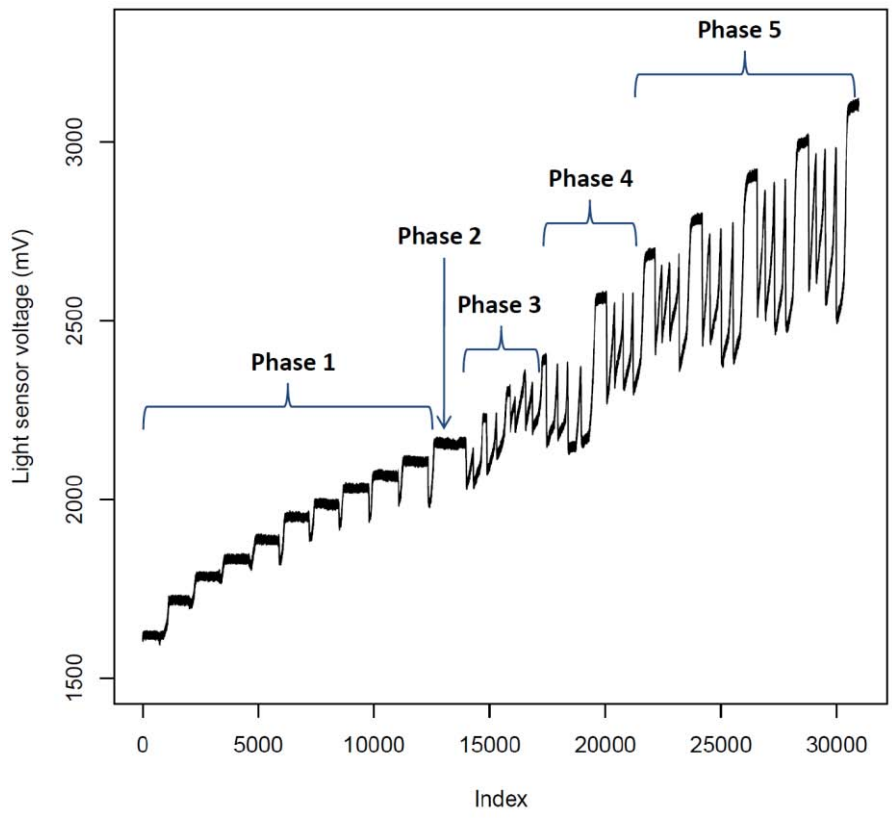



**Figure 4.4.** Zoom view of an entire ImagePrep deposition encompassing five phases. The individual drops in light sensor voltage were caused by matrix solution being sprayed onto a dry matrix layer, which was partially re-dissolved by the new matrix. Subsequently light scatter decreased from the initial "dry" plateau sensor voltage. The matrix then dried completely and a final plateau was reached where sensor voltage and therefore matrix crystal density was higher. Light scatter curve data was extracted as a text file from the ImagePrep station and plotted using R software **[76]**.

**Figure 4.4** shows the complete light scatter curve for all five phases of a typical SA preparation. Each phase of the preparation contained several spray cycles. Phase one was always independent of the light sensor to correct for inversions of the sensor signal, which occurred because matrix droplets sprayed onto a clean slide always generated more light scatter than the empty slide. This effect typically caused a signal spike, which, if not removed by the end of phase one, interfered with subsequent depositions. In **Figure 4.4**, a SLSC was established in phase one of the SA preparation and generated a matrix layer which could be partially re-dissolved and reduce light scatter. Phases three to five of the preparation then relied on the SLSC, because these phases applied matrix layers immediately before the previously applied layer dried. To do this in a reproducible way requires instrument controlled measurement of the SLSC and estimation of the appropriate time to deposit the next layer. This was

only possible when the SLSC was established, allowing extraction of peptides and protein analytes to be maximized while preventing pooling of matrix on the tissue section, which results in analyte delocalization.

**Figure 4.5a** further demonstrates the importance of establishing a SLSC. The data plotted for this SA deposition showed that although light scatter increased incrementally in the first phase, there was no consistent, significant drop in light scatter followed by drying and plateau formation. Subsequently, when phase three was reached, automated increases in spray power occurred to force a drop in light scatter and matrix solution began to pool on the slide. Over the course of several depositions, the failure to generate a SLSC was observed to be random and would thus only occur for a fraction of all sample preparations. It was thus hypothesized that although the washed ITO slide appeared macroscopically clean, trace contaminants including lipids were still present on the slides surface. These contaminants would impede the controlled matrix crystallization required for reproducible generation of a SLSC. Two method modifications were used to address this issue. Firstly, a clean cover slip was always placed on the slide directly over the light sensor. In addition, the minimum number of cycles in phase one was increased from four to eight. These two modifications were observed to be beneficial to establishing a SLSC. **Figure 4.5b** shows two stable preparations which utilized these modifications. The first was a typical SA deposition (blue curve). The second preparation repeated phase one prior to continuing the preparation as normal (red curve). Both preparations were able to establish a SLSC in phase one, with the repeated phase one preparation exhibiting a higher sensor voltage due to a larger total volume of matrix being deposited.

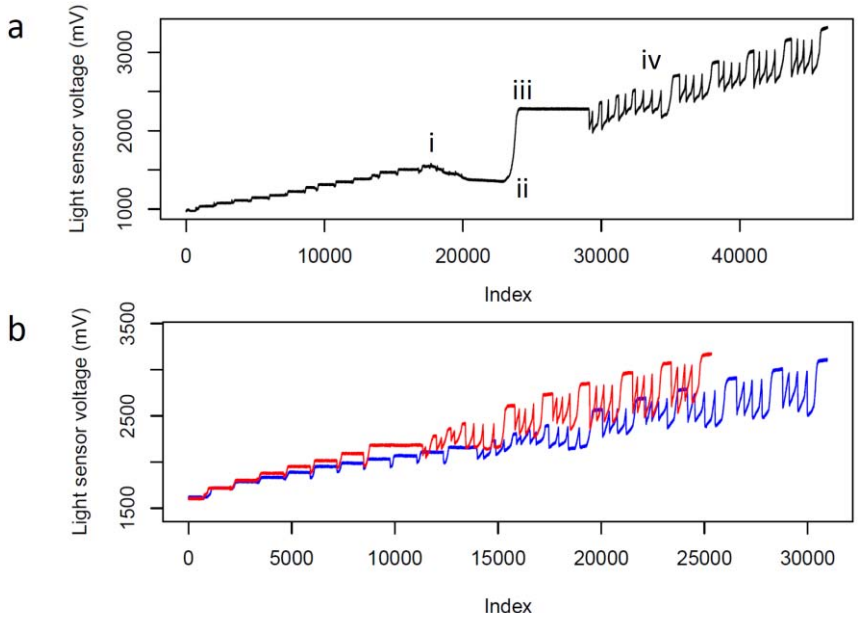



**Figure 4.5.** Light scatter curves are presented for three depositions of 10 mg/mL sinapinic acid (in 50% acetonitrile, 5% isopropanol and 0.2% trifluoroacetic acid) on slides containing murine brain sections. Standard sinapinic acid ImagePrep method was used to generate the curve shown in panel **a**. Points i-iv show an increase in spray power, a large drop in light scatter, a dramatically higher final plateau and normal remainder of the preparation, respectively. Panel **b** shows two independent preparations using a glass cover slip over the light sensor. One preparation used the standard method (blue) and the other repeated phase one before continuing as normal (red). Light scatter curve data was extracted as a text file from the ImagePrep station and plotted using R software **[76]**.

**Figure 4.6** demonstrates the results of the modified stable 10 mg/mL SA preparation when applied to a section of coronal murine brain (section scan in **Figure 4.6a**). A standard SA ImagePrep method (full details in methods section, see **Table 2.1**) was modified to incorporate a repeat of phase one, with a minimum number of cycles set to eight. MS data from the subsequent analysis was used to generate an ion intensity map for *m/z* 14123.9, using a rainbow (**Figure 4.6b**) and green colour scale (**Figure 4.6c**). Ion *m/z* 14123.9 was selected as it has been identified in a previous MALDI-IMS study on rat brain as myelin basic protein (MBP) **[63]**. Importantly, the distribution of ion *m/z* 14123.9 was typical of myelin, as supported by **Figure 4.6e** [2-D high resolution mouse (C57B) brain atlas **[96]**], which shows an overlay with a myelin stained section from the same brain region as **Figure 4.6a**. It was thus established that application of stable matrix spray deposition allowed generation of biologically relevant protein distributions.

Unfortunately, it was subsequently discovered that increasing MS sensitivity (i.e. depositing more matrix solution) while minimizing analyte delocalization (i.e. depositing too much matrix solution) was difficult to achieve with the ImagePrep. It was hypothesized that the small average droplet size generated on tissue by the ImagePrep did not efficiently extract protein for incorporation into the matrix at all locations on a tissue section. Thus, only peptides and small proteins (< 10 kDa) would ionize efficiently using MALDI. This readily explained the acquisition of spectra where very few protein ions could be detected above the noise level. Importantly, the variability in protein detection across tissue sections reduced the number of biologically meaningful ion intensity maps which could be generated for a typical experiment. In many cases only the top 10-20 major peaks in a sum spectrum were useable for section annotation. However, the droplets printed by the ChIP-1000 were larger than the droplets generated by the ImagePrep spray, as evidenced by the large area of tissue they covered (100-200 μm). In order to determine if this increased droplet size could improve the number of detected ions in a typical murine brain experiment, protein MALDI-IMS was attempted using the ChIP-1000.

NOTE:
This figure/table/image has been removed
to comply with copyright regulations.
It is included in the print copy of the thesis
held by the University of Adelaide Library.

**Figure 4.6.** A 10 µm coronal murine (C57B) brain section was washed in 70% and 100% isopropanol (IPA) for 60 sec each. An ImagePrep station was used to deposit 10 mg/mL sinapinic acid (in 50% acetonitrile, 5% IPA and 0.2% trifluoroacetic acid) using a modified standard ImagePrep method with a repeat of phase one and use of a cover slip. MS data was acquired at a centre to centre distance of 100 µm using an ultraflex III MALDI-TOF/TOF instrument operating in linear positive ion mode. Included are a scan of the tissue (**a**) and ion intensity maps for m/z 14123.9 using rainbow (**b**) and green colour scales (**c**). Spectra were normalized to noise. Overlay of the tissue with the green scale ion map (**d**) and a myelin stained section of C57B brain from the high resolution mouse brain atlas (**e**) are included (www.hms.harvard.edu/research/brain/). Scale bars are 1 mm.

## 4.4 Application of the Chemical Inkjet Printer for protein imaging

The ChIP-1000 print head uses a solution reservoir surrounded by piezoelectric material. When current is passed through the piezoelectric material its shape alters slightly, generating force inside the reservoir, which in turn causes ejection of droplets from a 55 μm diameter print head. The print head and droplet generation are illustrated in Supplementary Figure 4.2. Solution is held inside the print head by a connected vacuum line which can be manually adjusted. Monitoring of ChIP-1000 prints was achieved via the in-built camera system which allowed direct observation of both test (Supplementary Figure 4.2a) and *in situ* prints (Supplementary Figure 4.2b). The test prints provided the opportunity to optimize the dwell time and voltage settings which affected the number and shape of the droplets generated. Optimization of dwell time and voltage while monitoring droplet shape in the test print window and as they impacted the tissue section allowed the quality of a ChIP-1000 preparation to be controlled with more confidence than the ImagePrep station. The volume of matrix printed at each position in a given array was chosen based upon previously reported preparations, observation of crystal appearance and subsequent MS results. A suitable amount of matrix was observed to be a white (2,5-DHB/SA) or pale yellow (CHCA) crystalline solid from which analyte ions, if present, could be generated with relatively minor modifications to the MALDI-TOF instrument settings optimized at installation.

Although the ChIP-1000 was implemented to solve S/N issues caused by poor analyte incorporation into the matrix, as reported in the previous chapters, matrix crystallization on and inside the print head of the ChIP-1000 drastically affected the ability to control reagent deposition **[93]**. For example, both SA and CHCA crystallized on and inside the piezo outlet at the concentrations required for IMS (5-25 mg/mL) and as a result consistently inhibited stable printing. Several troubleshooting steps were applied in attempts to better control ChIP-1000 printing of SA and CHCA.

These steps included:

1. Application of numerous print head washes (with 50% and 100% v/v IPA and MeOH) both pre and post printing to remove left over matrix,
2. Forced removal of air bubbles from the print head with a syringe plunger unit,
3. Alteration of instrument specific settings such as base voltage and vacuum as well as,
4. Alteration of the matrix concentration and composition.

Consistently washing as well as clearing air bubbles from the piezo print head proved to be the most successful steps. In addition to these factors, printing consistently deteriorated in the presence of a saturated filter paper at the test print/wash position. When the filter was exchanged the difficulty of droplet optimization was dramatically reduced. Changes to the vacuum were only necessary when the solution composition was changed. For example, for an aqueous solution (80-100% v/v aqueous), a setting of -0.15 to -0.25 kPa was preferable, while a 50% v/v organic solution (E.g. MeOH or ACN) typically required a decrease to about -0.4 to -0.5 kPa. In terms of composition, addition of IPA (5-10% v/v) improved print performance slightly, although this was subject to use of the troubleshooting steps outlined above. Despite troubleshooting, constant monitoring of the printing process was always necessary when using CHCA or SA. Previous publications recommended solid ionic matrixes to improve print stability. These ionic matrixes combined a molar excess (≥1.5 x that of the matrix) of aniline with either CHCA or SA **[61]**. **Figure 4.7a** and **4.7b** show the matrix crystals generated on murine brain using 10 nL of 20 mg/mL SA (in 6:4 ACN:0.1% TFA) with and without a 1.5x molar excess of aniline respectively. Print stability was increased in the presence of aniline, producing more homogeneous, crystalline spots. As previously reported, the S/N of subsequent MS spectra was also improved **[61]**. **Figure 4.7d** and **4.7e** show the sum spectra generated from 10 independent acquisitions on the SA/aniline and SA only crystals respectively, which demonstrated this increase.

Unfortunately, neither of the SA preparations matched the reproducibility of 2,5-DHB (**Figure 4.7c**). Stability for SA and CHCA preparations, even in the presence of aniline, could not be guaranteed and

consequently a reproducible method for protein MALDI-IMS could not be established for these matrixes. However, for 2,5-DHB concentrations of 20 mg/mL (in 50% MeOH and 0.2% TFA), deposition of 12 nL total at each print position was sufficient to generate reproducible arrays of “donut” shaped crystals approximately 150 μm in diameter. More importantly, due to its solubility in water, 2,5-DHB did not crystallize heavily on the piezo and as a result did not impair print stability. The caveat was that 2,5-DHB is suited to analysis of analytes smaller than 10 kDa **[26]**. Thus, despite the ability of the ChIP-1000 to deposit larger matrix volumes and thereby incorporate more protein analytes into the matrix, the ideal matrix for the ChIP-1000, 2,5-DHB, was not amenable to ionization of proteins. It was thus concluded that the most effective way to use the ChIP-1000 reproducibly, and maximize data quality, was to analyze low molecular weight molecules such as endogenous or tryptic peptides.

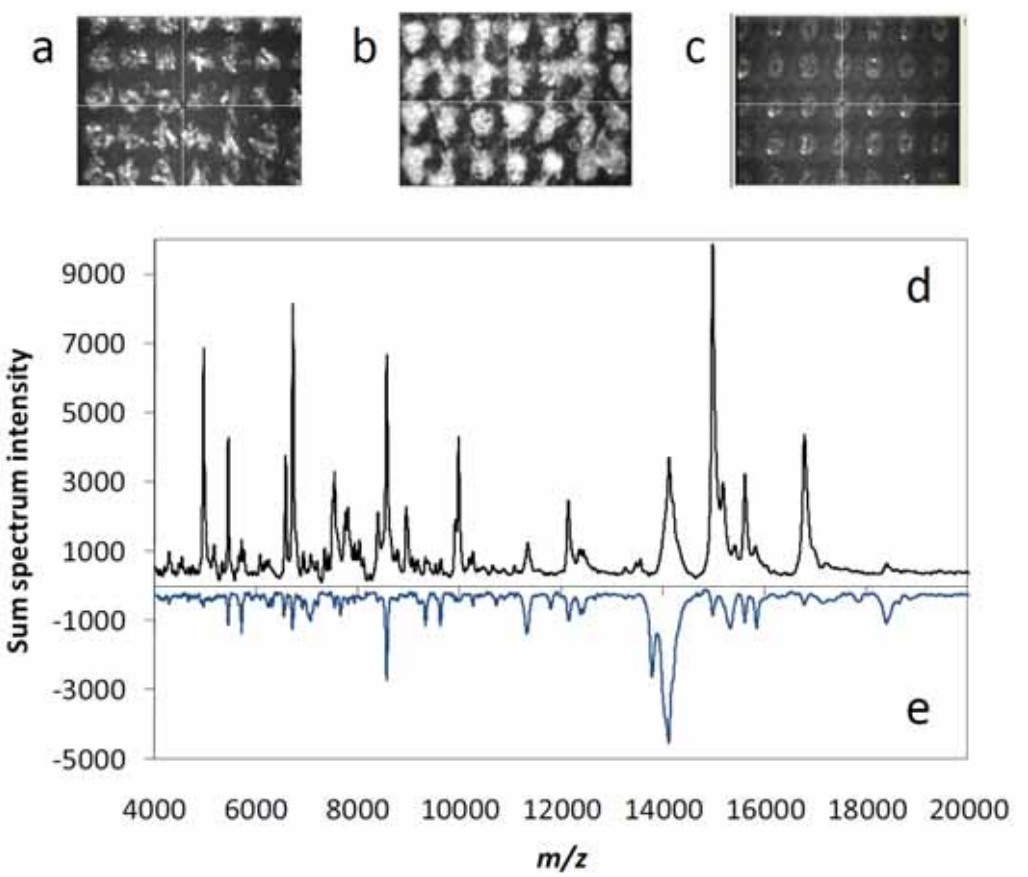



**Figure 4.7.** Ten µm thick coronal murine brain sections were washed horizontally with 70% isopropanol (+0.1% trifluoroacetic acid [TFA]) and 100% isopropanol for 10 s each. Two sections were covered with printed arrays of 20 mg/mL sinapinic acid (in 60% acetonitrile and 0.1% TFA) with (**a**) and without (**b**) 8.1 µL aniline per mL. For comparison, an image of 12 nL 20 mg/mL 2,5-DHB (in 50% methanol and 0.2% TFA) appears in **c**. MS data was acquired for arrays at 10 separate positions on each section (500 shots per position) using an ultraflex III MALDI-TOF/TOF instrument operating in linear positive ion mode. Spectra were processed by Gaussian smoothing and TopHat baseline subtraction before being exported as text files. Intensities for the two preparations were summed in MS Excel (2007) for each data point measured. Sum spectra of the individual acquisitions are shown for SA/aniline (**d**) and SA only (inverted in panel **e**).

## 4.5 Application of methods for tryptic peptide imaging

### 4.5.1 Tryptic peptide MALDI imaging mass spectrometry

Performing MALDI-IMS on tryptic peptides resulted in the first major deviation away from the aims described in chapter one. However, the change was made because peptide level MALDI-IMS came with several advantages, which stemmed from the limitations of protein MALDI-IMS as well as the difficulty of connecting protein distributions to protein identifications. The first challenge for protein imaging was that even if protein extraction for incorporation into the matrix was maximized, MS sensitivity decreased dramatically for masses larger than 20 kDa (see **Figure 4.1**). Only 29.4% of all proteins in the SwissProt protein database are smaller than 20 kDa (Supplementary Figure 4.3) and as a result this mass limitation excludes a substantial percentage of the proteome. Of those proteins that are larger than 20 kDa, very few can be detected by IMS **[31]**. The second challenge was that connecting a MALDI-IMS mass to an identity requires protocols to fractionate an associated cellular lysate. Each fraction can then either be further fractionated or digested with proteolytic enzymes for peptide identification by LC-MS/MS. Reverse-phase LC has been used for fractionation of lysates and correlation to IMS data. In these cases, either a) diagnostic MS of all fractions followed by digestion and LC-MS/MS identification **[28]** or b) top down protein fragmentation **[39]** were employed to assign identity. Protein fractionation methods were not pursued for this thesis due to the relatively long time and complex protocols required for sample preparation, fractionation and analysis combined with the advantages of "bottom up" tryptic peptide MALDI-IMS. As already discussed in chapter one, these advantages were the ability to image proteins larger than the demonstrated maximum of 70 kDa **[31]** and the relative ease of peptide extraction from tissue for identification by LC-MS/MS **[64]**.

### 4.5.2 Tryptic peptide imaging using a ChIP-1000

Printer systems have already been used to generate high quality peptide IMS data in the literature **[60, 62, 63]**. A published method was therefore used as a starting point for optimization **[62]**. The method selected used 4.8 nL of 45 ng/µL trypsin in 100 mM $NH_4HCO_3$ deposited at each print position **[62]**. The relatively high $NH_4HCO_3$ concentration maintained solution pH at a value suitable for trypsin (~8.5-8.8) and also neutralized the 50 mM acetic acid used as a trypsin storage buffer **[62]**. Keeping the concentration of introduced chemicals and salts to a minimum was critical for IMS applications, because any introduced substances can not be washed away and thus need to be compatible with MS. For example, the formation of homogeneous and small crystals, which are typical for an ideal MALDI matrix preparation, is prevented by salts in solution (data not shown). This consideration was especially important for IMS experiments because the tissue section surface already presented a less than ideal surface for MALDI given that 1) even after chemical section treatment it contains trace amounts of salt and lipid and 2) is an insulating surface that impedes efficient ionization. In order to ensure reproducible matrix crystallization, low chemical noise and high analyte signal, the digest method presented above was optimized to reduce $NH_4HCO_3$ concentration to 10 mM. This modification also required the addition of small volumes (0.5-1 µL) of 0.1 M NaOH to the trypsin solution to compensate for the 50 mM acetic acid buffer used to store the trypsin. For comparison, the pH of a solution containing 50 mM acetic acid and 10 mM $NH_4HCO_3$ was approximately 6, where as a solution containing 50 mM acetic acid, 100 mM $NH_4HCO_3$ and 12% v/v ACN was approximately 8.2-8.5. The reasons for ACN addition were two fold; to improve digestion efficiency by slight protein denaturation **[97]** and to improve print consistency for the otherwise aqueous trypsin solution.

The composition of the published tryptic digest protocol was:

- 500 µL 100 mM $NH_4HCO_3$ containing 45 ng/µL trypsin
- pH 8.2-8.5 (no pH adjustment)

Following modifications:

- 500 µL 10 mM $NH_4HCO_3$ with 12% v/v ACN containing 40 ng/µL trypsin
- pH 8.2-8.5 (adjusted from pH 6 with 0.1 M NaOH)

To evaluate the effect of this modified digestion solution on matrix crystallization, 250 µm centre to centre arrays of 5 nL trypsin gold (40 ng/µL) in either 10 mM (**Figure 4.8 a/c**) or 100 mM (**Figure 4.8 b/d**) $NH_4HCO_3$ (with 12% v/v ACN) were printed onto coronal murine brain sections using the ChIP-1000. CHCA/aniline (20 mg/mL in 6:4 ACN:0.1% TFA with a 1.5x molar excess of aniline) or 2,5-DHB (20 mg/mL in 50% MeOH and 0.2% TFA) were printed onto these digest arrays. MALDI closed circuit television (CCTV) images for these crystallized print locations are shown in **4.8a/b** and **4.8c/d** respectively. CHCA/aniline was included because it has been demonstrated in the literature to be superior to CHCA and improve print stability on a ChIP-1000 **[60, 61]**. As such it was important to evaluate any potential benefits to using CHCA/aniline instead of 2,5-DHB for peptide MALDI-IMS. Both matrix preparations were compared using images from the MALDI instrument CCTV system. The images showed that a ten-fold reduction in $NH_4HCO_3$ concentration made the matrixes more crystalline and homogeneous as well as improving the coherence of the matrix foci. In the case of CHCA/aniline, 9 and 10 nL were compared as a result of difficulty controlling print quality. However, despite the difference in volume, which negated direct spectral comparison, a clear difference in crystal morphology was observed. The unfocused, rounded crystals observed in the presence of 100 mM $NH_4HCO_3$ merged with adjacent print locations (**Figure 4.8b**). This was significant because any contact between adjacent print locations can cause delocalization of peptides. Furthermore, the rounded appearance of the crystals was similar to that typically observed on solid sample supports when a contaminant such as salt or polymer was present. However, discrete, homogeneous crystal foci were observed when the salt

concentration was reduced to 10 mM $NH_4HCO_3$ (**Figure 4.8a**). A similar improvement in crystallization was observed for 12 nL 2,5-DHB in the presence of 10 mM $NH_4HCO_3$ (**Figure 8c**) when compared to 100 mM $NH_4HCO_3$ (**Figure 8d**). The sum of these data suggested that a ten-fold reduction in salt was beneficial for stabilizing sample preparation and generating discrete printed foci *in situ.*

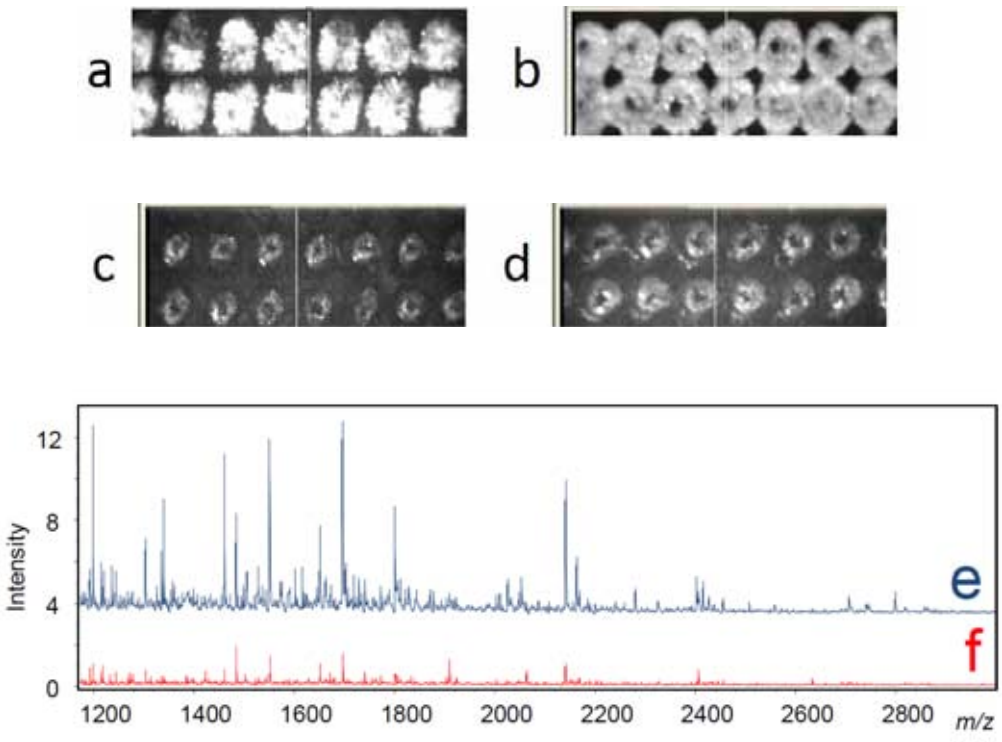



**Figure 4.8.** Crystallization of 9 nL (**a**) and 10 nL (**b**) CHCA/aniline (20 mg/mL in 6:4 ACN:0.1% TFA + 1.5 molar excess aniline) and 12 nL (**c/d**) 2,5-DHB (20 mg/mL in 50% MeOH and 0.2% TFA) in the presence of ChIP-1000 printed tryptic digests (5 nL 40 ng/µL trypsin and 12% v/v ACN, pH 8.2-8.5) with either 10 mM (**a/c**) or 100 mM (**b/d**) $NH_4HCO_3$. Crystallization was evaluated using MALDI closed circuit television images. MS data was acquired on an ultraflex III MALDI-TOF/TOF instrument operating in reflectron positive ion mode. Sum spectra from MALDI-TOF MS analysis of identical digests in either 10 mM (**e**) or 100 mM (**f**) $NH_4HCO_3$ on identical regions of coronal murine brain sections are included. Digest spots were crystallized using 2,5-DHB (12 nL 20 mg/mL 2,5-DHB in 50% MeOH and 0.2% TFA). All experiments utilized a ChIP-1000 to print solutions and all prints were 250 µm centre to centre.

To determine if analyte signal could be improved by the modified preparation, MS data from the two tryptic digest methods was also compared. **Figure 4.8e** and **4.8f** show the non-normalized sum spectra acquired from identical areas of coronal murine brain tissue digested using either 10 mM (**Figure 4.8e**) or 100 mM (**Figure 4.8f**) $NH_4HCO_3$ in the trypsin solutions and crystallized with 12 nL 2,5-DHB (composition identical to that used above). Importantly, the reduction in salt concentration caused a

marked increase in MS sensitivity. A 100 mM $NH_4HCO_3$ solution should maintain *in situ* pH more effectively and as such should generate more peptide product. Therefore any increase in sensitivity was most likely due to removal of salt and as a consequence improved peptide ionization efficiencies. Because MALDI-IMS will only access the high abundance portion of a proteome it was critical for this project to increase MS sensitivity where possible. This ensured that as many peptides as possible were imaged and therefore contributed to downstream molecular annotations of the tissue.

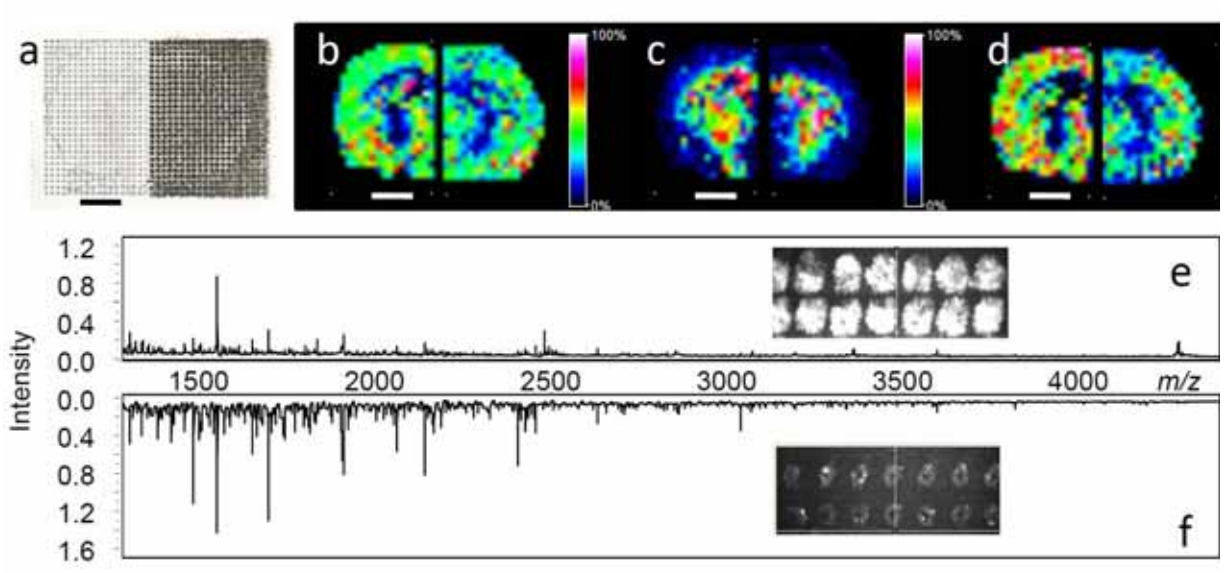



**Figure 4.9.** A comparison of CHCA/aniline (10 nL 20 mg/mL CHCA in 6:4 ACN:0.1% TFA + 1.5 molar excess aniline) and 2,5-DHB (12 nL 20 mg/mL 2,5-DHB in 50% MeOH and 0.2% TFA) matrix for imaging following tryptic digestion (5 nL 40 ng/µL trypsin gold in 10 mM $NH_4HCO_3$ and 12% v/v acetonitrile) on a coronal murine brain section. MS data was acquired on an ultraflex III MALDI-TOF/TOF instrument operating in reflectron positive ion mode. The printed matrix arrays on the murine brain section are shown in panel **a** (DHB on the left, CHCA/aniline on the right). Panel **b**, **c** and **d** show the ion intensity maps for *m/z* 1198, 1551 and 2063 respectively. The printed arrays were 250 µm centre to centre and the scale bars are 2 mm. Also included are flexImaging generated sum spectra for CHCA/aniline (**e**) and 2,5-DHB (**f**) with insets of the matrix crystal appearance acquired using the MALDI close circuit television system.

CHCA/aniline and 2,5-DHB were also directly compared for MS data quality to determine the optimal matrix for the digest protocol. This was necessary given the previously demonstrated high *in situ* S/N generated for tryptic peptides crystallized with CHCA/aniline **[60, 61]**. Thus, a single coronal section of murine brain was digested using a 250 µm centre to centre array of 5 nL 40 ng/µL trypsin gold (10 mM $NH_4HCO_3$ and 12% v/v ACN). Each brain hemisphere was then crystallized using either 12 nL of 2,5-

DHB or 9 nL CHCA/aniline **[60, 61]**. **Figure 4.9a** shows the prepared tissue section with the 2,5-DHB array on the left half of the brain section and CHCA/aniline on the right half. Panels **4.9b, 4.9c** and **4.9d** show the ion intensity maps for the example peptides *m/z* 1198, 1551 and 2063 respectively. The ion maps generated were almost equivalent following normalization to noise (see **panels 4.9b-d**), with the 2,5-DHB data showing a slight improvement in signal quality and therefore removing some of the granularity seen in the CHCA/aniline maps. Extraction of sum spectra from flexImaging showed that 2,5-DHB (**Figure 4.9f**) provided much greater non-normalized signal intensity than CHCA/aniline (**Figure 4.9e**). The inset images in **4.9e** and **4.9f** show the matrix crystals arrays as observed through the MALDI CCTV system. Similar to previous observations, 2,5-DHB crystals (**4.9f**) were more discrete than CHCA/aniline (**4.9e**), and far easier to generate reproducibly.

As trypsin stocks were generated fresh for each experiment and therefore not stored for longer periods of time acetic acid was not used as a trypsin storage buffer for subsequent experiments. All lyophilized trypsin stocks were subsequently re-suspended in 5 mM $NH_4HCO_3$. In a final experiment to demonstrate the reproducibility of the digest method (in the absence of acetic acid) as well as the lower limit for improving MS spectra quality by reducing salt concentration (**Figure 4.10**), non-normalized sum spectra were compared for duplicate coronal murine brain sections digested using 5 nL 40 ng/µL trypsin gold in the presence of either 5 mM or 10 mM $NH_4HCO_3$ (+12% v/v ACN). ChIP-1000 camera images of the 5 mM and 10 mM $NH_4HCO_3$ trypsin prints are shown in **Figure 4.10a** and **4.10b** respectively. The 12 nL 2,5-DHB matrix (composition identical to that used above) prints for these digests are shown in **Figure 4.10c** (5 mM $NH_4HCO_3$) and **Figure 4.10d** (10 mM $NH_4HCO_3$). The trypsin and matrix prints both indicated virtually identical preparation stability, as measured by the qualitative similarity of printed droplets. Furthermore, the appearance of the generated matrix crystals was also identical when examined using MALDI CCTV images in Figure **4.10e** (5 mM $NH_4HCO_3$) and **Figure 4.10f** (10 mM $NH_4HCO_3$). Importantly, the sum spectra were also reproducible (**Figure 4.10g**), with similar complexity and overall non-normalized intensity.

These data suggested that the modified tryptic digest method, using a 5-10 mM $NH_4HCO_3$ buffer and crystallized with 12 nL 2,5-DHB matrix, was sufficient to reproduce high S/N tryptic digests on murine brain. As such this digest preparation method was tested for applicability to tumour tissue from an ovarian serous epithelial carcinoma.

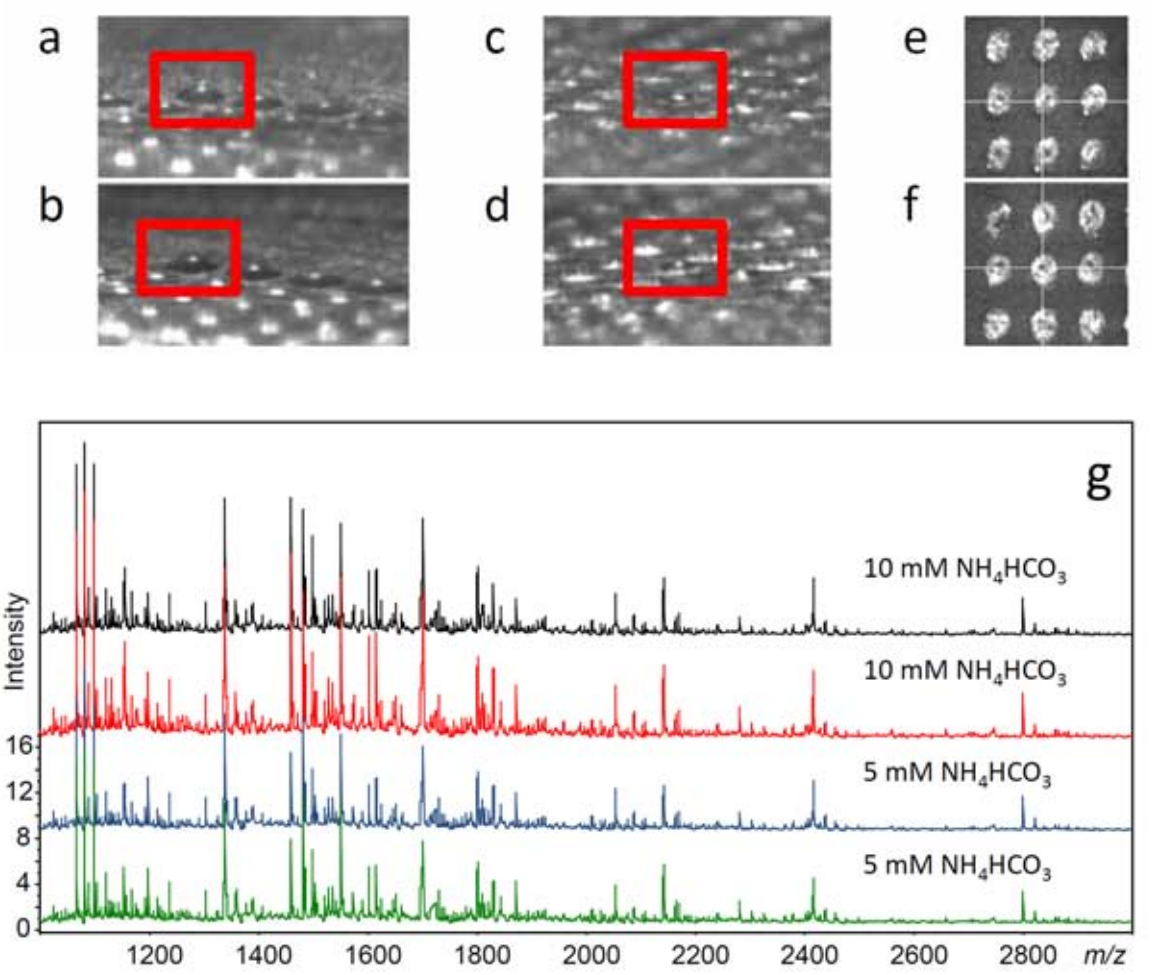



**Figure 4.10.** Duplicate coronal murine brain sections (total n = 4) were prepared by printing 5 nL 40 ng/µL trypsin in either 5 mM (**a**) or 10 mM (**b**) NH4HCO3 with 12% v/v ACN. Twelve nL 2,5-DHB (in 50% MeOH and 0.2% TFA) was printed onto the spots for the 5 mM (**c**) and 10 mM (**d**) digests. Images **a-d** were acquired using the ChIP-1000 instrument camera. MALDI-closed circuit television images of the 5 mM and 10 mM $NH_4HCO_3$ digest crystal arrays (250 µm centre to centre) are shown in **e** and **f** respectively. MS data was acquired on an ultraflex III MALDI-TOF/TOF instrument operating in reflectron positive ion mode. Non-normalized sum spectra generated from MALDI-TOF/TOF MS analysis are shown for the four sections (**g**) digested using either 5 mM (blue and green) or 10 mM (black and red) $NH_4HCO_3$.

## 4.6 Application of tryptic peptide imaging to ovarian cancer

In order to evaluate the ability of the in-house optimized digest method to generate meaningful data from ovarian cancer tissue, two 10 µm sections of a serous ovarian carcinoma tumour were washed in 70% IPA twice and once in 100% IPA (60 s each). They were dried and digested in a 250 µm centre to centre array of 40 ng/µL trypsin in 5 mM $NH_4HCO_3$ and 12% V/V ACN. The digest array was crystallized with 12 nL 20 mg/mL 2,5-DHB (in 50% MeOH and 0.2% TFA) and analyzed on an ultraflex III MALDI-TOF/TOF system. Non-normalized sum spectra for the duplicate sections appear in **Figure 4.11a** and **4.11b**. Haematoxylin and eosin stained scans of the sections are shown in **Figure 4.11c** and **4.11g**. Ion intensity maps for the two ions *m/z* 1629 and 1640 were combined in **Figure 4.11d** and **4.11h** with the individual ion maps for *m/z* 1640 in **Figure 4.11e** and **4.11i** while *m/z* 1629 is shown in **4.11f** and **4.11j**.

Similar to the data obtained for the murine brain sections, the ovarian cancer data was reproducible, both in terms of complexity and the intensity of spectra generated (**Figure 4.11a**). Comparing these spectra in the *m/z* range 1000 to 2000 (as in the figure) showed that the data was virtually identical for both sections. In addition, the anatomical structures outlined in the H&E stain, including the cancerous tissue (**4.11e/i**) and stroma (**4.11f/j**) could be discriminated using the ion intensity maps given as examples in **Figure 4.11**. The sum of these data indicated that the printed digest method optimized for use with the ChIP-1000 was applicable to real samples.

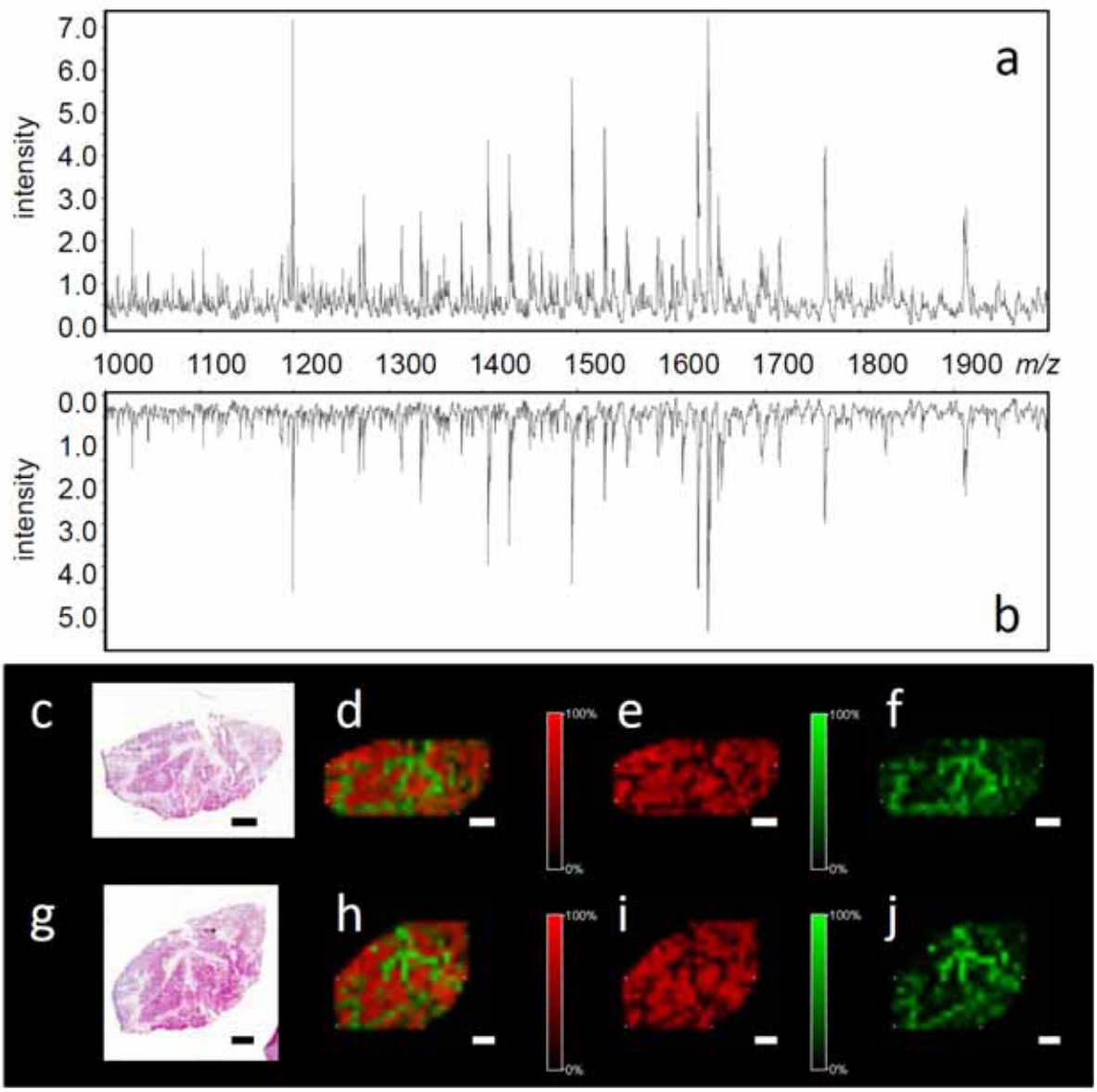



**Figure 4.11.** Ten µm sections of frozen serous ovarian tumour were mounted onto ITO slides and washed twice with 70% IPA and once with 100% IPA. A ChIP-1000 printer was used to deposit 5 nL trypsin gold (in 5 mM $NH_4HCO_3$ and 12% v/v ACN) and 12 nL 2,5-DHB (20 mg/mL in 50% MeOH and 0.2% TFA) in 250 µm centre to centre arrays. MS was performed on an ultraflex III MALDI-TOF/TOF instrument in reflectron positive ion mode. Spectra were combined in flexImaging (V2.1, Bruker Daltonics) to generate sum spectra for duplicate sections in panel **a** and **b**. Haematoxylin and eosin stains of the sections are shown in **c** and **g**. Combined ion maps for *m/z* 1629 (green) and 1640 (red) appear in **d** and **h**. Individual ion maps appear in **e** and **i** for *m/z* 1640 and **f** and **j** for *m/z* 1629. Scale bars are 1 mm. Ion intensity scales are included.

## 4.7 Concluding remarks

The successful demonstration of a reproducible *in situ* tryptic digest method marked the completion of the first aim of this thesis. As a result, tryptic peptide MALDI-IMS data could be performed at a spatial resolution of 250 µm (centre to centre) on sections derived from frozen ovarian cancer specimens. However, it was important to make further method improvements to achieve the following:

1. Resolution of analysis of 100 µm centre to centre or less and,
2. Reduction of sample preparation time.

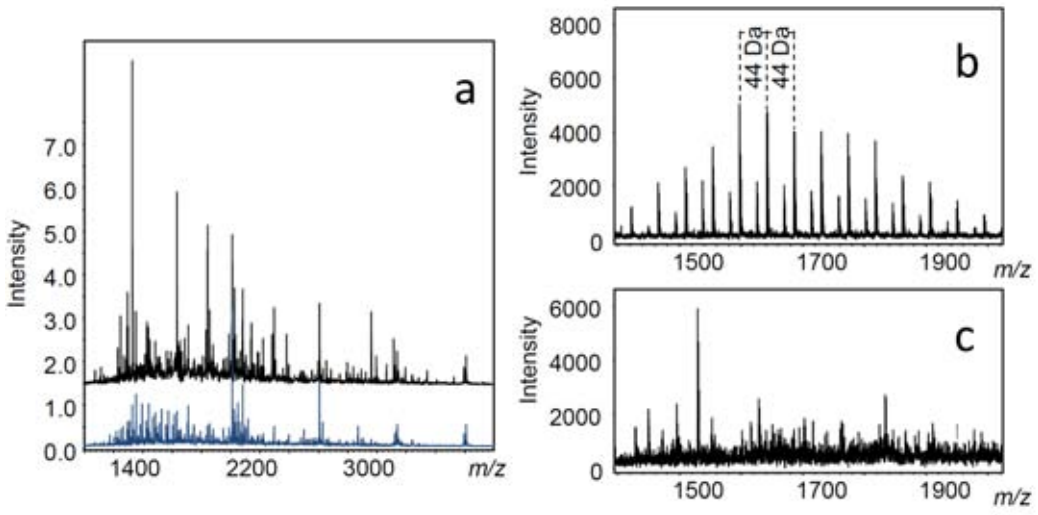
Firstly, higher resolution analyses were required to complement printed acquisitions because of the inherent internal heterogeneity of cancer tissues. If the spatial resolution could be decreased to 100 µm or less it would be possible to visualize much finer anatomical regions, as demonstrated for protein MALDI-IMS of the corpus callosum in murine brain (**Figure 4.6**). A side by side comparison of the improvement possible for protein ion intensity maps produced using the ImagePrep was also presented in Figure **3.2f** and **3.2g** (see **chapter 3**).

Secondly, it was necessary to implement a more rapid automated method for sample preparation alongside the ChIP-1000 because a full slide of four murine brain sections (approximately 1.0 x 0.5 cm each) took eight to ten hours to prepare using the printer. If large tumour sections spanning several square centimeters were prepared in this way it would be impractical to prepare a whole slide in one day.

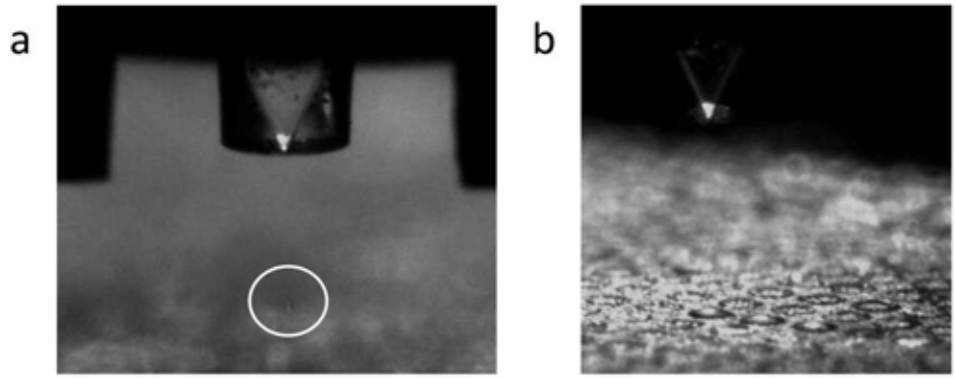
As a result of these requirements, and in addition to the work presented in this chapter, methods for ImagePrep tryptic digestion were employed. The sensitivity issues encountered with the ImagePrep station for protein MALDI-IMS could be circumvented by working at the peptide level, as peptides were easier to incorporate into the matrix and ionized more efficiently. Furthermore, the speed of preparation could be increased significantly. For example, a complete ImagePrep method including trypsin

incubation and matrix deposition required five hours for preparation of a single slide. Most importantly, the application of an ImagePrep digest method coincided with the optimization of an antigen retrieval (AR) method for accessing FFPE tissues **[64]**. These developments and their contribution to achieving aim two of this thesis project are discussed in detail in **chapter 5**.

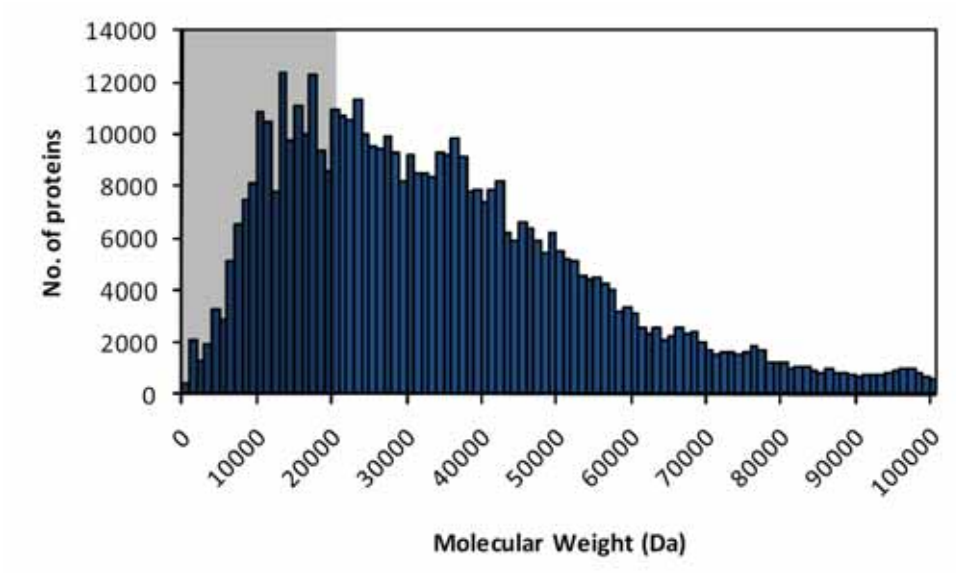
## 4.8 Supplementary Information




Supplementary Figure 4.1. Sum spectra for two sections of human serous papillary ovarian carcinoma prepared by either a wash in 50 mL of 70% isopropanol (IPA) and 100% IPA for 60s each (**a**, black spectrum), or 2x 200 µL rinses with chloroform (**a**, blue spectrum). Sections were digested using a ChIP-1000 printer delivering 5 nL 40 ng/µL trypsin (in 5 mM $NH_4HCO_3$ and 12% acetonitrile) per position. Digest spots were crystallized using 12 nL 20 mg/mL 2,5-DHB each (in 50% methanol and 0.2% trifluoroacetic acid). MS data was acquired using an ultrafleXtreme MALDI-TOF/TOF instrument (Bruker Daltonics) operating in reflectron positive ion mode. Sum spectra were generated in flexImaging (V2.1, Bruker Daltonics). Individual spectra from the chloroform rinsed (**b**) and the IPA washed sections (**c**) are included.




Supplementary Figure 4.2. ChIP-1000 printing. Panel **a** shows the ejection of a droplet at the test print position onto the filter paper wash station. Panel **b** shows an in situ matrix print. Print positions in panel **b** are 250 µm centre to centre.




Supplementary Figure 4.3. Histogram showing protein mass frequencies in the SwissProt protein database (up to 100 000 Da). The database listed 525 207 proteins in total as of February 2011. The grey box shows the typical mass region that could be analyzed by IMS without using protocols optimized for higher mass proteins. This represented 29.4% of annotated proteins in the database.

# Chapter 5 Manuscript Context

The preceding chapter demonstrated that tryptic peptide MALDI-IMS could be performed reproducibly for tissue sections sourced from frozen ovarian cancer tissue. Thus, aim one of this thesis project was completed. To achieve aim two, the methods developed as part of **chapter 4** needed to be applied to FFPE tissue and modified where necessary to ensure reproducible MALDI-IMS data acquisition. Methods to access FFPE tissues were critical because of the existing large FFPE archives which have been stored alongside patient specific data and histological annotations for decades. In the context of the final aim of this thesis – the application of a MALDI-IMS method to characterizing ovarian cancer metastasis – analyzing archived FFPE tissues presented a unique opportunity because of the ease of sectioning FFPE material, the availability of numerous archived FFPE tissue samples and the inherent longevity of a formalin-fixed sample. It would thus be of benefit to create a workflow combining tissue from these archives with a reproducible, high resolution tryptic peptide MALDI-IMS method.

In order to address this aim, a method was established which combined antigen retrieval (AR), tryptic digestion and MS in one workflow. Methods for AR combined with printed digestion and matrix application have been presented in the literature **[62]**. However, because the tissue specific effects of AR buffer systems were unknown, it was considered worthwhile to adapt an AR method which was typically used for IHC on FFPE ovarian cancer sections. As a result, a citric acid antigen retrieval (CAAR) method was modified to ensure compatibility with MALDI-IMS and its performance was compared to that of another commonly used AR buffer, Tris/EDTA **[62]**. Furthermore, as outlined at the end of **chapter 4**, additional modifications were made to the already established tryptic peptide MALDI-IMS method to achieve higher resolution analyses (≤ 100 µm centre to centre). This was required because of the heterogeneous and complex internal architecture of cancer tissues. These developments were published in the Journal of Proteome Research and are presented for this thesis in **chapter 5**.

# Chapter 5 Citric Acid Antigen Retrieval (CAAR) for Tryptic Peptide Imaging Directly on Archived Formalin-Fixed Paraffin-Embedded Tissue

Johan O. R. Gustafsson,† Martin K. Oehler,‡ Shaun R. McColl,§ and Peter Hoffmann*,†

† Adelaide Proteomics Centre, School of Molecular and Biomedical Science, The University of Adelaide, SA 5005, Adelaide, Australia,

§ Chemokine Biology Laboratory, School of Molecular and Biomedical Science, The University of Adelaide, SA 5005, Adelaide, Australia, and

‡ Robinson Institute, School of Paediatrics and Reproductive Health, The University of Adelaide, SA 5005, Adelaide, Australia

*J. Proteome Res.*, 2010, 9 (9), pp 4315–4328

**DOI:** 10.1021/pr9011766

Copyright © 2010 American Chemical Society

Received December 19, 2009

Publication Date (Web): July 5, 2010

**Keywords:** antigen retrieval • MALDI • trypsin • proteomic methods • imaging mass spectrometry

# Statement of authorship for chapter 5

**Ove Johan Ragnar Gustafsson (Candidate)**

Statement of contribution (in terms of the conceptualization of the work, its realization and its documentation)

Performed analysis on all samples presented, interpreted data and wrote the manuscript

Certification that the statement of contribution is accurate

Signed.....Date.....

**Martin K. Oehler (co-author)**

Statement of contribution (in terms of the conceptualization of the work, its realization and its documentation)

Provided tissue samples and manuscript evaluation

Certification that the statement of contribution is accurate and permission is given for the inclusion of the paper in the thesis

Signed.....Date.....

**Shaun R. McColl (co-author)**

Statement of contribution (in terms of the conceptualization of the work, its realization and its documentation)

Manuscript evaluation

Certification that the statement of contribution is accurate and permission is given for the inclusion of the paper in the thesis

Signed.....Date.....

**Peter Hoffmann (co-author)**

Statement of contribution (in terms of the conceptualization of the work, its realization and its documentation)

Supervised development of work, helped in data interpretation and manuscript evaluation and acted as corresponding author

Certification that the statement of contribution is accurate and permission is given for the inclusion of the paper in the thesis

Signed.....Date.....

Johan O. R. Gustafsson, Martin K. Oehler, Shaun R. McColl and Peter Hoffmann. (2010) Citric Acid Antigen Retrieval (CAAR) for Tryptic Peptide Imaging Directly on Archived Formalin-Fixed Paraffin-Embedded Tissue.
*Journal of Proteome Research, v. 9 (9), pp. 4315-4328, September 2010*

NOTE: This publication is included in the print copy of the thesis held in the University of Adelaide Library.

It is also available online to authorised users at:

http://dx.doi.org/10.1021/pr9011766

# Chapter 6 Evaluation of internal calibrants for tryptic peptide MALDI-IMS data

## 6.1 Introduction

Aims one and two of this thesis were achieved in **chapters 4** and **5** respectively. The manuscript presented in **chapter 5** also encompassed the tentative correlation of tryptic peptides mapped by MALDI-IMS to tryptic peptides extracted from adjacent sections and identified by LC-MS/MS. However, the confidence in assignment of peptide identification was not as good as it could be, given that the variation in *m/z* for the peptides mapped by MALDI-IMS was not known.

To address this issue and complete aim three, it was first necessary to develop a processing workflow for automatically processing and graphically presenting the results from several gigabytes of MALDI-IMS data.

The development of an automated processing method was required for several reasons. Firstly, although IMS data reduction and processing tools have been presented**[112]**, no software tool was readily available which allowed automated analysis of tryptic peptide MALDI-IMS data sets. In addition, an automated processing methodology would remove operator-induced bias for selection of peaks of interest. Finally, an automated strategy circumvented the generation of sum spectra, which can mask low abundance or highly localized peak features. The result of this work, with the code writing assistance of Mr. James S. Eddes (Adelaide Proteomics Centre), was the IonMapper tool, which is described in full in results section **6.2.1**. Using customized flexAnalysis scripts for peak feature extraction and IonMapper, data set size was dramatically reduced, thousands of ion intensity maps could be rapidly generated as well as organized and the *m/z* deviation for peak features across a data set could be investigated en masse for the first time.

The ability to use IonMapper to evaluate *m/z* deviation across a whole data set provided the opportunity to increase the confidence of the tentative identifications assigned in **chapter 5** and subsequently complete aim three of this thesis. Improving confidence in peak *m/z* was critical because the ultimate aim of the MALDI-IMS experiments in this thesis was to connect peak feature distribution to peak feature identity. While this has been demonstrated using accurate mass for small molecules such as lipids **[113]**, identification of mapped peptides typically requires extraction and LC-MS/MS identification **[64]**. The main issue affecting the matching of data from a peptide MALDI-IMS experiment with high accuracy LC-MS/MS peptide data was that the TOF mass analyzer used for MALDI-IMS assumed a constant distance from target (i.e. tissue) to detector. The slide-tissue interface and internal heterogeneity of the tissue section introduced variations in target height and thus caused a spread in the TOF derived *m/z* values for single ion species measured across a section. Consequently, matches between TOF based IMS experiments and LC-MS/MS data could not be made with high confidence. For example, in **chapter 5**, the matches made between MALDI-IMS and LC-MS/MS data used a match window of ± two Da. This tolerance was purposely made large because the typical *m/z* deviation across an entire tissue section was unknown. Furthermore, a window of this size would be more than large enough to encompass any variability in measured *m/z*. The ideal solution to these *m/z* deviations would be to use a mass analyzer which relies on measurement of an ion characteristic not influenced by sample height on the slide, for example, Q-TOF, FTICR or Orbitrap instruments. However, MALDI-TOF systems were those available for this thesis. Furthermore, MALDI-Orbitrap, MALDI-FTICR and/or MALDI-Q-TOF systems are expensive and suffer from issues including long acquisition times (due to the low repetition rate lasers), high purchase/running costs as well as a lack of proprietor support for MALDI-IMS applications. Finally, TOF mass analyzers are the standard platform for MALDI-IMS **[26]**; a preference which is based on high sensitivity (atto to femtomolar), large mass range (0-8 kDa for reflectron mode) and rapid analysis times when compared to either FTICR or Ion Trap instruments **[26]**.

As a result, an understanding of the extent of tissue induced *m/z* deviation on a MALDI-TOF system was considered critical.

In this chapter the typical *m/z* deviation across a MALDI-IMS data set was shown to be ± 0.2 Da. This equates to a 166 ppm error at 1200 Da. A recent publication showed that 160 *unmodified* tryptic peptides could exist in the range *m/z* 1235.624 ± 100 ppm for an *in silico* digest of the *Mus musculus* SwissProt database **[114]**. These findings indicated that methods were required which could reduce the MALDI-IMS *m/z* data spread to at least 50 ppm and as a result increase confidence in assignments between MALDI-IMS and LC-MS/MS data. As a result, an internal calibration mixture was developed and added to the CAAR tryptic peptide MALDI-IMS method presented in **chapter 5**. The *m/z* deviation across tissue sections before and after internal re-calibration of MS spectra was investigated using IonMapper. It was shown that the mass deviation of tryptic peptides, from their abundance weight mean (AWM) *m/z*, could be reduced to less than 50 ppm at the same time as calibrating the entire data set.

## 6.2 Results and Discussion

### 6.2.1 Automated generation of ion maps, mass deviation maps and grouped peak lists

When analysing peptide MALDI-IMS data there were several confounding variables that needed to be considered. Firstly, MALDI-IMS data sets are difficult to analyse because of their large size (up to $10^4$ spectra, 1-2 gigabytes per section), particularly if data was acquired at high resolution. Secondly, when generating a sum spectrum in flexImaging (MALDI-IMS data visualization software) peptides which are not detected frequently, or which are highly localized to a small tissue region, exist below the apparent sum spectrum noise level. Finally, the complexity of a tissue tryptic peptide digest causes significant

overlap between peptide species of similar mass. In this case, ion intensity maps generated from the flexImaging sum spectrum could contain more than one peptide, giving a false distribution. In an effort to address these issues, an in-house analysis method was developed in conjunction with Mr. James S. Eddes (Adelaide Proteomics Centre). The method encompassed a customized flexAnalysis script applied to each MS spectrum in a MALDI-IMS data set. The flexAnalysis script processed spectra to remove chemical noise (width *m/z* 5), baseline subtracted the spectra (TopHat), picked peaks (SNAP II) and internally re-calibrated the spectra. Following processing a peak list (ASCII format) was exported for each MS spectrum with the X and Y coordinates within the file name. The peak lists were subsequently processed by a java software tool, called IonMapper, which combined all the lists into one master list before iteratively grouping all peaks within a set mass tolerance (e.g. 0.2 Da). An abundance weighted mean (AWM) *m/z* was calculated and used as a label for each group. All groups were then exported as ASCII files containing X and Y locations, *m/z*, intensity and S/N. The S/N of groups existing in 50 or more MS spectra (minimum peak appearance) was plotted at each spatial coordinate. This was the equivalent of the ion intensity maps discussed previously. Furthermore, the *m/z* deviation of all peaks within a group from their AWM *m/z*, at each spatial coordinate, was plotted using a red (-0.1 Da), white (0 Da), green (+0.1 Da) colour gradient. This was referred to as a mass deviation map. Both plot types were exported as portable network graphic (.png) files labelled with the AWM *m/z* of the peak group. A schematic of the process used to group peaks and generate ion intensity and mass deviation maps is shown in **Figure 6.1**.

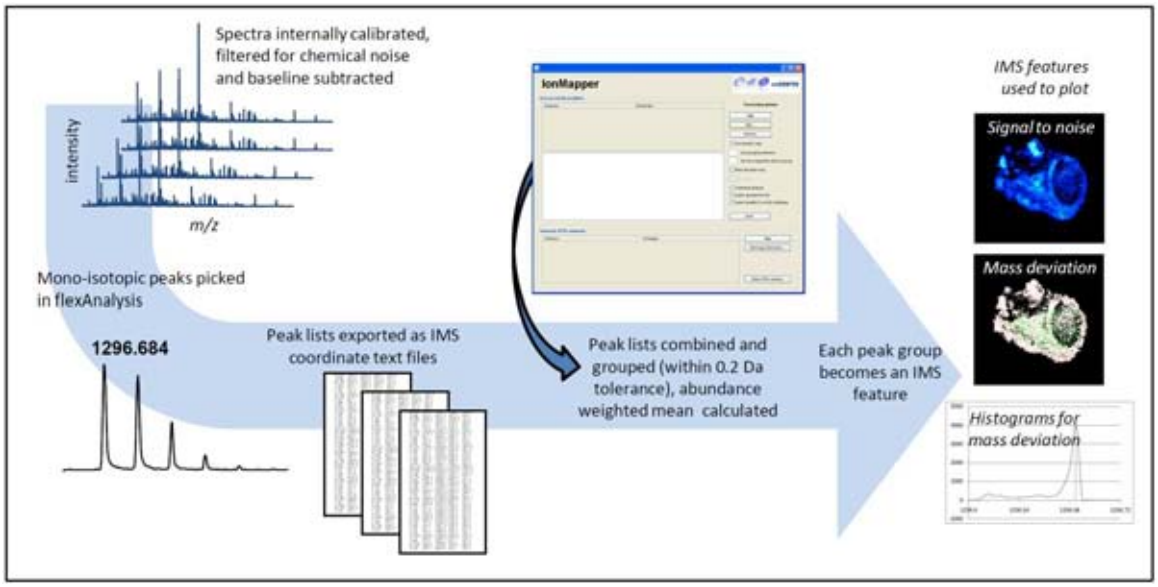



**Figure 6.1.** Workflow for generating and processing peak lists: Data acquired on an ultrafleXtreme MALDI-TOF/TOF MS instrument was processed by internal calibration, Gaussian smoothing and TopHat baseline subtraction of spectra in flexAnalysis software (V3.3, Bruker Daltonics, Bremen, Germany). Mono-isotopic peaks from each spectrum/IMS coordinate were exported as text files and used to generate peak groups. Abundance weighted mean (AWM) *m/z* values were calculated for each peak group. The signal to noise (S/N) of *m/z* values within the peak groups were used to plot S/N maps, while the *m/z* values within a peak group were used to plot mass deviation maps (deviation from AWM *m/z*).

The application of this process was beneficial for several reasons. Firstly, exporting peak lists from the processed profile spectra reduced the data volume by several orders of magnitude. Secondly, automated generation of ion S/N maps ensured that all peaks existing in a user defined number of spectra were included for analysis. This meant that peaks hidden in a sum spectrum were plotted and potentially important peptide masses were thus not excluded. Furthermore, because the process was automated, thousands of ion S/N maps could be efficiently generated, organized and compared across multiple sections. In addition, assuming that SNAP II always correctly picked the mono-isotopic peptide peak, unless two peptides with differential distributions had exactly the same *m/z* value, the existence of two unique peptide distributions could be inferred from these ion S/N maps and validated by further experiments. Finally, the mass deviation maps provided, for the first time, a graphical presentation of the *m/z* deviation within peak groups. For further analysis of *m/z* deviation, the peak group text files could

be used to plot *m/z* frequencies. The ability to investigate *m/z* deviations within peak groups proved of great value for implementation of internal calibrants, which is described further in **6.2.2**.

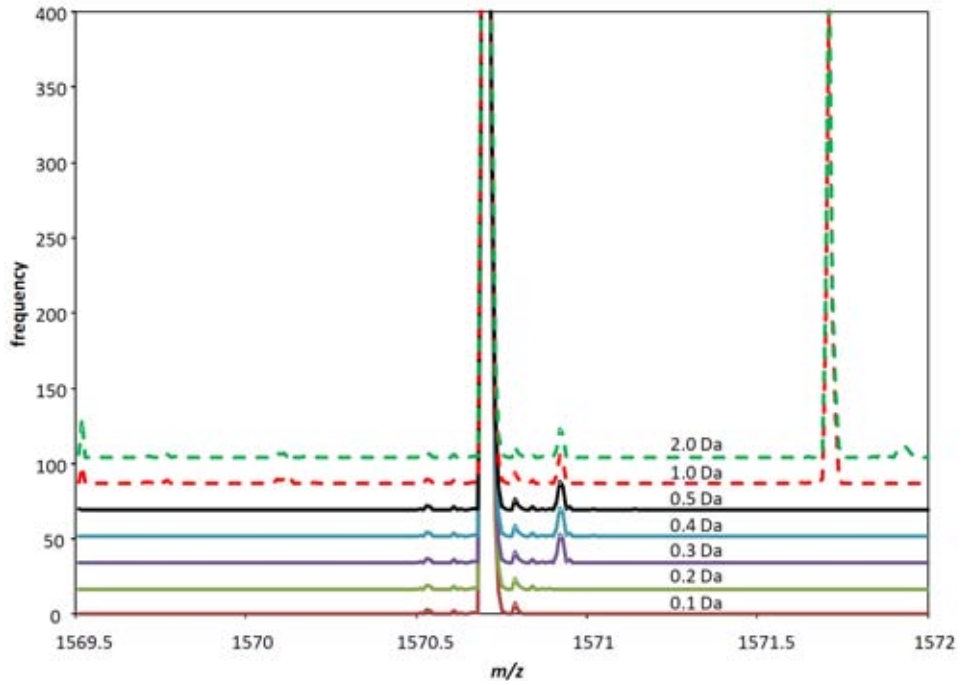



**Figure 6.2.** Peak lists for a single section of ovarian peritoneal implant tumour were processed using variable peak grouping tolerances (0.1 to 2.0 Da), and the frequency of the *m/z* values in the subsequent peak groups were plotted for the [Glu$^1$]-Fibrinopeptide B internal calibrant of *m/z* 1570.6.

Two key variables – peak grouping tolerance and minimum peak appearance – could be modified in the IonMapper graphical user interface. **Figure 6.2** shows the effect of altering the peak grouping tolerance. Peak lists from a single FFPE ovarian cancer tumour section analyzed by tryptic peptide MALDI-IMS were processed with seven different grouping tolerances (ranging from 0.1 Da to 2.0 Da). The *m/z* frequencies within the peak group for the internal calibrant m/z 1570.6 ([Glu$^1$]-Fibrinopeptide B) were subsequently plotted for **Figure 6.2**. Tolerances of 0.1 and 0.2 Da gave the best result, as this tolerance prevented the appearance of other masses in the peak group. Ideally, only one mono-isotopic peptide species should contribute to one peak group. As such, a tolerance of 0.2 Da was selected as standard

for processing because it appeared to effectively group the peak lists such that only one mono-isotopic mass was present.

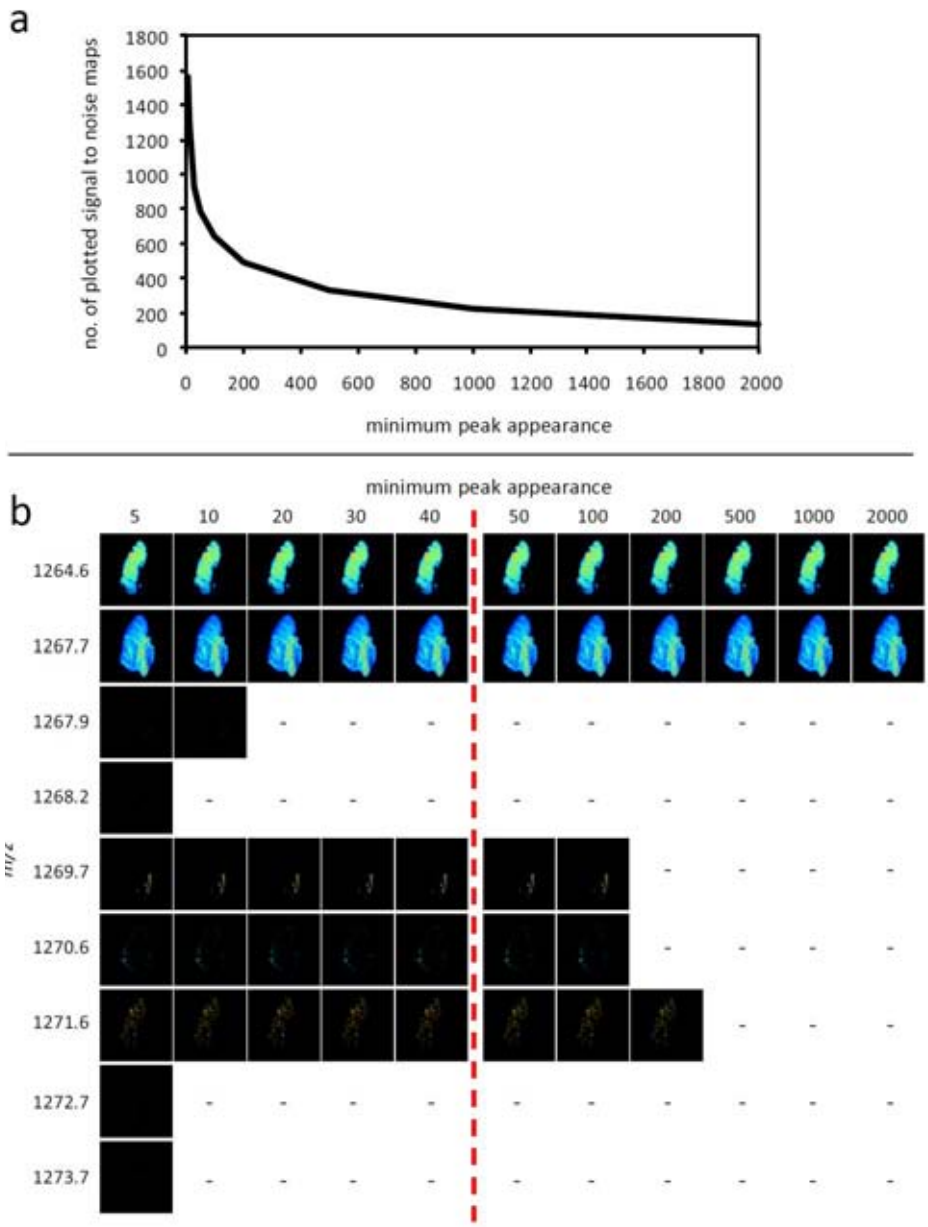



**Figure 6.3.** Peak lists for a single section of ovarian peritoneal implant tumour were processed using variable minimum peak tolerances (with grouping tolerance of 0.2 Da). The number of plotted ion signal to noise maps for each minimum peak tolerance appears in panel **a**. An example of the maps created for a sub-set of nine peak groups is shown in panel **b**. The red line indicates the selected standard value minimum peak tolerance of 50.

**Figure 6.3a** shows the effect of modifying the minimum peak appearance. As expected, there was a consistent decrease in the number of plotted ions for increasing minimum peak appearance values. Qualitatively, a minimum of 50 appeared to give the best balance between removing pixelated noise and keeping important peak groups (see **Figure 6.3b**). The caveat was that despite the usefulness of the minimum peak appearance in removing noise features, some features localized to small areas could be missed. Assuming a square shape for a tissue section area, a minimum peak appearance of 50 will filter out areas with 707 µm edges, or approximately seven cell diameters. Thus, for identifying small, unique tumour areas, this minimum peak appearance value would need to be combined with measurements of the distance between plotted signal centres. In this way, highly localized peak groups could be kept in favour of widely dispersed groups. This would allow a lowering of the minimum peak tolerance while still removing noise features which are not localized to distinct tissue regions.

With demonstration of an automated tool for processing peak lists from a tryptic peptide MALDI-IMS experiment, the *m/z* deviation across tissue sections could now also be addressed.

### 6.2.2 Characterizing m/z deviation across tissues using internal calibrants

The majority of mass variation in a MALDI-TOF MS measurement will be caused by variations in sample height. For a typical MALDI target (e.g. Bruker Daltonics AnchorChip solid sample support) this is caused by crystal heterogeneity and worsens for more angular matrixes such as 2,5-DHB. Imaging experiments need to additionally compensate for variable surface height caused by both the presence of tissue and changes in tissue composition. To determine where the majority of mass variation occurs in a given IMS experiment, a section of FFPE serous ovarian carcinoma peritoneal implant tissue was mounted onto an ITO slide (see **Figure 6.4a**), the paraffin was stripped with xylene and the section washed with EtOH. An ImagePrep station was used to deposit a calibration mix (50 fmol/µL each of

Angiotensin I, Renin substrate and ACTH (clip 18-39) in 200 µL 0.1% TFA) followed by 7 mg/mL CHCA (in 60% ACN and 0.2% TFA) onto the section using default preparation methods. MS data was acquired on an ultrafleXtreme MALDI-TOF/TOF instrument in reflectron positive ion mode. Because no tryptic digest was performed, only the calibrants themselves were present during acquisition. This allowed determination of the *m/z* measurement variability without intereference from endogenous analytes. The in-house-developed analysis method described above was used to plot mass deviation maps for each of the calibrants, which appear in **Figure 6.4b** (Angiotensin I), **6.4c** (Renin substrate) and **6.4d** (ACTH clip 18-39).

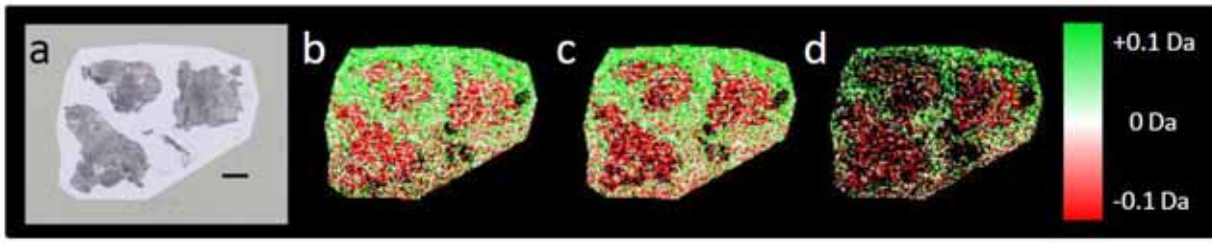



**Figure 6.4.** The deviation (Da) of Angiotensin I (**b**), Renin substrate (**c**) and ACTH [clip 18-39] (**d**) from their calculated abundance weighted mean (AWM) m/z plotted as mass deviation maps for a single section of formalin-fixed ovarian cancer tissue (**a**). Scale bar is 2 mm and a mass deviation intensity scale is included.

The AWM *m/z* values for all three calibrant peptides were 1296.913 (Angiotensin I), 1759.171 (Renin substrate) and 2465.505 Da (ACTH clip 18-39). These values were all slightly greater than their theoretical *m/z* (see **Table 6.1**). This was most likely caused by the inclusion of the off-tissue spectra. The ITO slide surface improves peptide ionisation efficiency relative to the tissue, which is an insulating surface. As a result, the greater intensity of the calibrant peptide ions off tissue means they contributed more to the AWM *m/z*, shifting its value higher. As expected, a consistent negative *m/z* deviation from the AWM *m/z* was also observed on tissue, confirming that the tissue sample presented a raised sampling surface for the MALDI-TOF/TOF instrument. This decreased ion TOF and as a result also decreased measured *m/z*.

**Table 6.1.** Name, theoretical mass ($[M+H]^+$) and sequence of the internal calibrants used for imaging mass spectrometry of peptides from formalin-fixed tissues.

| *Calibrant peptide* | *Theoretical mass $[M+H]^+$* | *Sequence* |
|---|---|---|
| Angiotensin I (34-43) | 1296.685 | DRVYIHPFHL |
| [Glu$^1$]-Fibrinopeptide B | 1570.677 | EGVNDNEEGFFSAR |
| Renin substrate | 1758.933 | DRVYIHPFHLVIHN |
| Dynorphin A | 2147.199 | YGGFLRRIRPKLKWDNQ |
| ACTH fragment (18-39) | 2465.198 | RPVKVYPNGAEDESAEAFPLEF |
| ACTH fragment (1-24) | 2932.588 | SYSMEHFRWGKPVGKKRRPVKVYP |

### 6.2.3 Use of internal calibrant peptides to reduce *in situ m/z* deviation

The deviations described so far demonstrated the expected effect of a raised sampling surface, such as tissue, to TOF measurements. Although the tissue-slide interface introduced the greatest *m/z* deviations, the *m/z* of endogenous masses was also variable across tissue sections (see **Figure 6.5**, raw calibration). Methods were thus developed for re-calibrating the MS data. The typical procedure for a MALDI imaging experiment was to externally calibrate prior to sample MS acquisition **[64]**. However, external calibrants will always be at a different sampling height to the tissue. It was possible to calibrate on an adjacent tissue section of identical thickness, but this would not compensate for heterogeneous tissue thickness and was not viable when using precious human tissue samples. The most straightforward way to re-calibrate MALDI imaging data was thus to include internal calibrant masses in the sample preparation steps. To achieve this, a formalin-fixed ovarian implant tumour section was processed by a modified CAAR protocol **[64]** and digested by ImagePrep Trypsin deposition. An optimized mixture of four readily available calibrants (100 µL total with 0.1% v/v TFA) was sprayed onto the section following digestion. The calibrants used in the optimized mixture were Angiotensin I, [Glu$^1$]-

Fibrinopeptide B, Dynorphin A and Adrenocorticotropic Hormone fragment (ACTH clip 1-24). The theoretical *m/z* values and amino acid sequences for these standard peptides are shown in **Table 6.1**. The internal calibrants were not added to the trypsin solution as three of the four calibrants contained lysine and/or arginine residues, rendering them sensitive to proteolysis. The prepared sections were coated with CHCA matrix as before and MS data was acquired at 100 μm spatial resolution. Following spectral processing (see **section 6.2.1**), the exported peak lists were used to generate mass deviation maps. Four example maps for the raw calibration indicated that endogenous masses, 1264.6 (**Figure 6.5b**), 1501.7 (**6.5e**), 1992.9 (**6.5h**) and 2719.2 (**6.5k**) deviated by up to 0.1 Da. Furthermore, a systematic decrease in *m/z* occurred across the measurement area, from top left to bottom right (green to red), indicating either a gradation in the slide surface or tissue thickness (see **6.5h** and **6.5k**).

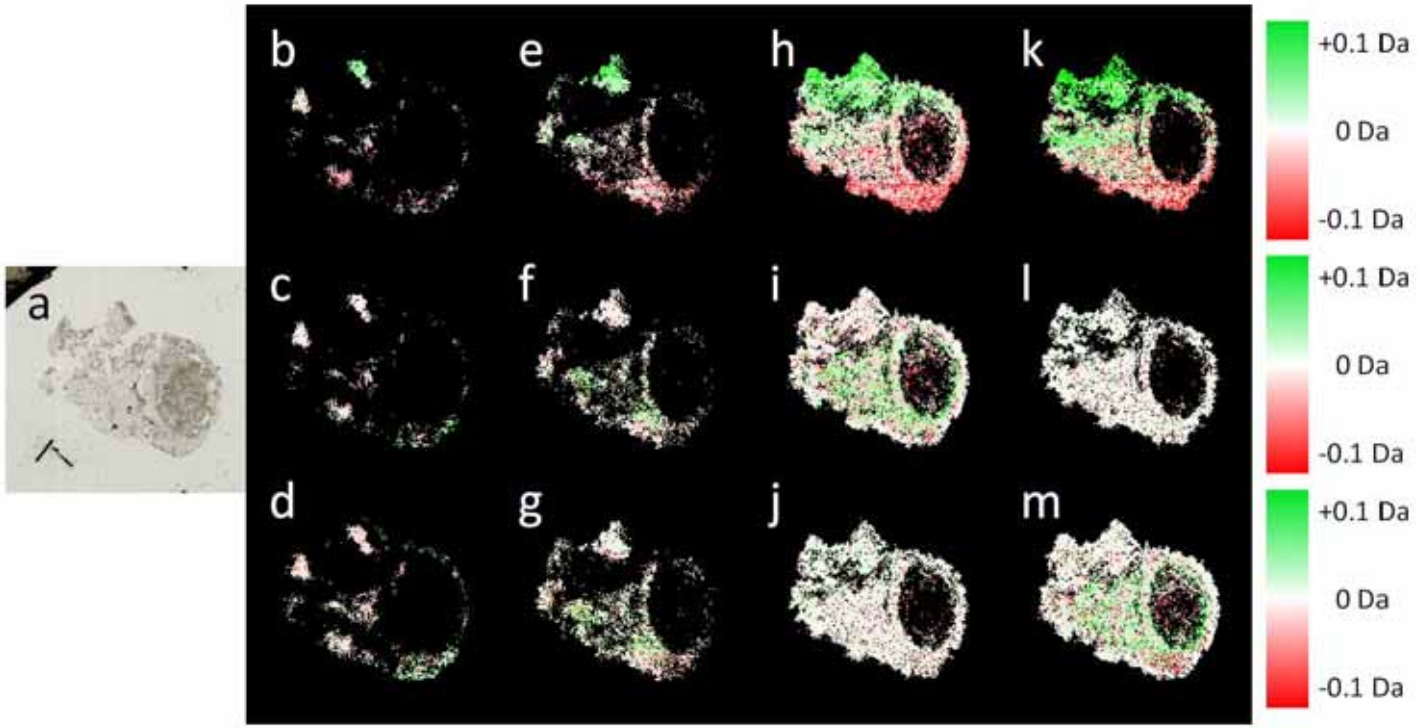



**Figure 6.5.** A single ovarian tissue section was prepared by antigen retrieval, tryptic digestion, deposition of internal calibrants and CHCA matrix followed by MALDI-TOF analysis. A tissue section scan (**a**) and mass deviation maps (±0.1 Da scale) for the peak groups (abundance weighted mean *m/z*) 1264.6, 1501.7, 1992.9 and 2719.2 are shown prior to internal calibration in **b**, **e**, **h** and **k** respectively. Post internal re-calibration mass deviation maps are shown for these same peak groups in the same order. Spectra were processed (prior to peak list extraction) either by a chemical noise filter in **c**, **f**, **i** and **l** respectively or smoothing (Gaussian) in **d**, **g**, **j** and **m** respectively. Peak grouping tolerance was 0.2 Da and minimum peak appearance was set to 20. Mass deviation scales are included.

To compensate for the apparent mass deviation of endogenous peaks from their AWM *m/z*, the internal calibrants deposited onto the tissue section were used to batch re-calibrate the entire data set. The four standard calibrant peptides as well as one trypsin autolysis product (*m/z* 2211) were used as points for a quadratic recalibration (flexAnalysis V3.3, Bruker Daltonics). **Figure 6.6** shows a sum spectrum from the analyzed tissue with the internal calibrants highlighted by red arrows. As the sum spectrum shows, the peptide calibration standards were selected to cover the *m/z* range where the majority of peptide ions were observed (see **Figure 6.6**). Importantly, the sum spectrum also indicated that the concentrations in the optimized mixture were sufficient for the standards to ionize well in the presence of a complex *in situ* mixture of tryptic peptides. As such no further changes were made to the calibrant mixture.

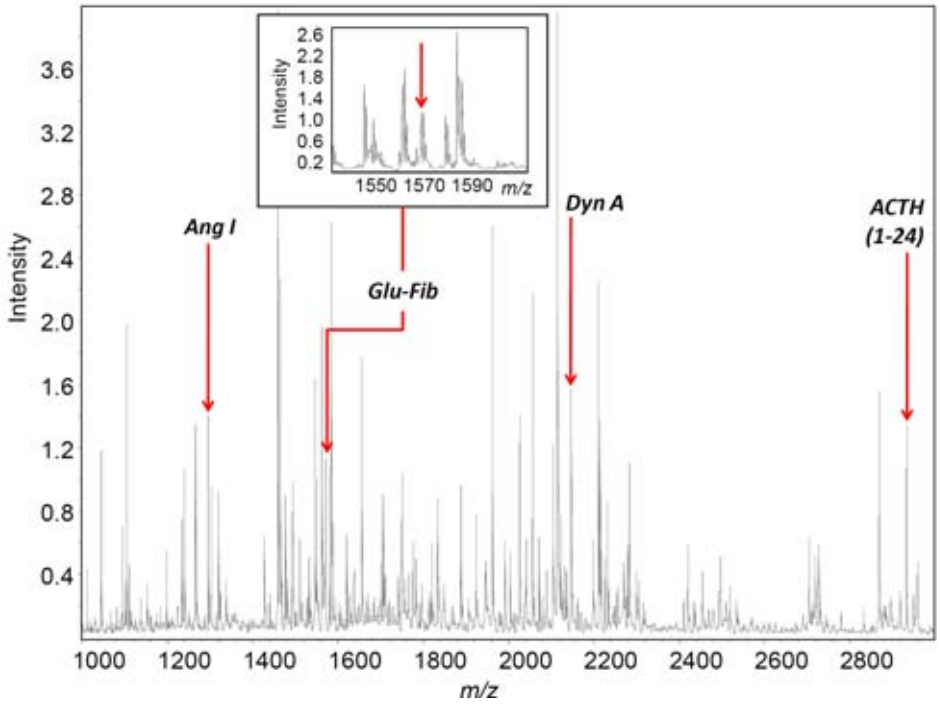



**Figure 6.6.** Example sum spectrum for MALDI-IMS data set with internal standards Angiotensin I, [Glu1]-Fibrinopeptide B, Dynorphin A and ACTH fragment (clip 1-24) spiked in. The red arrows indicate the locations of the various ions in the sum spectrum, with an inset zoom view for Glu-Fib, at *m/z* 1570.6.

Following internal re-calibration, repeat processing to generate mass deviation maps (identical to raw calibration processing) showed that the internally calibrated data exhibited a marked decrease in mass deviation, as indicated by the intensity decrease for the red and green zones in the mass deviation maps (see **Figure 6.5c, 6.5f, 6.5i and 6.5l**). Furthermore, the mass error gradation observed across the analyzed area disappeared, correcting what appeared to be a systematic bias in the data set. To further demonstrate the effect of the re-calibration, the same four peak groups were plotted as pre and post calibration histograms in **Figure 6.7a-d**. The figure indicated that the internal calibration reduced data spread and as a result also changed the AWM *m/z*. The re-calibration appeared to increase in effectiveness for larger peptides, as indicated by **Figure 6.7c** and **6.7d**. *M/z* histograms for the internal calibrant peptides are shown in **Figure 6.8a-e** for peak groups 1296.6 (Angiotensin I), 1570.7 ([Glu1]-Fibrinopeptide B), 2147.2 (Dynorphin) and 2932.6 (ACTH) respectively. Importantly, the internal calibrant peak groups aligned well with their theoretical masses.

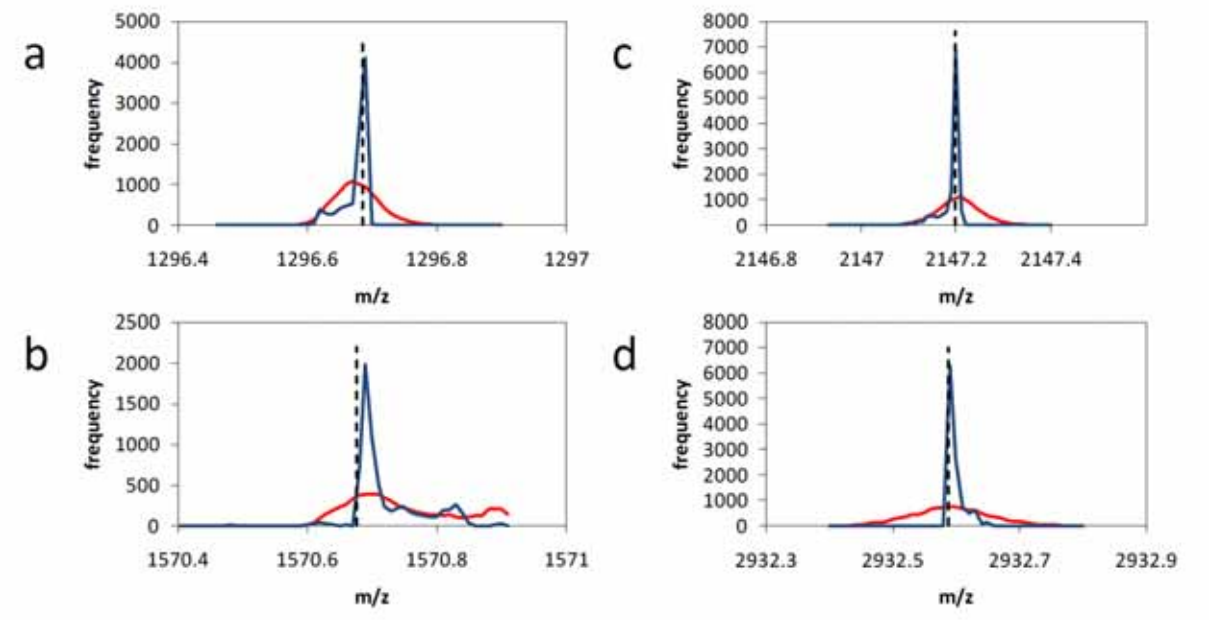



**Figure 6.8.** *m/z* histograms for the internal standards Angiotensin I (**a**), [Glu1]-Fibrinopeptide B (**b**), Dynorphin A (**c**) and ACTH fragment (clip 1-24) [**d**], pre (red) and post (blue) internal re-calibration. Black dotted line shows the theoretical mass of the peptide.

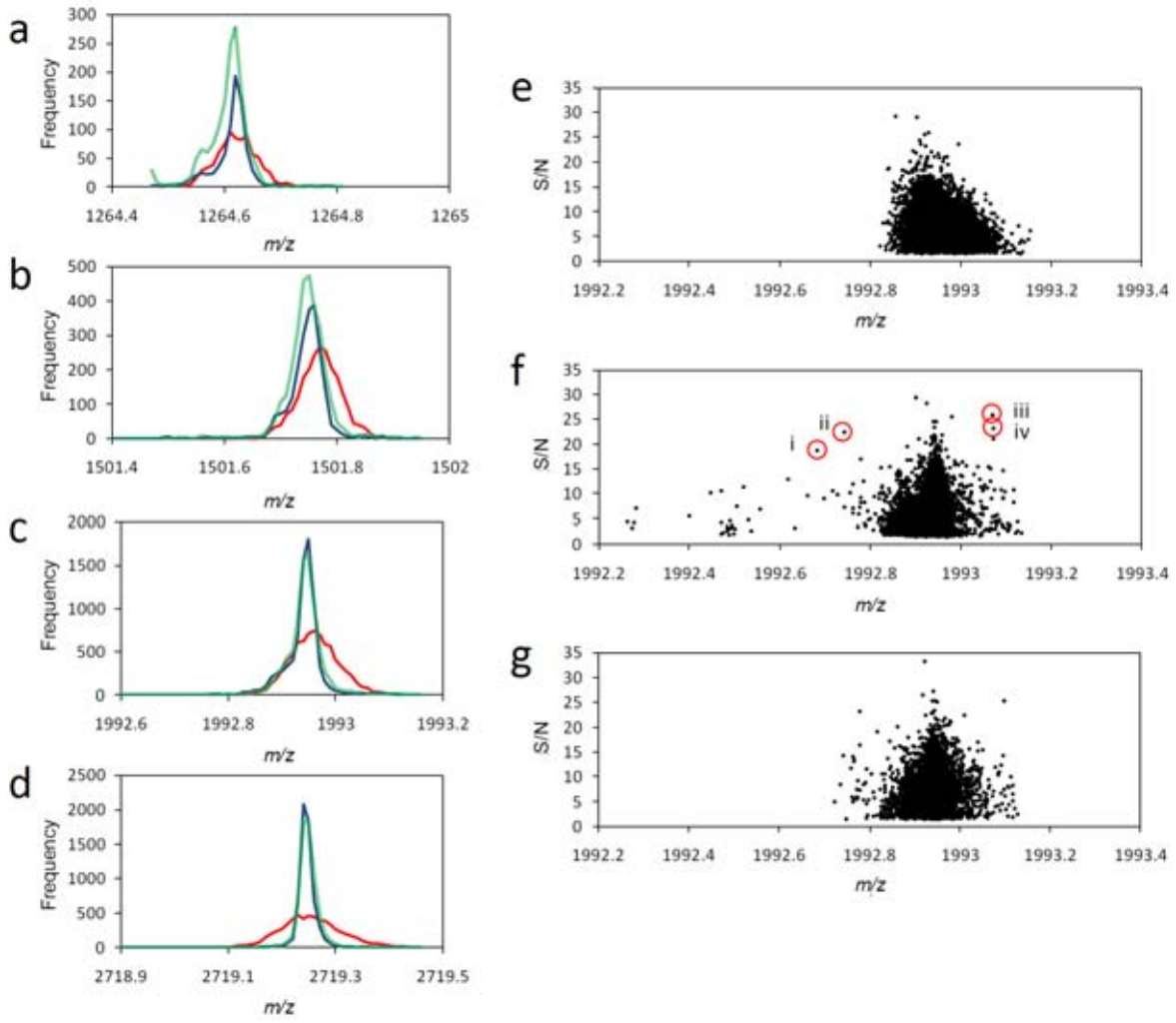



**Figure 6.9.** Histograms of the m/z for the peak groups 1264.6 (**a**), 1501.8 (**b**), 1992.9 (**c**) and 2719.3 (**d**). Histograms for the raw (red line), chemical noise filtered internally re-calibrated (blue line) and Gaussian smoothed internally re-calibrated data (green) are shown in each graph. Signal to noise (S/N) scatter plots for all *m/z* values in the raw (**e**), chemical noise filtered internally calibrated (**f**) and Gaussian smoothed internally calibrated (**g**) data for the 1992.9 peak groups are also shown. Example spectra of high relative S/N that are mis-calibrated are indicated by roman numerals **i-iv** in **d**.

While the re-calibration was successful for the majority of peaks in a given group, some re-calibrations resulted in increased deviation from the AWM *m/z*. The S/N of the *m/z* values comprising the 1992.9 peak group were plotted in **Figure 6.9e** for the raw data. The data spread evenly across 0.3-0.4 Da of the *m/z* range surrounding 1992.9. **Figure 6.9f** shows the *m/z* spread for the re-calibrated data on an

identical *m/z* scale. Four individual spectra (i-iv) were selected as examples because they exhibited a high S/N coupled to poor re-calibration relative to the AWM of 1992.9. Spectra i and iii did not contain Dynorphin A, the trypsin *m/z* 2211 autolysis product or ACTH (clip 1-24), and as such the *m/z* was re-calibrated to a lower value. Conversely, Angiotensin I was missing and Dynorphin A was of low intensity for spectra iii and iv resulting in a re-calibration to a higher *m/z*. These data demonstrated imperfect deposition of the internal calibrants across the analyzed tissue section. As a result it was not possible to re-calibrate all spectra. Slight variation in deposition quality was expected. It was not expected, however, that significant calibration errors would occur, given that the calibration portion of the flexAnalysis script was conditional, such that a spectrum would only be re-calibrated if the new calibration had a smaller standard deviation than the original.

It was subsequently found that the processing of spectra prior to peak grouping was important to improve the calibrations. The reason being that all spectra were processed with a chemical noise filter to reduce noise and allow for easier visualization of more peak features in the flexImaging sum spectrum. However, the noise filter does not smooth peaks and as a result peak picking efficiency was affected. Application of a gaussian smoothing algorithm changed the number of peaks picked for peak groups across the data set and subsequently affected the calibration of the data. **Figure 6.9g** shows a S/N plot of the peaks picked with the gaussian smoothed internally re-calibrated data for peak group 1992.9. Because the peak group now showed a significant reduction in the number of outliers, the optimum procedure for the automated processing and peak list extraction was changed to include gaussian smoothing prior to peak picking. To further demonstrate the effect of changing to a gaussian smoothing algorithm prior to peak picking, additional mass deviation maps for the smoothed internally calibrated data appear in **Figure 6.5d** (peak group 1264.6)**, 6.5g** (peak group 1501.7)**, 6.5j** (peak group 1992.9) and **6.5m** (peak group 2719.2). Mass deviation histograms for the smoothed data set are shown for the same peak groups in **Figure 6.9a-d** respectively.

Subsequently, if sum spectrum visualization in flexImaging was required, the spectra could be re-processed using the chemical noise filter. This could be done without affecting downstream results as mass selection was typically performed manually in flexImaging and the intensity of a selected mass was always taken by the software as the maximum value in the mass range selected.

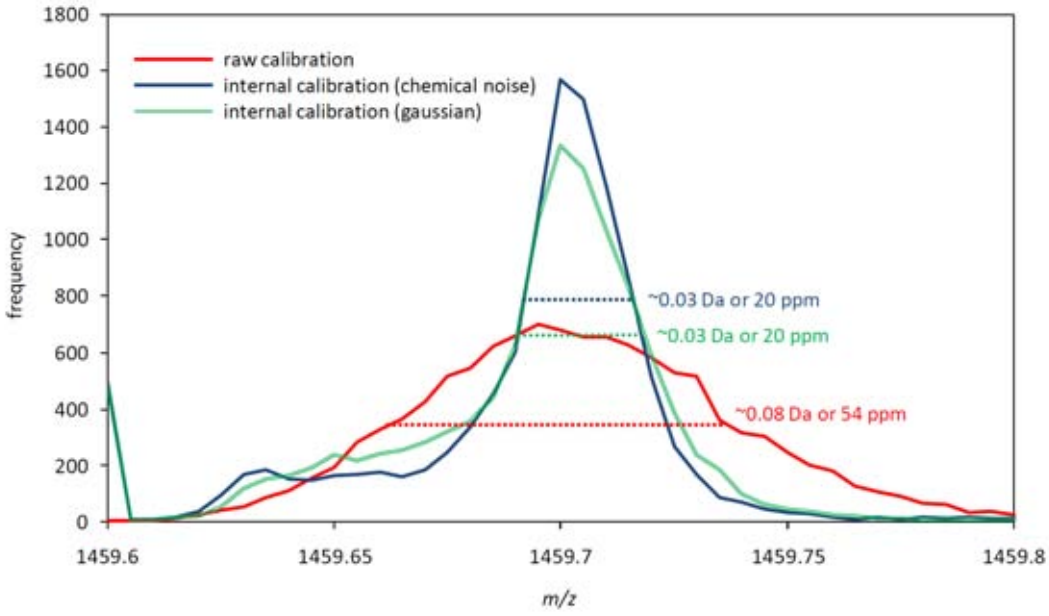



**Figure 6.10.** Histogram above shows the raw and internally calibrated data for peak group 1459.7. Shown are the histograms for the raw data (red line) as well as data processed with a chemical noise filter (blue line) and a Gaussian smoothing algorithm (green line) prior to internal re-calibration.

As a result of the observed qualitative improvement in *m/z* deviation all existing MS spectra were re-processed using the internal re-calibration. However, it was important to demonstrate the quantitative improvement to the mass deviation as a result of calibration. Unfortunately, the standard deviation of *m/z* did not indicate a marked improvement. The most likely cause of this was the shoulder present in the majority of the re-calibrated *m/z* histograms. **Figure 6.10** exemplifies the effect of the internal re-calibration on peak group 1459.7, which was selected due to a wide shoulder region located on the low mass end of the re-calibrated histogram (blue and green histograms in **Figure 6.10**). To quantitatively demonstrate the calibration improvement, the *m/z* deviation across this peak group was taken at the full

width half maximum (FWHM) *m/z* deviation. As shown in **Figure 6.10**, the *m/z* deviation was reduced from 54 ppm at the FWHM for the raw data to 20 ppm for the internally re-calibrated data. An examination of the peak groups presented in **Figure 6.9** clearly showed similar but not identical raw to re-calibration *m/z* deviation improvements.

It was thus demonstrated in this chapter that re-calibration can be used to increase confidence in AWM *m/z* values derived from a tryptic peptide MALDI-IMS experiment. The disadvantage of the approach originates from the imperfect deposition, and subsequent MS detection, of all calibrants. When using a MALDI-TOF instrument, assuming adequate S/N for the calibrants, the quality of a calibration typically increases with the number of calibration landmarks. Here it was shown that using a FWHM *m/z* value derived from a *m/z* histogram, the mass deviation across a data set could be reduced by about half for the peak group example presented. Key to this was the exclusion of those spectra for which calibration was not successful due to absence of one or more calibrants. Because it was shown that a *m/z* deviation decrease similar to the tryptic peptide peak groups was observed for the internal calibrants (see **Figure 6.8**), it was concluded that calibration was successful for most spectra, but additional developments to sample preparation and re-calibration processing were required to ensure re-calibration of all spectra. The improvements to sample preparation will undoubtedly include alterations to current methods and potentially also different deposition instrumentation. In terms of sample preparation, increasing the volume of the calibrant solution will most likely increase the homogeneity of calibrant deposition. In effect, because a greater number of spray cycles would be required to deposit the solution, there would also be a greater chance that all areas of the tissue are exposed to the calibrant spray. However, this volume increase may need to be accompanied by a slight concentration increase for the calibrants to compensate for the solution lost to spray onto the areas surrounding the ITO slide. In the context of new equipment, the HTX TM sprayer (HTX Technologies), for example, appears to allow very fine control over solution deposition parameters as demonstrated by the

homogeneity of matrix deposition demonstrated in the literature **[92]**. The demonstration of homogeneous matrix deposition indicates that a homogeneous coating of calibrant solution would not be difficult to achieve using this instrument

Regardless of the improvements made to sample preparation, to make the most of the internal calibrants also requires the alignment of spectra to each other based on landmark features. The approach used for this would be similar to alignment of MALDI-TOF profiling data **[115, 116]**. Unfortunately, these approaches do not adequately address the increased computational workload required to align the thousands of spectra generated in a complete MALDI-IMS experiment **[115, 116]**. Furthermore, most of the work to date has been achieved for profiling data of centroided linear MALDI-TOF data, not reflectron data which contains the isotopic profiles of low mass features like peptides **[116]**. For example, a 2007 manuscript described options for re-alignment including the *msalign* tool available in Matlab. Unfortunately, the processing time per spectrum can reach a full minute for this Matlab tool **[116]**. The authors rightly concluded that the processing time required to align thousands spectra makes the use of such tools impractical **[116]**. Other available alignment approaches including software such as ClinProTools (Bruker Daltonics) and even freeware developed software tools such as SpecAlign face the same issue **[117]**. Thus, the challenge lies in finding a way to minimize computational workload and balance alignment quality with the time required for data processing.

Although more work remains to address the complexities of matching MALDI-IMS data to LC-MS identifed peptides, the data presented in this chapter demonstrated the principle advantage and utility of internal calibrants in the context of a tryptic peptide MALDI-IMS workflow; the ability to internally re-calibrate reflectron MS data en masse for downstream analysis. In the context of future research, assuming that the spectra from a MALDI-IMS data set can be re-aligned to each other in a sensible time frame, these internal calibrants could be used again to re-calibrate the aligned data.

## 6.3 Concluding Remarks

The FWHM value for *m/z* confidence was only illustrative at this point and was not the final solution to calibration errors introduced by TOF measurement from a heterogeneous surface. The ultimate solution, which will be applied in future, is the re-alignment of all spectra to each other. The alignment could then utilize the internal calibrants to correct the calibration problems discussed above. Unfortunately, the application of alignment algorithms to the data presented here required work that lay beyond the immediate scope of this thesis. For the purposes of this thesis the internal calibrants were able to reduce the *m/z* deviation within peak groups at the same time as calibrating the data sets (see **Figures 6.5 and 6.10**). Aim three was thus completed, as the matching of LC-MS/MS identified peptides to MALDI-IMS peak groups could now be achieved with greater confidence than before. Therefore, the complete tryptic peptide MALDI-IMS work flow was coupled with internal re-calibration, IonMapper processing and LC-MS/MS to identify key peptide markers of ovarian peritoneal metastasis. The results of these analyses and their contribution to achieving aim four of this thesis are discussed in detail in **chapter 7**.

# Chapter 7 – Characterization of peritoneal ovarian cancer metastasis by tryptic peptide imaging mass spectrometry

## 7.1 Introduction

The majority of patients that present with ovarian cancer are diagnosed with advanced stage disease (see **Table 1.1** in **chapter 1**). The diagnosis of advanced stage (stage 3) disease requires histological confirmation of metastasis outside the pelvis and/or in the regional lymph nodes **[9]**. An ovarian carcinoma metastasizes by shedding of tumour cells from the primary tumour and implantation into the inner lining of the abdominal cavity, the peritoneum **[118]**. The implantation process is initiated through interactions between mesothelial surface proteins and receptors on the metastatic cancer cell **[119]**. Subsequent tumor invasion into the peritoneal stroma occurs via tissue disruption as well as promotion of tumor angiogenesis and growth **[119]**. A dramatic increase in morbidity is associated with the dissemination of tumor cells throughout the peritoneal cavity, despite first line treatment consisting of radical surgery and adjuvant chemotherapy **[120, 121]**. The morbidity increase is the result of the bulky peritoneal disease typical of the advanced stage cancers. The tumour bulk causes interruption of bowel motility leading to irreversible bowel obstruction, which is the terminal event in the majority of patients **[122, 123]**. Peritoneal metastasis of ovarian carcinomas thus represents an important area of research, in particular for application of novel technologies that can increase our understanding of the processes underlying metastasis and more importantly improve treatment options, which are currently not optimized for treatment of specific tumour sub-types or grades **[9, 120]**.

In this context, several manuscripts have been published which detail the application of molecular MS profiles for characterizing cancer. For example, a 2003 manuscript described successful discrimination of non-small cell lung cancer (NSCLC) sub-types based upon protein MS profiles generated directly from tumour and normal lung tissue **[46]**. Several more MALDI-IMS studies have been published since

detailing the use of MS based markers for grading of soft tissue sarcomas **[47]** and follicular lymphoma **[49]** as well as classification of HER2 receptor status in breast cancer **[39]** and prediction of short and long term patient survival for patients with malignant gliomas **[54]**. These studies all demonstrated that useful information could be extracted from MS profiles of human cancer, including grade, classification and prognosis **[124]**. All the aforementioned manuscripts analyzed the protein profile of tissue sections by MALDI-IMS. However, a key study published in 2008 analyzed FFPE tissues by printed tryptic peptide MALDI-IMS **[62]**. The tryptic peptide MS profiles generated from NSCLC and normal lung tissue sections were able to successfully classify adenocarcinoma and squamous cell carcinoma. The assignment of tumour sub-type based on a tryptic peptide MS profile emphasized the potential wealth of information which could be generated by application of the methods developed in **chapters 5** and **6** to metastatic ovarian tumours. The MALDI-IMS methods developed in these chapters may prove critical to generating a more comprehensive tissue classification scheme for ovarian metastasis and at the same time provide a measure of standardization for individualized patient treatment.

In order to create a new classification system for ovarian cancer metastasis it would first be necessary to demonstrate the existence of key tryptic peptide markers which discriminate a metastatic ovarian tumor from surrounding tissue. The molecular changes that occur across the margin between invasive tumour and “normal” stroma could then be exploited to identify metastatic tumours of different composition within a training set of patients. Depending on the classification desired the training set could then be segregated based upon molecular markers that correlate to tumour grade and/or response to treatment. In this context, key tryptic peptide markers that correlate to patient prognosis in advanced stage ovarian cancer could be identified. Furthermore, it may be possible to identify tryptic peptide(s) that correlate to patient response during adjuvant chemotherapy. The implication is that the most effective treatment could be given to a patient immediately, thereby providing much needed standardization for treatment of ovarian malignancies.

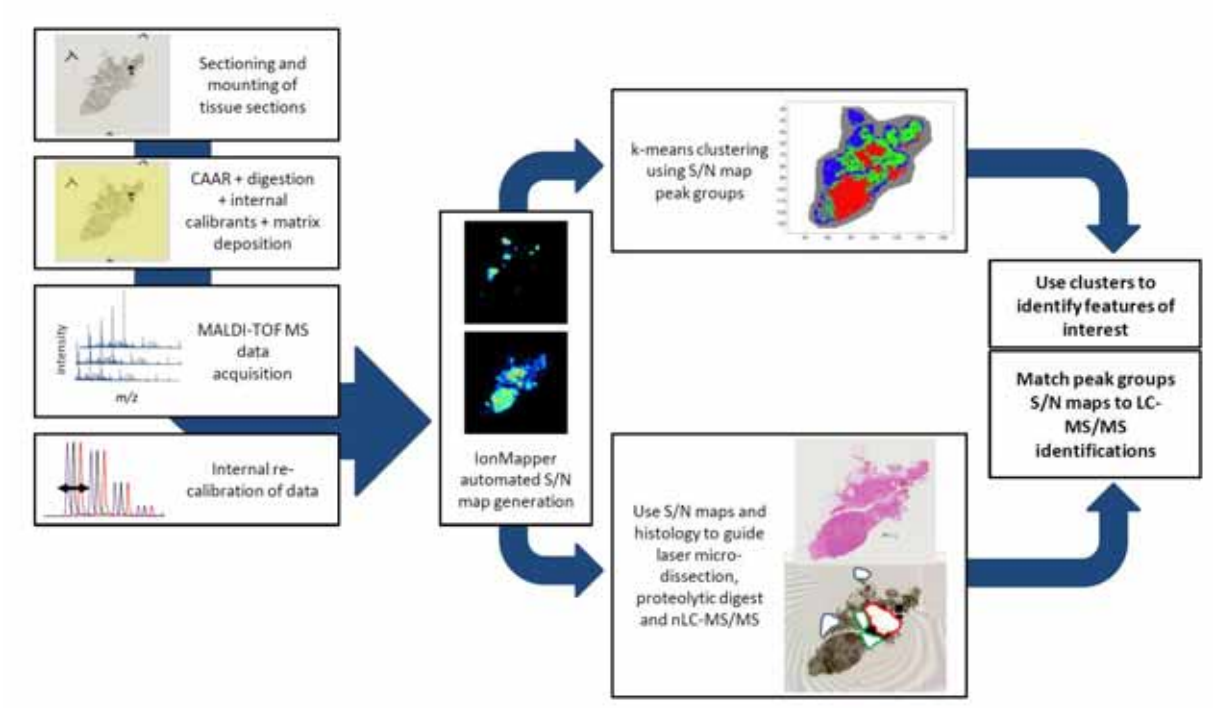



**Figure 7.1.** Comprehensive outline of methods applied for characterizing peritoneal implant tumours from ovarian epithelial carcinoma.

As a result, the MALDI-IMS analysis methods developed in **chapters 5** and **6** were used to identify tryptic peptides markers which could identify the metastatic tumor in sections from seven patients with FIGO stage III serous ovarian carcinoma. A CAAR tryptic peptide MALDI-IMS workflow was applied to FFPE tissue sections coupled with a set of internal calibrant masses (see **Figure 7.1**). The acquired MALDI-IMS data was processed by smoothing and baseline subtraction prior to export of spectral peak lists. Using the IonMapper tool (also developed in **chapter 6**), peaks from the thousands of peak lists were grouped within a mass tolerance of 0.2 Da. An AWM *m/z* was calculated for each group and peaks in each group were plotted as S/N maps. The peak groups were subsequently used for a k-means clustering analysis to identify MS spectra containing similar peak groups. Cluster membership of the spectra was then used to sum the intensity of peak groups for all spectra belonging to a given cluster. The percentage feature intensity in a cluster was used to identify peak groups of interest, which were subsequently identified by using LCM and tryptic digestion of FFPE tissue (cancer, stroma and adipose tissue) coupled to LC-MS/MS identification. These data were used to characterize the tryptic peptide

profile of cancer and adjacent stroma in a subset of the patients. This final aim represented a key first step in characterizing ovarian cancer metastasis using spatial proteomics.

## 7.2 Results and Discussion

### 7.2.1 Tryptic peptide MALDI-IMS analysis of ovarian cancer metastases

A set of seven metastatic peritoneal tumours from high grade FIGO stage III serous ovarian carcinoma obtained during surgery at the Royal Adelaide Hospital were formalin-fixed and paraffin-embedded. The tumours were sectioned at a thickness of six µm and water bath-mounted (water temperature set at 39ºC) onto ITO slides. Multiple slides were prepared so that serial sections from the same patient were analyzed a minimum of three times. The section mounting was also staggered so that the replicate sections occupied each of the three possible ITO slide locations (see **Figure 7.2**). Sample preparation quality typically decreases with increasing distance from the ImagePrep spray generator. As such, the sections were mounted in this way to ensure that inter-patient differences were not due to differences in preparation quality at the different locations on a slide.

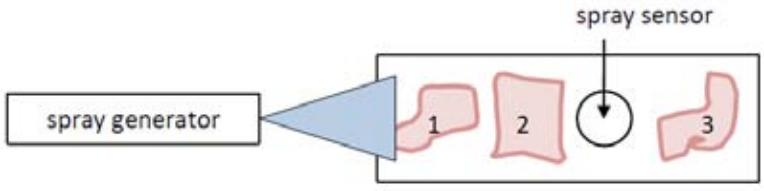



**Figure 7.2.** Schematic of the slide layout relative to the ImagePrep spray sensor and spray generator.

Tissue sections were prepared for MALDI-IMS using the methods described in **chapter 6**. Briefly, paraffin was stripped from all sections, prior to rinses in ethanol and treatment with a modified CAAR

protocol (see **section 2.5.3**). All sections were digested by ImagePrep deposition of trypsin gold in 25 mM $NH_4HCO_3$ and incubation at 37°C for two hours in a humid container. An internal calibrant mixture of four peptides (see **section 2.5.5** and **chapter 6**) was deposited using the ImagePrep, followed by CHCA matrix (7 mg/mL in 50% ACN and 0.2% TFA). MS analysis was performed on an ultrafleXtreme MALDI-TOF/TOF instrument in reflectron positive ion mode. The order of slide analysis was randomized to control for differences in the day to day quality of sample preparations and MS measurement. Following MS analysis, the matrix was removed, the tissue sections were H&E stained and sent to Dr. Fergus Whitehead (Adelaide Pathology Partners, Mile End, South Australia) for annotation and grading.

Two patients were determined to be unsuitable and were excluded from the analysis. Patient 181 sections were excluded prior to data acquisition because the sections consistently fragmented during CAAR treatment. Patient 372 sections were excluded, following data acquisition and H&E staining, due to necrotic tissue dominating the non-adipose regions of the sections. For the remaining five patients, 18 sections (n = 3 for patients 363 and 544, n = 4 for patients 44, 173 and 540) were included in the final analysis.

### 7.2.2 Linking peak groups to histology using IonMapper generated signal to noise maps

Analysis of MALDI-IMS data was initially performed in flexImaging (V3.0). The flexImaging software combines all spectra into one representative spectrum refered to as a sum spectrum. Mass filters of a defined width (*m/z*) are applied to this sum spectrum and the highest intensity value within this filter, for all the spectra in the data set, are presented as an ion intensity map. The sum of the data collected for the five patients indicated good overlap between generated ion intensity maps and subsequent H&E stains of the same sections. High resolution H&E stains for one section from each patient are shown in Supplementary Figure 7.1 to 7.5. **Figure 7.3** shows the H&E stains and a three ion intensity map overlay for *m/z* 1390.7 (red), 1495.7 (blue) and 1962.0 (green) for all five patients. The maps were

produced by manual selection of mass filters (width; 0.1 Da) in the flexImaging sum spectra for patients 44 (**7.3a**), 173 (**7.3b**), 363 (**7.3c**), 540 (**7.3d**) and 544 (**7.3e**). Importantly, the figure shows that even manual selection of peaks of interest discriminated the tissue region marked by a pathologist as high grade cancer (red in **Figure 7.3**, see also Supplementary Figure 7.6) as well as surrounding tissues, including the peritoneal stroma (green) and adipose tissue (blue).

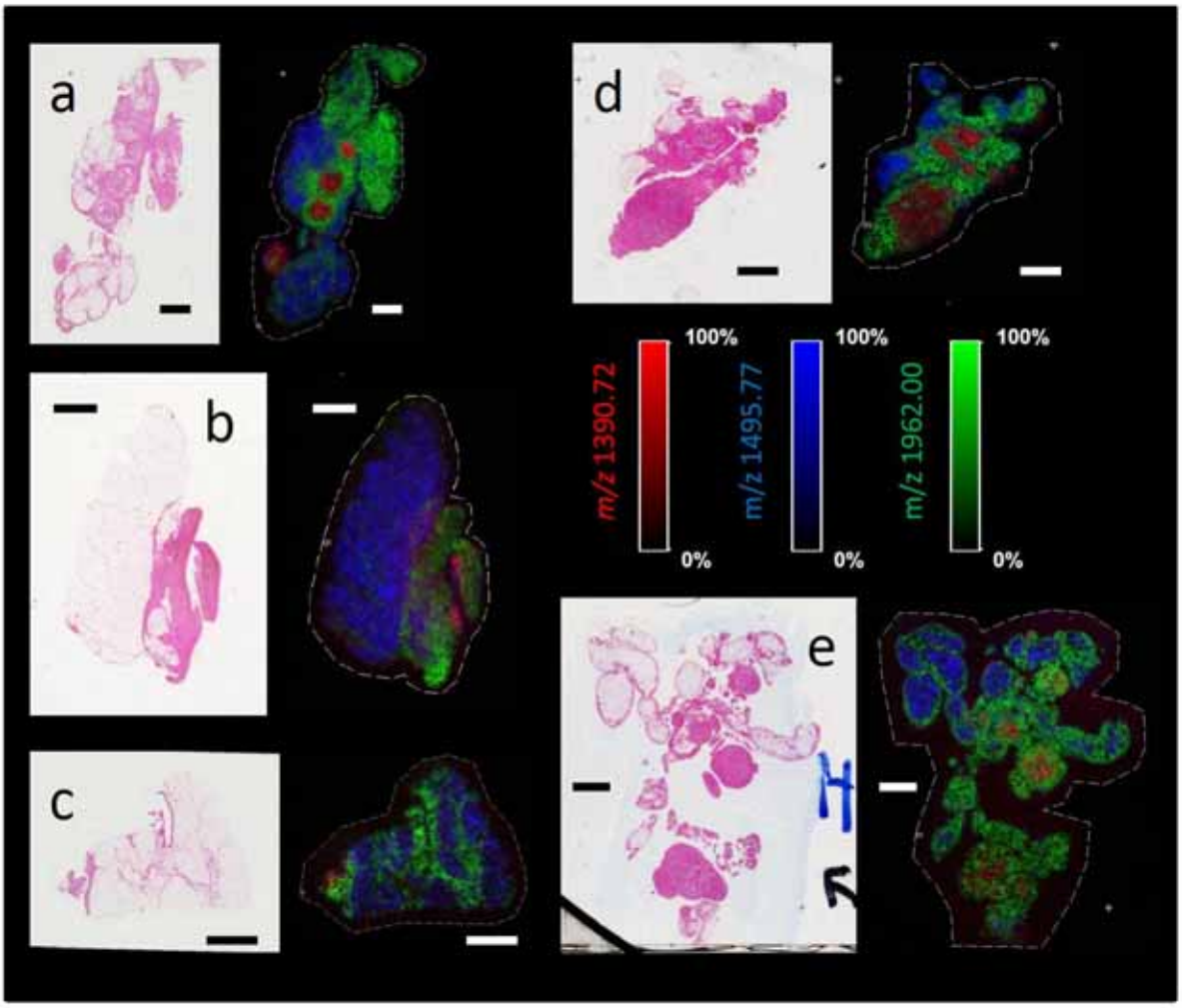



**Figure 7.3.** Sections from five patients with high grade peritoneal implants (serous ovarian carcinoma) were processed by antigen retrieval, tryptic digestion, internal calibrant and matrix deposition, followed by MALDI-TOF MS analysis on an ultrafleXtreme instrument (reflectron positive ion mode). Matrix was eluted from the sections prior to haematoxylin and eosin (H&E) staining. The H&E stains for patient 44 (**a**), 173 (**b**), 363 (**c**), 540 (**d**) and 544 (**e**) are shown. Adjacent to each H&E stained image is an overlay of the ion intensity maps for m/z 1390.7 (red), 1495.7 (blue), and 1962.0 (green). Ion intensity maps were produced in flexImaging (V3.0) using spectra normalized to total ion current. Scale bars are 2 mm and intensity scales for each ion are included.

As described in **chapter 6**, there were advantages to working with peak lists rather than complete profile spectra. For example, data size could be reduced by at least two orders of magnitude. Furthermore, because mono-isotopic peaks were always picked for inclusion in the peak lists any peptides that overlapped in a sum spectrum, but existed in different tissue locations could be distinguished.

As such, peak lists were generated from each data set (n = 18) using a customized flexAnalysis script. These peak lists were subsequently processed by the IonMapper software tool developed in **chapter 6**. This tool grouped all peaks in a data set together based on a set peak grouping tolerance (0.2 Da). An AWM *m/z* was calculated for all peak groups and the S/N of the peaks within each group were plotted as a S/N map. The S/N map was effectively the in-house replacement for the ion intensity map produced by flexImaging. For this project, in order for a S/N map to be plotted, the peaks within each group were required to exist in a minimum of 50 spectra (minimum peak appearance value). A minimum peak appearance of 50 was appropriate because 50 spectra always represented less than 2% of the total number of spectra acquired for the sections used in this project. This meant that the chance of removing an important feature with this setting was minimal.

As expected, the total number of S/N maps created for each section was variable (Supplementary Figure 7.7). Generally, those sections occupying the sample preparation positions closest to the ImagePrep spray generator yielded more S/N maps, indicating denser and more homogeneous reagent coating in this region. This was confirmed by manual test MS acquisitions prior to starting each automated MALDI-IMS experiment (data not shown). In terms of the total number of S/N maps generated there was a marked increase for the larger sections. This indicated that the minimum peak appearance value may have been too high for the smaller sections (e.g. patient 363 or 540). However, closer inspection of the S/N maps generated for all sections suggested that the larger numbers were included due to diffusely detected “noise.” This noise either represents true noise, such as picked matrix peaks, or true peak groups which were in such low abundance that they were only detected in a small fraction of MS spectra. These data suggested that for any one data set more than 400 individual S/N maps could be generated with most sections generating more than 1000 S/N maps. Those sections

which generated the smallest number of S/N maps were typically located at position three on the slide, which was furthest from the ImagePrep spray generator during sample preparation.

To fully demonstrate the efficacy of automated S/N map generation, **Figure 7.4** outlines a comparison between flexImaging generated ion intensity maps and IonMapper generated S/N maps for three different highly localized peak groups. **Figures 7.4a** and **7.4b** show enlarged images of the H&E stains for the patient 44 tissue presented in **Figure 7.3a**. These H&E stains are annotated with the three major tissue regions found in the sections analyzed, these being cancer (red outline), peritoneal stroma (green outline) and adipose tissue (blue outline). Three maps generated either by IonMapper (S/N maps) or flexImaging (ion intensity maps, normalized to total ion current) were also included. S/N maps for peak groups 1390.7 (**7.4c**), 1495.8 (**7.4e**), and 1962.0 (**7.4g**) demonstrated annotation of the high grade cancer, adipose tissue and peritoneal stroma respectively. The corresponding ion intensity maps for *m/z* 1390.7, 1495.8 and 1962.0 are shown in **Figure 7.4d**, **7.4f** and **7.4h** respectively. These data demonstrated that the IonMapper tool generated virtually identical, albeit less noisy, distributions to flexImaging. The noise difference was the result of IonMapper using peak lists rather than a maximum MS intensity within a set mass window, as flexImaging does. The most obvious result was the removal of background (i.e. all features below the designated S/N used for processing the original spectra). For example, the low level signals for peak group 1390.7 were removed for the S/N map in **Figure 7.4c** as compared to the ion intensity map in **Figure 7.4d**, which included all signals regardless of whether they represented a peptide peak, raised baseline or chemical noise. Notably, key S/N maps for the peak groups presented in **Figure 7.4** were reproducible across replicate sections from the same patient tissue block (see Supplementary Figure 7.8). This was important as it suggested that the sample preparation developed in previous chapters and applied here was robust and did not need further optimization.

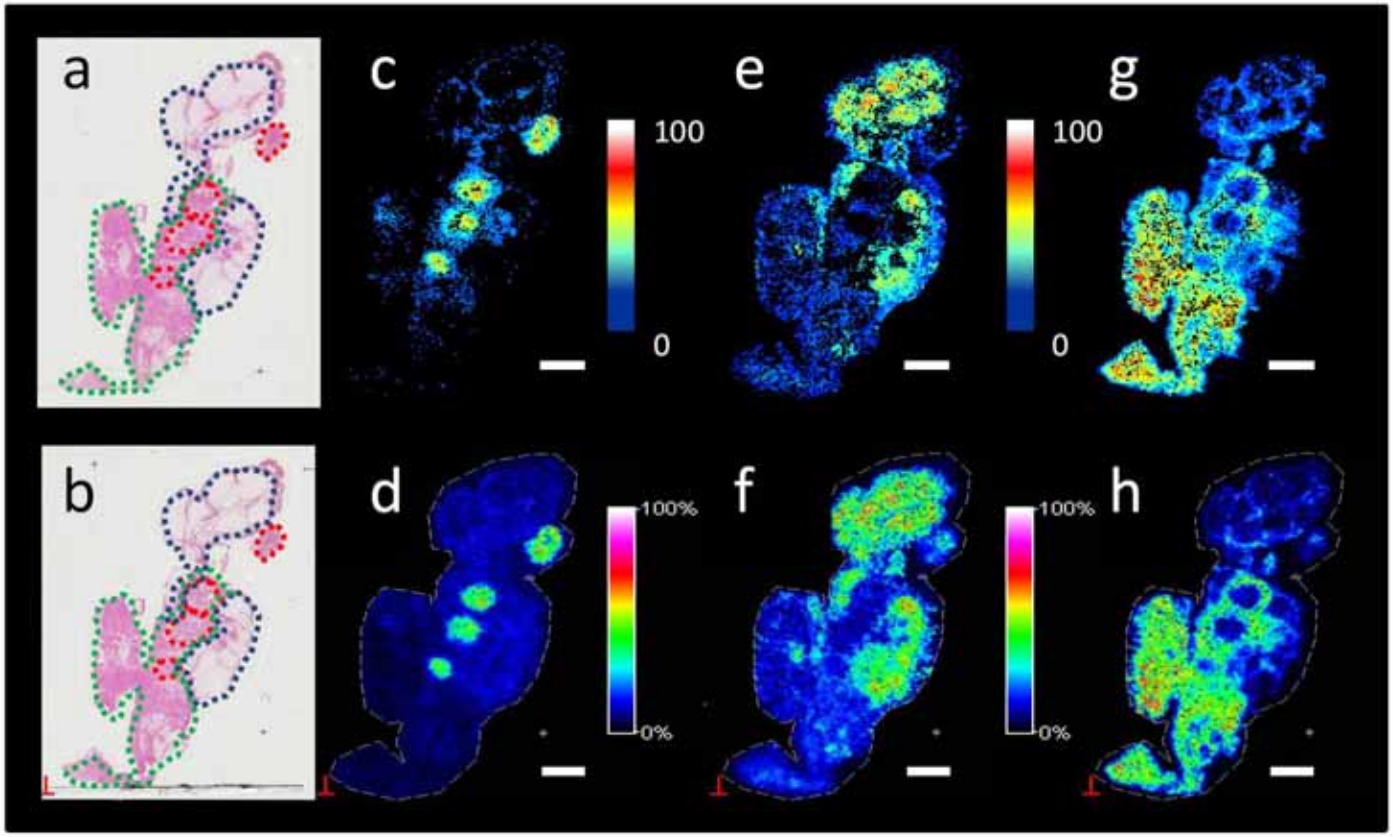



**Figure 7.4.** Tryptic peptide MALDI-IMS data was performed on a section from patient 44. The haematoxylin and eosin (H&E) stain for the section is included in panels **a** and **b**. The H&E stain is annotated with dotted lines to show the major tissue regions, high grade cancer (red), peritoneal stroma (green) and adipose tissue (blue). MS data acquired from the tissue was used to generate IonMapper signal to noise maps for peak groups 1390.7 (cancer), 1495.7 (adipose tissue) and 1962.0 (peritoneal stroma), which are shown in **c**, **e** and **g** respectively. The corresponding flexImaging ion intensity maps (normalized to total ion current) for approximate m/z 1390.7 (**d** - cancer), 1495.7 (**f** - adipose tissue), and 1962.0 (**h** – peritoneal stroma) are also shown. Scale bars are 2 mm and intensity scales are included.

### 7.2.3 Peptide identification and the application of k-means clustering

In order to provide peptide identities to match to peak groups of interest, LCM was used to isolate tissue from sections adjacent to those used for MALDI-IMS. Three regions were isolated, including the area marked by a pathologist as cancer (**Figure 7.5**, red outlines). The other two regions were the peritoneal stroma (**Figure 7.5**, green outline) and adipose tissue (**Figure 7.5**, blue outline). The regions isolated for patients 44, 173, 363, 540 and 544 are outlined using the appropriate colour in **Figure 7.5a-e** respectively. The tissue samples (total n = 15) were independently treated with CAAR and digested in-solution using trypsin. Conditions for these treatments were similar to those used for the MALDI-IMS experiments apart from the in-solution digest; a requirement given that the tissue samples could not easily be re-mounted on an ITO slide and digested using the ImagePrep station. The tryptic peptides produced were purified using C-18 filter centrifuge columns and subjected to LC-MS/MS identification on a nLC-LTQ-Orbitrap system (Thermo-Fisher Scientific). Subsequent MS/MS data sets were searched against the Human (*Homo sapien*) SwissProt database using the MASCOT search engine. Supplementary Figure 7.9 shows the false discovery rates and number of peptide hits above the identity and homology thresholds for each of the 15 samples at a MASCOT significance threshold of 0.05. Importantly, the false discovery rate was not higher than 1.6% for any of the searches, which indicated that the acquired data was high quality. Peak groups for which a S/N map was generated (for each patient) were then matched within a ± 0.2 Da mass tolerance to the peptides identified by LC-MS/MS with an ion score equal to or higher than their corresponding MASCOT identity threshold.

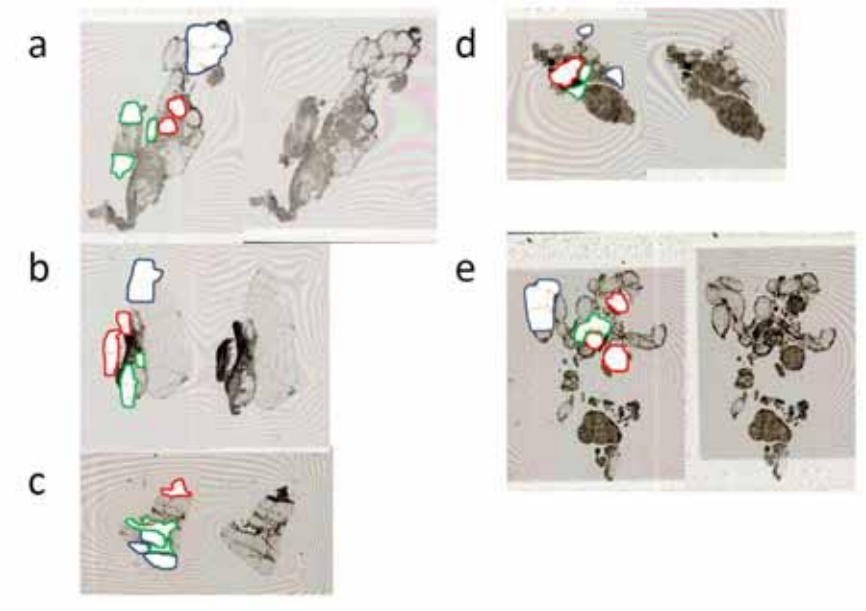



**Figure 7.5.** Formalin-fixed sections from patients 44 (**a**), 173 (**b**), 363 (**c**), 540 (**d**) and 544 (**e**) were mounted onto PEN membrane slides, paraffin was stripped and sections were washed in 100% EtOH. Areas of interest, adipose tissue (blue), peritoneal stroma (green) and cancer (red) were excised by laser capture micro-dissection (LCM) as shown by the coloured outlines. An image of a consecutive section from each patient block is included for comparison. Excised material was treated by citric acid antigen retrieval and digested in-solution using trypsin. Tryptic peptides were purified using C-18 spin columns and identified using nLC-LTQ-Orbitrap MS/MS and MASCOT database searches.

It was considered less than ideal to manually select peptide peak groups from these match lists which appeared to only exist in the cancer tissue. Selecting peak groups of interest in this way would be biased depending on the individual performing the analysis. As a result, the selection of peak groups of interest was based upon the clustering of spectra with similar peak group features. The use of available proprietary software such as ClinProTools (Bruker Daltonics) was avoided, as this software was not written to use profile spectra with isotopic resolution. Consequently, this software requires manual picking of the mono-isotopic peptide peaks prior to hierarchical clustering **[65, 125]**. This was also less than ideal as it required manual selection of mono-isotopic peptide peaks from potentially complex profile spectra where multiple peptide isotope profiles could be overlapping. A k-means clustering approach utilizing Matlab was selected for this project to make the number of candidates more manageable by focusing on those peak groups which make the greatest contribution to a user defined number of clusters. The smaller list of candidates could then be compared to the list of tentative

matches described above. The k-means clustering was performed by A/Prof. Inge Koch (School of Mathematics, The University of Adelaide, Australia) in collaboration with Prof. Steve Marron (Statistics and Operations Research, The University of North Carolina – Chapel Hill). A comprehensive discussion of the methods used is presented elsewhere **[80]**. Briefly, k-means clustering partitions a data set, in this case spectra, into a defined number of clusters. The key idea was to split the available MS data into the defined number of clusters such that the variability within a cluster was as small as possible and the distance between clusters was as large as possible.

In order to apply the k-means clustering to the peptide MALDI-IMS data, the peak groups used to generate S/N maps (for one section from each patient using the minimum peak appearance filter described previously) were converted into two dimensional feature vector data files (ASCII text). These files contained the spectral coordinates (x, y) as one dimension and the peak group intensities as the other. All spectra were thus represented in this file along with the intensities of the peak groups. In this context, a k-means clustering assigned the raw data vectors within this feature vector file to one of a defined number of clusters.

Prior to analyzing all five patient sections, patient 540 was used to troubleshoot the clustering approach by comparing k-values (i.e. number of clusters) of three, four and five. It was found that a k value of four gave the best clustering result relative to the histology of the sections and the corresponding pathologist annotations. Furthermore, it was found that transforming the original non-zero intensity values in the feature vector file to one and leaving all zero values as zero improved subsequent clustering results (data not shown). Although the conversion to binary was not typical for a clustering approach, the improvement for the binary transformation led to adoption of this file format for the exploratory tissue characterization described in this chapter. Koch clustered these binary data into four clusters using the Matlab k-means algorithm with the cosine distance and the Hamming distance, rather than the default (Euclidean) distance. The cosine distance calculates the cosine of the angle between two vectors, and vectors with small angles between them are regarded as ‘close’, and form a cluster. For the Hamming distance, two vectors are ‘close’ and become part of the same cluster, when the proportion of peak

groups values at which they differ is small. Thus, if two vectors have non-zero counts for the same peak group values, and zero-counts at the same peak group values, their Hamming distance is zero.

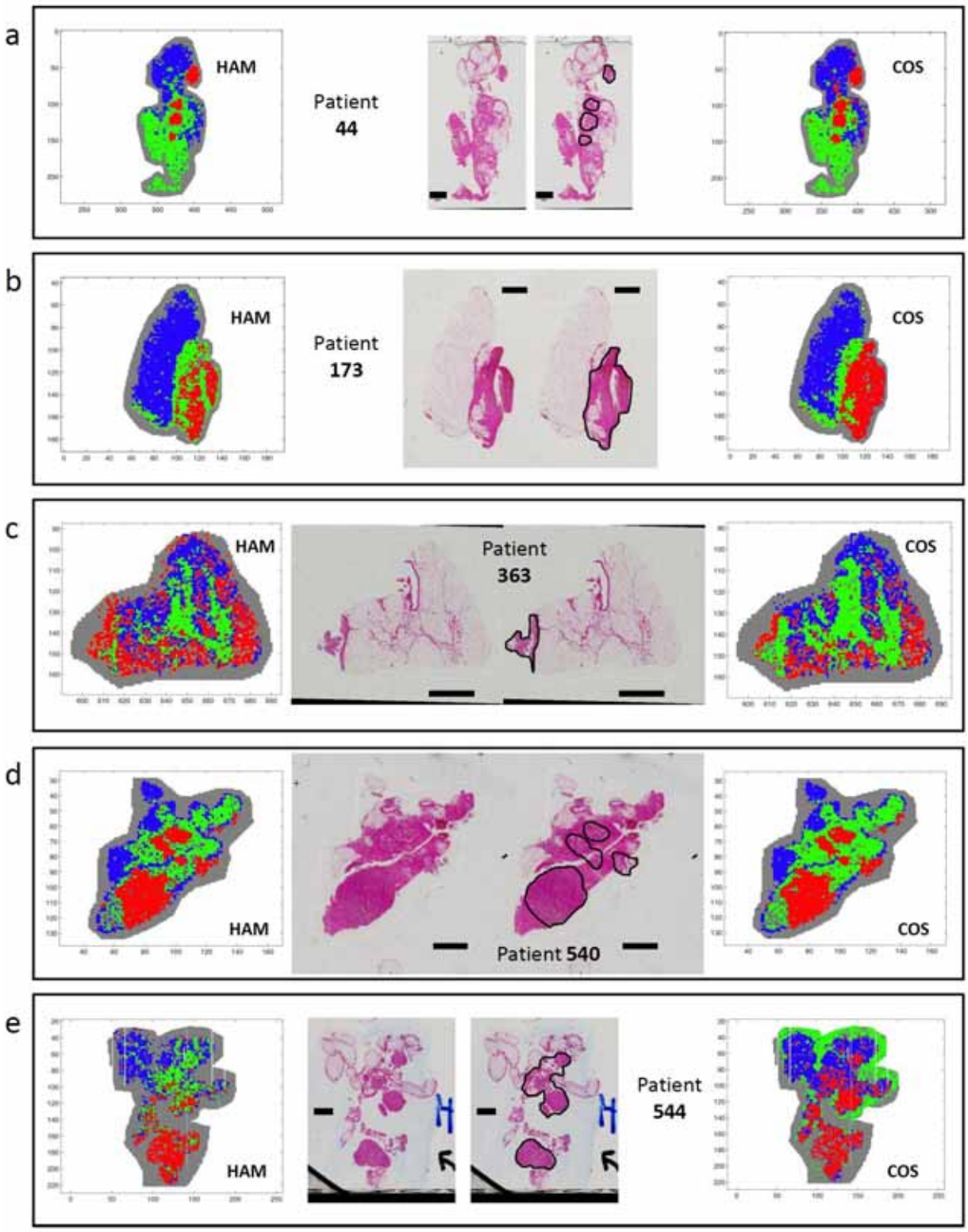



**Figure 7.6.** Plots showing the cluster membership (Green – 1, Gray – 2, Blue – 3, Red – 4) for spectra clustered using the Hamming (far left) and cosine (far right) distance. Plots are included for patients 44 (**a**), 173 (**b**), 363 (**c**), 540 (**d**) and 544 (**e**). Haematoxylin and eosin stains of the same section are provided in the centre along with a black outline indicating the approximate cancer region (as annotated by a pathologist). Scale bars are 2 mm.

Feature vectors from each patient were analyzed using both clustering distances with a k-value equal to four and each spectrum was subsequently assigned a cluster membership. The membership of each spectrum was plotted as a cluster map showing the high grade cancer as red, peritoneal stroma as green, adipose tissue as blue and off-tissue spectra as grey (**Figure 7.6a-e**). All the cluster maps in **Figure 7.6** are accompanied by H&E stains annotated with the high grade cancer region (as annotated by a pathologist). The following section outlines the results of matching the cluster maps to histology for all five patients.

### 7.2.4 Matching clusters to histology

As already described, the MS spectra cluster memberships derived from the k-means analyses were used to illustrate the existence of four distinct spectral groups; cancer, peritoneal stroma and adipose tissue as well as spectra collected off-tissue. Because the final aim of the analysis was the characterization of the invasive tumour and its interface with the peritoneal stroma these two clusters were analyzed further to select peak groups of importance to each region.

A comparison of the two clustering approaches indicated virtually identical cluster assignments for cosine and Hamming distances applied to patients 44 (**Figure 7.6a**) and 540 (**Figure 7.6d**). The clusters for both sections clearly discriminated cancer (red) from peritoneal stroma (green) and adipose tissue (blue).

The Hamming distance appeared to provide clearer discrimination between the different tissue regions when applied to patient 173 (**Figure 7.7b**). However, the border region between stroma and cancer was not as clearly defined for this section as for patients 44 (**7.6a**) and 540 (**7.6d**); a fact demonstrated by the relatively conservative pathologists annotation for patient 173 in **Figure 7.6b** (see also Supplementary Figure 7.6). Because of the lack of a definite tumour border, a higher resolution scan of

the H&E stain (**7.7a**) was compared to the Hamming (**7.7b**) and cosine distance (**7.7c**) cluster maps. It was clear from this comparison that the intimate association of cancer and stroma in patient 173 made selection of a preferable clustering method difficult. As a result, both the Hamming and cosine distance were used to identify peak groups of interest, which were subsequently compared to the results for patients 44 and 540.

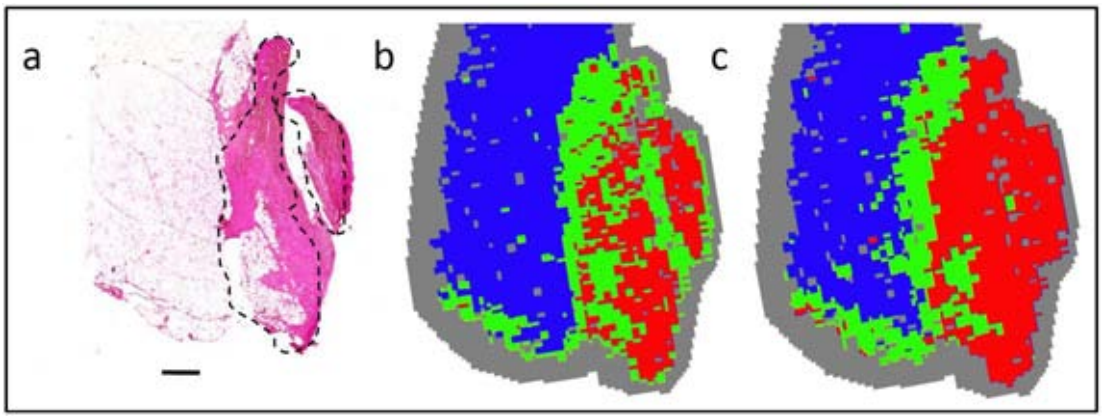



**Figure 7.7.** Haematoxylin and eosin (H&E) stain (**a**) for patient 173 along with the cluster maps for the feature vectors clustered using the Hamming distance (**b**) and cosine distance (**c**). Black outlines in panel **a** show the approximate outline of the cancerous red cluster in panel **b**. Scale bar is 1 mm.

Cluster analysis of the patient 363 data split the spectra associated with adipose tissue into two groups (both Hamming and cosine distance in **Figure 7.6c**) and failed to discriminate the cancer. It was determined that the most likely cause was a failure to pick a sufficient number of cancer specific peaks to overcome the minimum peak appearance of 50. This was unexpected given that the cancerous region contained more than 90 MS spectra. Because the clustering failed to discriminate the cancer, patient 363 was not included for further analysis to identify peak features important to the individual clusters.

Patient 544 presented an interesting case due to the small amount of stroma present in the tissue section (**Figure 7.6e**). This discrepancy appeared to confound the clustering using the Hamming

distance, separating the cancerous spectra into two different clusters (red and green in **Figure 7.6e**). This indicated that different peak groups may have existed in these tissue regions. After examination of the S/N maps localized to the red cluster of the Hamming cluster map it was found that the cancer specific peak groups were of low S/N. Low S/N will result in fewer picked peaks for a single peptide species across a data set. In an experimental setting, if this peptide was important to the cancer, the reduced peak picking efficiency would result in the peptide not being picked in all the spectra where it existed. This would subsequently reduce the robustness of the clustering result in terms of assigning the correct region as cancerous. This was most likely the case for patient 363 described above. However, this was considered unlikely for patient 544 given that a pattern that closely matched to histology was produced when applying the cosine distance clustering to the patient 544 feature vector (**Figure 7.6e**). **Figure 7.6e** demonstrates this result, where the red and blue cluster regions distinguished cancerous and adipose regions respectively, while the off-tissue region was split into the remaining two clusters; grey and green. Importantly, the cosine distance result agreed with the apparent absence of significant amounts of stroma. As a result of these findings, the cancer spectra as defined by both the Hamming and cosine distance clustering were used to compare peak groups of interest from all four patients.

### 7.2.5 Cancer peak group features – Hamming distance clustering results

To find peak groups critical to the distinction between cancer and stroma the intensity of a peak group in the original feature vector was summed across all the spectra belonging to each cluster. This was repeated for all the peak groups in the vector, converting the thousands of spectral coordinate columns in the feature vector to four cluster-specific intensity columns. The four intensities for each peak group were normalized to the number of spectra in their respective cluster and these normalized intensity values were divided by the summed peak group intensity across all four clusters. The end result was a percentage intensity contribution (PIC) to each cluster for all the peak groups. The percentage of

spectra (POS) in each cluster containing the peak group was also calculated. This process was first completed for the cancerous cluster as indicated by the Hamming distance for all four patients.

The complete list of peak groups used for clustering was filtered to remove those with PIC values less than 80% and POS values less than 20%. The PIC value of 80% was selected as an arbitrary minimum to ensure that only those peak groups with a significant contribution to the cancer cluster were considered for subsequent identification. The POS value of 20% was selected to ensure that peak groups were not excluded from analysis due to poor peak picking efficiency for cancer specific peak groups, which were typically present at a low S/N. Future work on the processing protocols for peptide MALDI-IMS data will need to improve upon this, for example, by employing other peak picking algorithms. However, for the current analysis the low (20%) POS filter was used to compensate for any problems with peak picking. In addition to the PIC and POS cut offs those peak groups which only existed in one patient were removed. Fourteen peak groups were included in the final list of peak groups of interest (**Table 7.1**).

To make collation of data easier, peak groups were annotated with a 0.1 Da mass error (100 ppm) for sorting and analysis of data. To allow comparison of the variability in peak group AWM *m/z* for the different patients **Table 7.1** also includes a 0.01 Da error margin (10 ppm). Knowing the AWM *m/z* to two decimal places was important for subsequent confirmation of peptide identifications for peak groups of interest.

**Table 7.1.** Patient number, peak group abundance weight mean (AWM) m/z, percentage intensity contribution to the cancer cluster (PIC) and percentage of spectra in the cancer cluster containing each peak group of interest (POS) are shown for the cancer region clustered using the Hamming distance for patients 44, 173, 540 and 544. All features were found in two or more patients and had PIC values greater than 80%. Peak groups in bold were found to have a PIC greater than 90%.

| Patient | Peak group AWM *m/z* | | PIC | POS |
|---|---|---|---|---|
| | ± 0.1 *m/z* | ± 0.01 *m/z* | | |
| **44** | **1406.7** | **1406.70** | **94.5** | **97.1** |
| 540 | 1406.7 | 1406.70 | 82.5 | 95.7 |
| 44 | 1413.7 | 1413.74 | 85.8 | 47.9 |
| 540 | 1413.7 | 1413.75 | 87.1 | 67.3 |
| 44 | 1543.8 | 1543.80 | 88.4 | 22.2 |
| **540** | **1543.8** | **1543.85** | **90.9** | **25.9** |
| **44** | **1589.8** | **1589.85** | **91.5** | **83.8** |
| **540** | **1589.9** | **1589.86** | **94.6** | **66.1** |
| **44** | **1609.8** | **1609.79** | **95.5** | **93.8** |
| 540 | 1609.8 | 1609.78 | 80.9 | 84.3 |
| 44 | 1703.9 | 1703.87 | 85.7 | 59.9 |
| **540** | **1703.9** | **1703.90** | **92.5** | **71.6** |
| **44** | **1739.9** | **1739.91** | **93.1** | **78.3** |
| 540 | 1739.9 | 1739.92 | 89.1 | 78.0 |
| 44 | 1762.9 | 1762.85 | 88.9 | 73.2 |
| 540 | 1762.9 | 1762.86 | 81.4 | 56.4 |
| **44** | **1936.0** | **1936.04** | **95.1** | **95.0** |
| 540 | 1936 | 1936.03 | 80.9 | 30.6 |
| **44** | **1997.9** | **1997.93** | **94.1** | **96.5** |
| 540 | 1997.9 | 1997.94 | 83.4 | 56.1 |
| 44 | 2171.1 | 2171.13 | 87.2 | 79.1 |
| **540** | **2171.1** | **2171.13** | **91.8** | **63.3** |
| **44** | **2189.9** | **2189.94** | **93.9** | **65.6** |
| 540 | 2189.9 | 2189.92 | 88.3 | 88.3 |
| 544 | 2189.9 | 2189.94 | 68.3 | 26.5 |
| **44** | **2354.1** | **2354.13** | **95.4** | **93.4** |
| 540 | 2354.1 | 2354.14 | 88.8 | 40.0 |
| 44 | 2854.4 | 2854.41 | 89.8 | 99.4 |
| 540 | 2854.4 | 2854.41 | 81.6 | 71.2 |

The isotopic pattern of the peptides was also examined at the individual profile spectrum level to confirm picking of the mono-isotopic peptide peak for each peak group. In most cases, the level of potential overlap, as indicated by assessment of the ratio of the mono-isotopic ($^{12}C$) and $^{13}C$ peaks, was minimal. However, some peak groups, including 1413.7, 1543.8, 1589.8 and 1703.9 displayed isotopic profiles indicative of overlap. As a result, it was not possible to unambiguously map the peak group distribution using either the S/N map or ion intensity map for these peak groups. The reason for this was that both maps of peak group distribution ultimately relied on intensity to generate their visual output, which was susceptible to contribution by isotopic profiles belonging to adjacent peak groups.

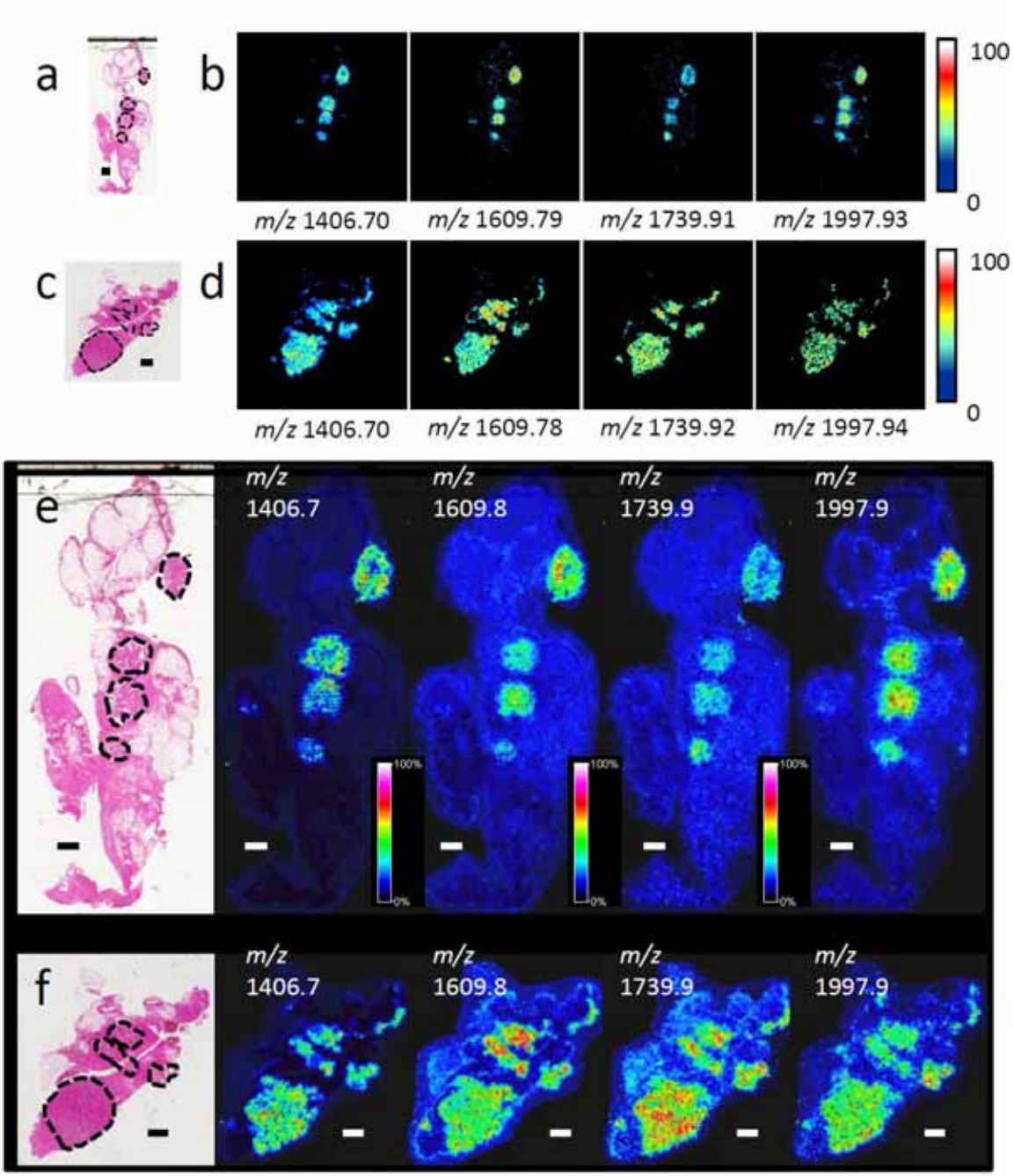



**Figure 7.8.** Haematoxylin and eosin stains are shown for patients 44 (**a**) and 540 (**c**) along with annotations indicating the high grade cancer (black dotted line). Signal to noise maps generated for patients 44 (**b**) and 540 (**d**) are shown for peak groups 1406.7, 1609.8, 1739.9 and 1997.9. Ion intensity maps generated in flexImaging are also shown for the same peak groups (0.1 Da mass filters) in **e** and **f** for patients 44 and 540 respectively. Scale bars are 1 mm and intensity scales are included. Ion intensity maps were normalized to noise.

**Figure 7.8** presents the S/N maps for four cancer specific peak groups from **Table 7.1**, which showed minimal peptide isotope profile overlap with adjacent masses. H&E stains for patients 44 and 540, along with annotations indicating the cancerous regions are shown in **Figure 7.8a** and **7.8c** respectively. The example S/N maps for peak groups 1406.7, 1609.8, 1739.9 and 1997.9 are shown in **Figure 7.8b** for patient 44 and **7.8d** for patient 540. For both patients the S/N maps for these peak groups showed convincing, highly localized distributions, which were subsequently confirmed by examination of the ion intensity maps generated in flexImaging for patient 44 (**Figure 7.8e**) and 540 (**Figure 7.8f**).

Unfortunately, no peak groups from patient 173 and only one peak group from patient 544 (peak group 2189.9) matched the criteria required for inclusion in **Table 7.1**. These data indicated that the contribution of peak groups was not reproducible for all the patients. There were several possible explanations for the lack of common peak groups across the patient data set.

Firstly, the sample preparation (tryptic digestion or matrix deposition) could have introduced variability in the tryptic peptide complement generated for patients 173 and 544. This was considered unlikely because both patient 544 and 173 were analyzed on slides where sections from patient 44 and 540 were immediately adjacent. Furthermore, patients 44 and 540 were analyzed on different slides. As such, any sample preparation issues should also have caused a discrepancy between patients 44 and 540. The mounting method used and the high level of overlap between patients 44 and 540 thus ruled out sample preparation as a cause of poor reproducibility.

The second and most likely explanation was possible tissue heterogeneity, caused either by the high variability in molecular composition of individual malignant tumours or intra-tumour variability in tryptic peptide detection. There were also other potential contributors to heterogeneity, including age, which was ruled out because the four patients included in the final analysis were all post-menopausal women of similar age. The tumour origin was also important. However, as indicated by the available patient specific information, the primary tumours for patients 540 and 544 were both classical serous ovarian carcinomas. In addition, the patient 44 primary tumour was also, most likely, a classical serous ovarian

carcinoma. This contention was supported by the molecular similarity of the patient 44 metastatic tumour to patient 540 at the MS level. Finally, although the metastatic tumours arose from the same type of primary tumour, when the metastases were excised they were dissected into more manageable pieces prior to FFPE treatment. Depending on the tissue piece and sectioning plane selected, a different portion of the tumour would be analyzed. This could explain the molecular heterogeneity between patients, as patient 44 and 540 shared morphologically similar regions, as did patient 540 and 544 (see Supplementary Figure 7.10).

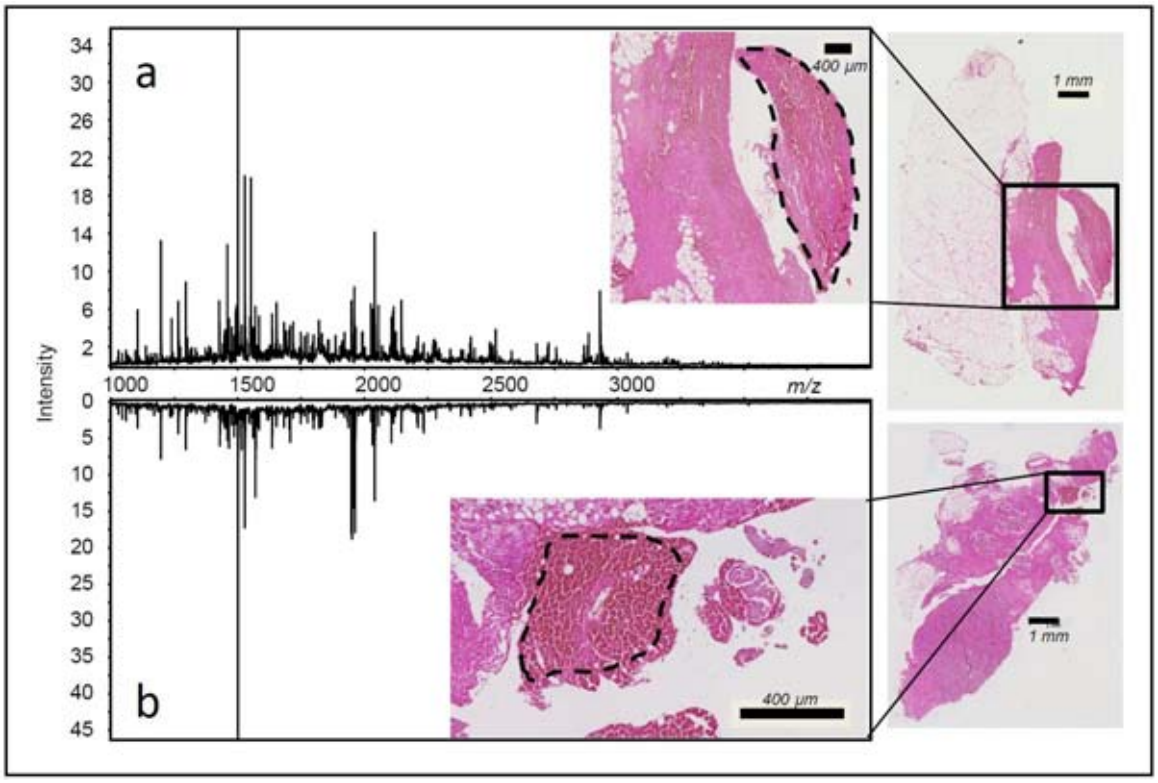



**Figure 7.9.** Sum spectra (flexImaging generated) for necrotic regions in sections from patient 173 (**a**) and 540 (**b**). The black dotted outline in the inset enlarged H&E stains show the location of the spectra used to generate the sum spectra. At the far right are the whole section H&E stains. The enlarged areas used for annotation are indicated on these stains. Scale bars are annotated with distances.

Patient 173, who presented with a serous tubal carcinoma, was the only case identified as fundamentally different by MALDI-IMS analysis. However, the complete lack of peak group overlap with

the other patients could be explained by the larger amount of necrotic tissue present in patient 173. The sections from patient 44 and 540 were largely free of necrosis and this would contribute greatly to any differences observed when trying to isolate common peak groups of interest for patient 173. **Figure 7.9** demonstrates this by comparing the MS sum spectrum from a necrotic region of patient 173 (**Figure 7.9a**) with a sum spectrum from a similar but smaller region in patient 540 (**Figure 7.9b**). Crucially, the MS peptide profiles in the sum spectra were virtually identical, supporting tissue heterogeneity as a cause for the molecular heterogeneity across the patient data set.

The final potential explanation for a lack of reproducibility was believed to be of relevance to patient 544. As mentioned previously, the Hamming distance did not correctly discriminate the cancerous region in the patient 544 section. The implication was that the importance of any common peak groups of interest (across patients) were potentially reduced by erroneous PIC and POS values derived from the spectra identified as cancerous using the Hamming distance. To determine whether the lack of common peak groups across patients was caused by molecular heterogeneity (due to tissue selection) or erroneous clustering, the PIC/POS processing and peak group filtering described above was applied to the cancer spectra from all four patients as determined by the cosine distance clustering.

### 7.2.6 Cancer features – Cosine distance clustering results

When the cosine distance clustering was used to select cancer spectra, application of the PIC/POS processing and filtering methods resulted in 14 peak groups of interest (see **Table 7.2**). Of these peak groups 13, 14 and 3 were discriminatory for cancer in patients 44, 540 and 544 respectively.

**Table 7.2.** Patient number, peak group abundance weight mean (AWM) *m/z*, percentage intensity contribution to the cancer cluster (PIC) and percentage of spectra in the cancer cluster containing each peak group of interest (POS) are shown for the cancer region clustered using the cosine distance for patients 44, 173, 540 and 544. All features were found in two or more patients and had PIC values greater than 80%. Peak groups in bold were found to have a PIC greater than 90%.

| Patient | Peak group AWM *m/z* | | PIC | POS |
|---|---|---|---|---|
| | ± 0.1 *m/z* | ± 0.01 *m/z* | | |
| 44 | 1269.7 | 1269.68 | 84.7 | 51.0 |
| 540 | 1269.7 | 1269.69 | 80.9 | 90.3 |
| **44** | **1406.7** | **1406.70** | **94.2** | **68.1** |
| 540 | 1406.7 | 1406.70 | 83.4 | 97.5 |
| 44 | 1413.7 | 1413.74 | 80.8 | 31.7 |
| 540 | 1413.7 | 1413.75 | 87.1 | 67.9 |
| 44 | 1589.8 | 1589.85 | 88.8 | 56.4 |
| **540** | **1589.9** | **1589.86** | **93.5** | **66.5** |
| **44** | **1609.8** | **1609.79** | **94.0** | **61.5** |
| 540 | 1609.8 | 1609.78 | 82.6 | 85.2 |
| **44** | **1739.9** | **1739.91** | **90.5** | **48.7** |
| 540 | 1739.9 | 1739.92 | 89.7 | 81.0 |
| 44 | 1762.9 | 1762.85 | 83.9 | 52.1 |
| 540 | 1762.9 | 1762.86 | 82.5 | 56.7 |
| **44** | **1936** | **1936.04** | **93.8** | **61.5** |
| 540 | 1936 | 1936.03 | 84.0 | 31.9 |
| **44** | **1997.9** | **1997.93** | **93.0** | **71.8** |
| 540 | 1997.9 | 1997.94 | 84.8 | 56.1 |
| **544** | **1997.9** | **1997.90** | **93.3** | **20.6** |
| 44 | 2171.1 | 2171.13 | 83.6 | 53.5 |
| **540** | **2171.1** | **2171.13** | **91.9** | **64.1** |
| **44** | **2189.9** | **2189.94** | **93.8** | **52.3** |
| 540 | 2189.9 | 2189.92 | 88.5 | 88.0 |
| 544 | 2189.9 | 2189.94 | 87.5 | 27.2 |
| **44** | **2354.1** | **2354.13** | **95.0** | **65.9** |
| 540 | 2354.1 | 2354.14 | 88.6 | 40.5 |
| 540 | 2429.3 | 2429.33 | 80.0 | 55.3 |
| 544 | 2429.4 | 2429.37 | 84.3 | 22.9 |
| 44 | 2854.4 | 2854.41 | 86.8 | 80.3 |
| 540 | 2854.4 | 2854.41 | 82.9 | 72.2 |

Of these, one additional peak group, 1269.7, showed potential isotopic profile overlap with an adjacent peak group. The four peak groups with potential overlap (groups 1269.7, 1413.7, 1589.8 and 1703.9) were not considered for further analysis as the true mono-isotopic peak could not be confirmed.

There was no significant change to the features selected for patients 44 and 540 when the cosine distance clustering was used to select the cancer spectra. This was expected given the virtually identical clustering results for the cluster maps generated previously (see **Figure 7.6a** and **7.6d**). The absence of any peak groups overlapping with patient 173 was also expected, due to the previously

described existence of necrotic tissue within the tumour/stroma region. Significantly, two additional peak groups, 1997.9 and 2429.3, were assigned to patient 544 when using the cosine distance. This confirmed the superiority of the cosine distance over the Hamming and brought the total overlapping peak groups for patient 544 to three (1997.9, 2189.9 and 2429.3), with two of these peak groups overlapping with patient 44 (1997.9 and 2189.9) and all three overlapping with patient 540. As expected given the correct cluster map generated using the cosine distance for patient 544, the additional peak groups of interest were an improvement over the Hamming distance results. However, the modest increase from one common peak group to three supported the contention that a portion of patient 544 tissue with minimal molecular similarity to patients 44 and 540 had been selected for MALDI-IMS analysis.

In attempting to identify morphological section features that confirmed this contention, the H&E stains for patients 540 (**7.10a**) and 544 (**7.10b**) were examined in more detail. It was found that the tissue region identified by a pathologist, and subsequently by clustering, as cancer (dotted black line in **7.10a** and **7.10b**) was morphologically different. **Figure 7.10c** and **7.10d** show zoomed views of these cancer regions for patients 540 and 544 respectively. It was clear from these histological stains that the sections from patients 540 and 544 did possess a morphologically similar cancerous component. However, as indicated by the black arrows in **Figure 7.10,** there was a secondary tissue component in the patient 544 cancer (**7.10d**) with a different morphology and less cellular density when compared to that found in patient 540 (**7.10c**). Crucially, this additional tissue was spread throughout the tumour mass and could not be discriminated using either S/N or ion intensity maps. The implication was that the intimate association of the cancer with this fibrous tissue allowed the region to be clustered as different to the adipose and off-tissue spectra for patient 544. However, this difference did not necessarily imply correlation with the cancer cluster from patient 540 where potentially different underlying metastatic mechanisms could have introduced morphological and therefore molecular differences.

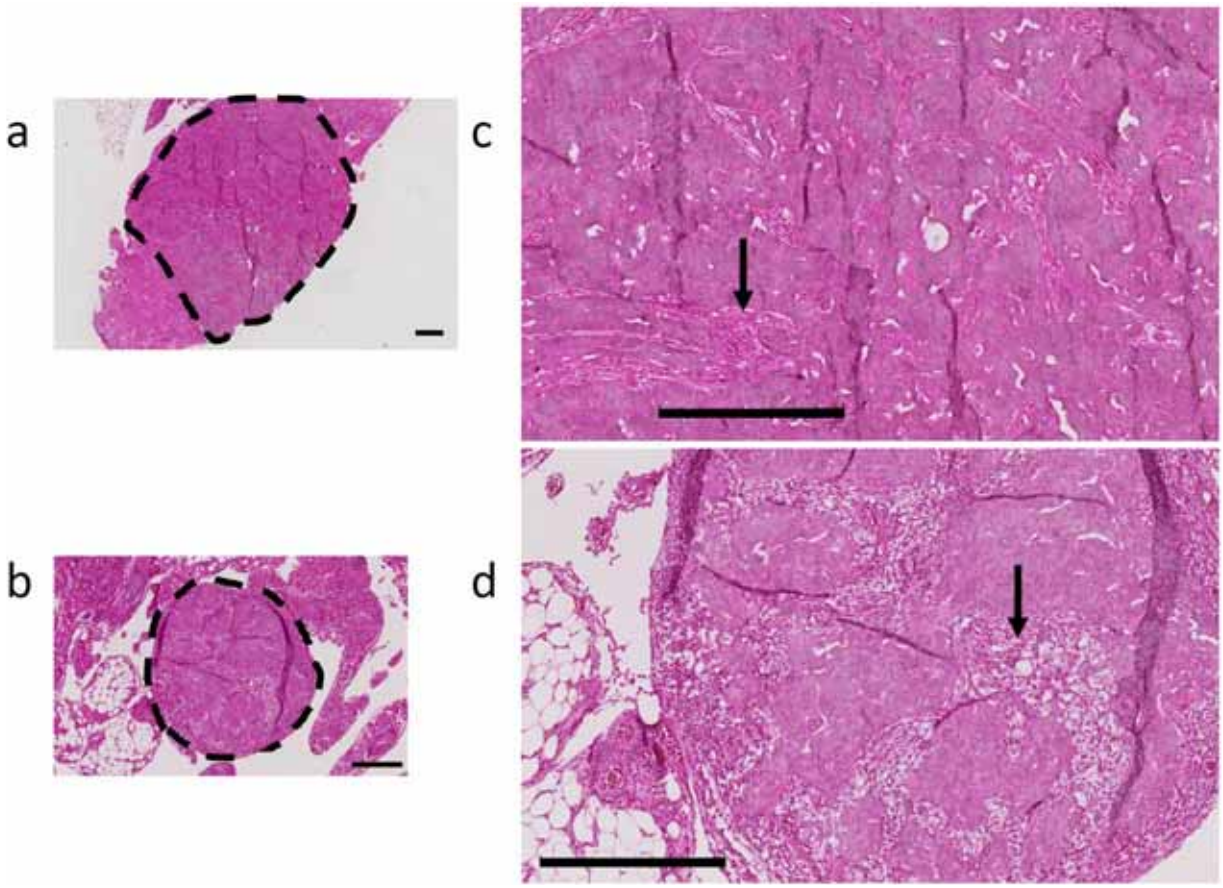



**Figure 7.10.** Haematoxylin and eosin stains for patient 540 (**a**) and 544 (**b**) following analysis by tryptic peptide MALDI-IMS. Green dotted outline shows a region on each section identified as cancerous by a pathologist and subsequently confirmed by the k-means clustering analysis. Zoom views from patient 540 and 544 are provided in panel **c** and **d** respectively to show the subtle differences in tissue morphology. All scale bars are 400 µm.

These data, including the different cancer morphology observed in patient 544, led to the conclusion that cancer heterogeneity was the most likely cause for the difference in molecular features when comparing patients 44 and 540 to patient 544. Those peak groups for patient 544 which did overlap with the other patients were thus sufficiently important to the cancer spectra to overcome the existence of additional tissue types within the tumour border. In order to address this, a more comprehensive characterization of each patient could be performed, from additional regions exhibiting necrotic, solid tumour and tumour margins. However, the heterogeneity within the patients would most likely not be identical, nor would biopsies collected for future patients always come from the same tumour region. Furthermore, it would be impractical to analyze multiple tumour regions in each patients, especially if the

number of patients included in a study increases. However, an increase in patient number would reduce the problem of heterogeneity across patients by providing a larger set of biological replicates **[7, 62]** and increase the chances of identifying common tumour markers.

### 7.2.7 Combining peak group features from Hamming and cosine distance results

The analysis of the spectra defined as cancerous (Hamming and cosine distance clustering) indicated twelve peak groups which were common to both (see **Table 7.1** and **Table 7.2**). The peak groups unique to the Hamming analysis were 1543.8 and 1703.9, while the peak groups unique to the cosine analysis were 1269.7 and 2429.3.

The MALDI-IMS experiments in this chapter were thus used to identify a set of peak groups discriminatory for cancer in three of the four patients. Two peak groups, 1997.9 and 2189.9, could distinguish cancer tissue in patients 44, 540 and 544. Prior to assignment of identifications to these peak groups, the data sets were further scrutinized to find features that could discriminate stroma from cancer.

### 7.2.8 Peritoneal stroma features

The assignment of peak groups of interest which could discriminate the peritoneal stroma was only performed for patients 44, 173 and 540. Patient 544 was excluded given the absence of significant amounts of stroma in the section, which caused the clustering to segregate cancer and adipose tissue spectra only.

The stroma peak groups of interest were reproducible for patients 44 and 540. **Table 7.3** shows the peak groups of interest contributing greater than 80% to PIC for at least two patients (same PIC/POS processing and filtering used for cancer data). The peak groups in **Table 7.3** were picked using the MS spectra memberships defined by the Hamming (**Table 7.3a**) and cosine distance (**Table 7.3b**) clustering analyses.

**Table 7.3.** Patient number, peak group abundance weight mean (AWM) *m/z*, percentage intensity contribution to the cancer cluster (PIC) and percentage of spectra in the cancer cluster containing each peak group of interest (POS) are shown for the cancer region clustered using the Hamming distance (**a**) and cosine distance (**b**) for patients 44, 173 and 540. All features were found in two or more patients and had PIC values greater than 80%. Peak groups in bold were found to have a PIC greater than 90%.

a

| Patient | Peak group AWM m/z | | PIC (%) | POS (%) |
|---|---|---|---|---|
| | ± 0.1 *m/z* | ± 0.01 *m/z* | | |
| 44 | 2284.1 | 2284.15 | 89.8 | 21.8 |
| 540 | 2284.1 | 2284.15 | 88.8 | 53.0 |
| 44 | 2301.1 | 2301.14 | 85.8 | 58.7 |
| 540 | 2301.1 | 2301.13 | 83.5 | 92.8 |
| 44 | 2454.2 | 2454.21 | 80.2 | 25.4 |
| 540 | 2454.2 | 2454.22 | 83.7 | 73.9 |
| 44 | 2495.2 | 2495.23 | 89.6 | 28.2 |
| 540 | 2495.2 | 2495.23 | 86.6 | 55.8 |
| 44 | 2581.3 | 2581.26 | 86.2 | 30.5 |
| 540 | 2581.3 | 2581.26 | 89.2 | 56.8 |
| **44** | **2611.3** | **2611.31** | **91.9** | **53.6** |
| **540** | **2611.3** | **2611.31** | **98.0** | **57.7** |
| 44 | 2703.3 | 2703.27 | 82.4 | 37.0 |
| 540 | 2703.3 | 2703.26 | 81.4 | 73.6 |
| **44** | **3008.4** | **3008.43** | **94.7** | **43.6** |
| **540** | **3008.4** | **3008.45** | **98.0** | **37.1** |
| 44 | 3240.5 | 3240.47 | 80.5 | 47.1 |
| **540** | **3240.5** | **3240.48** | **97.8** | **31.3** |
| 44 | 3377.7 | 3377.73 | 86.6 | 57.3 |
| **540** | **3377.8** | **3377.76** | **95.3** | **27.9** |

b

| Patient | Peak group AWM m/z | | PIC (%) | POS (%) |
|---|---|---|---|---|
| | ± 0.1 m/z | ± 0.01 m/z | | |
| 44 | 2284.1 | 2284.15 | 85.9 | 24.4 |
| 540 | 2284.1 | 2284.15 | 86.3 | 44.6 |
| 44 | 2301.1 | 2301.14 | 83.8 | 67.0 |
| 540 | 2301.1 | 2301.13 | 83.7 | 81.2 |
| **44** | **2495.2** | **2495.23** | **91.0** | **32.4** |
| 540 | 2495.2 | 2495.23 | 85.0 | 48.3 |
| 44 | 2611.3 | 2611.31 | 84.1 | 60.6 |
| **540** | **2611.3** | **2611.31** | **99.0** | **46.6** |
| 44 | 3008.4 | 3008.43 | 84.9 | 49.2 |
| **540** | **3008.4** | **3008.45** | **98.7** | **29.6** |

No peak groups listed in **Table 7.3** overlapped with peak groups from patient 173. This was expected, because the subtle boundary between the cancer, necrotic tissue and stroma in the patient 173 section would have made any overlap of stroma specific peptide peak groups unlikely. Therefore, no conclusive

assignment of peak groups to the patient 173 stroma was achieved. The peak groups in **Table 7.3**, which showed isotopic peptide profile overlap with adjacent peptides were 2495.2, 2581.3 and 2703.3. These were removed from the final list of peak groups of interest as the true mono-isotopic peak could not be confirmed. As before, four example peak groups of interest, which showed apparent minimal overlap, were presented as examples in **Figure 7.11**. Panels **7.11a** and **7.11c** show the H&E stains for patients 44 and 540 respectively. Annotations (black dotted outline) showing the area believed to be the peritoneal stroma are also included in this figure. Patient 44 (**7.11b**) and 540 (**7.11d**) S/N maps for the example peak groups 2611.3, 3008.4, 3240.5 and 3377.8 are shown. The flexImaging ion intensity maps for all four of these features are included in **7.11e** and **7.11f** for patients 44 and 540 respectively.

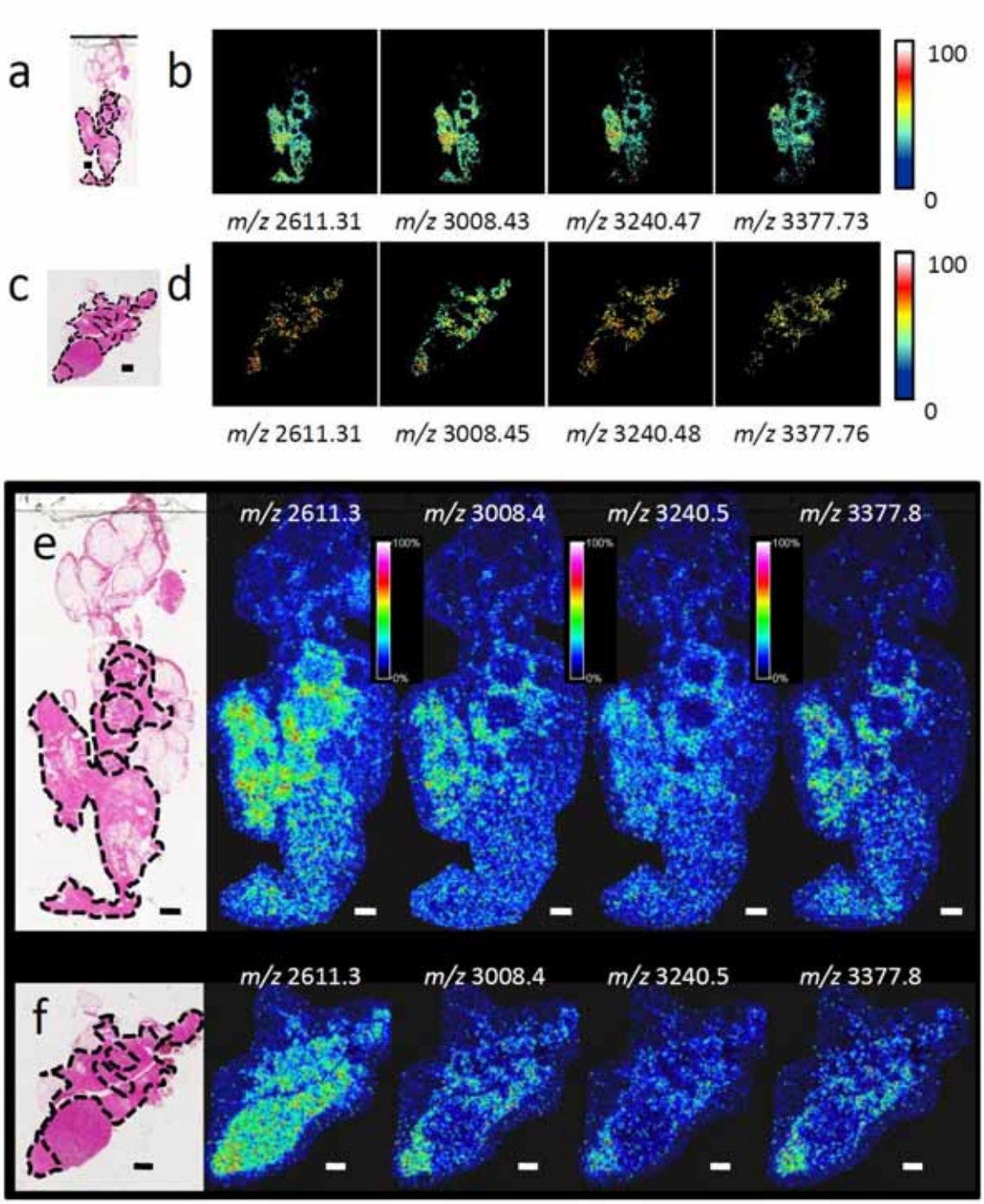



**Figure 7.11.** Haematoxylin and eosin stains are shown for patients 44 (**a**) and 540 (**c**) along with annotations indicating the peritoneal stroma (black dotted line). Signal to noise maps generated for patients 44 (**b**) and 540 (**d**) are shown for peak groups 2611.3, 3008.4, 3240.5 and 3377.7. Ion intensity maps generated in flexImaging are also shown for the same peak groups (0.1 Da mass filter) in panels **e** and **f** for patients 44 and 540 respectively. Scale bars are 1 mm and ion intensity scales are included. Ion intensity maps were normalized to noise.

There were two important observations made concerning the peak groups specific to the stroma in these two patients.

Firstly, and similar to the cancer data, the peak groups which matched the filtering criteria for PIC and POS showed highly localized, convincing distributions in their respective S/N maps. This was an important result as it confirmed that a series of peak groups could be isolated which clearly discriminated both stroma (**Figure 7.11a** to **7.11d**) and cancer (see **section 7.2.6**).

Secondly, the peak groups of interest for the stroma in both patients 44 and 540 were also present at a low S/N. While the S/N maps were of similar quality to those presented for the cancer features, the low S/N of the features caused the intensity maps for all four features to be noisy. Peak group 2611.3 even appeared to be present in the cancerous region for the patient 540 ion intensity map (see **Figure 7.11f**). Manual inspection of the individual MS spectra showed that the flexImaging ion intensity map mass filter was also plotting a trailing isotope peak for a more abundant peptide of lower mass. Crucially, without the S/N map, this feature would not have been identified as important. It was hypothesized that the low S/N of common features across the patients could have resulted from higher abundance peptides existing across the cancer/stroma boundary defined by the clustering. Using patient 44 as an example, several high abundance masses (relative to the original flexImaging sum spectrum) were compared to section histology. H&E stains annotated with the approximate stroma region (black dotted outline) and S/N maps for peak groups 1962.0, 2027.0 and 2519.2 are presented in **Figure 7.12a**, **7.12b** and **7.12c** respectively. It was immediately clear from the S/N maps for these features that they existed across the entire tissue section but were higher abundance in the region annotated as stroma. Peak groups 1962.0 and 2027.0 did indeed both exhibit POS values greater than 97% for both patients 44 and 540. However, neither of these peak groups contributed greater than 80% to PIC and were thus excluded from the final list of peak groups of interest. Conversely, peak group 2519.2 had a PIC value of 92.7% in patient 540, but only 61.8% in patient 44, indicating that this peak group was removed during processing because it was unique to the stroma cluster in only one patient. As such, low PIC values and

a lack of reproducibility across the two patients for the high S/N peak groups caused low S/N peak groups to be selected for inclusion in **Table 7.3**.

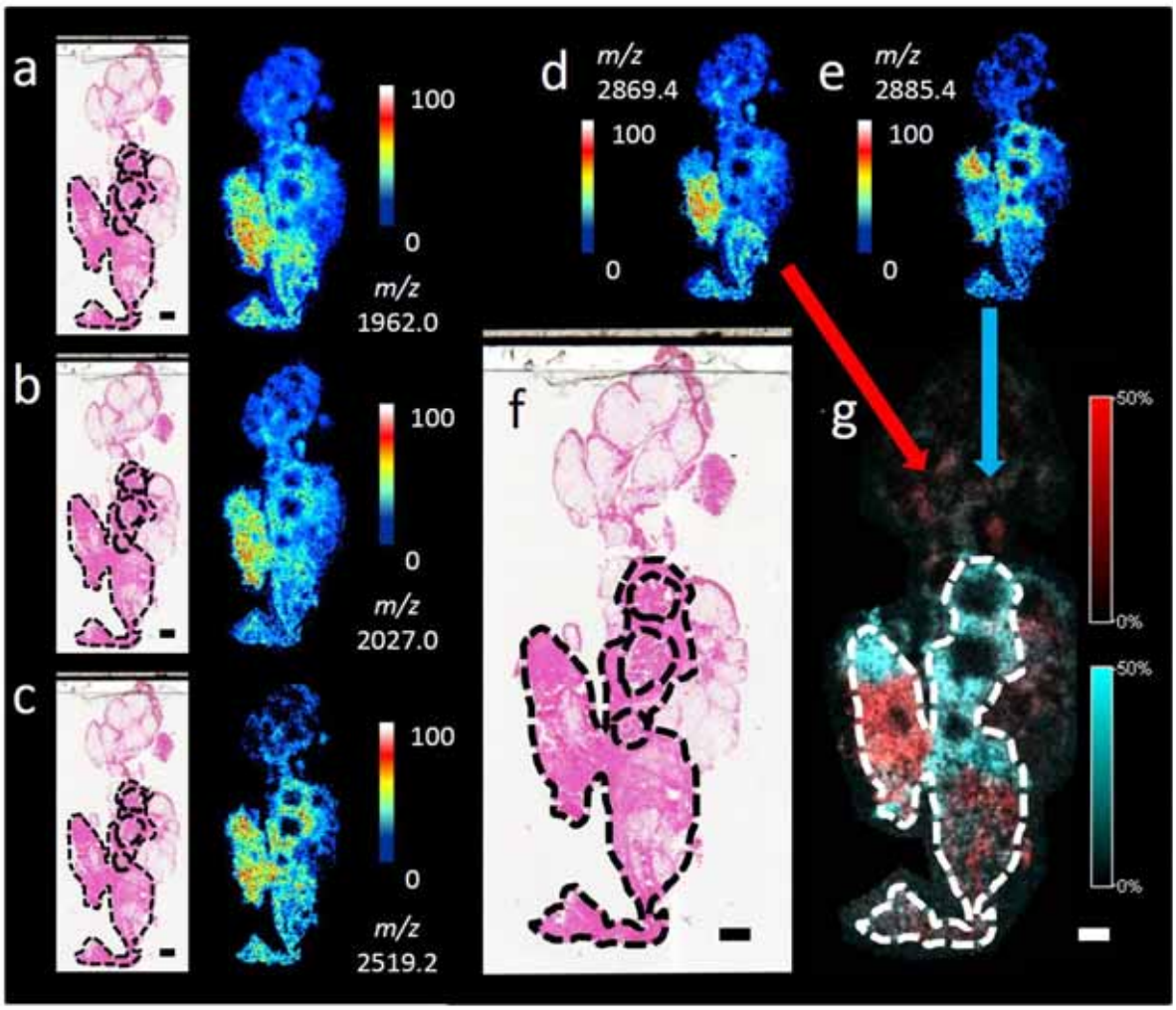



**Figure 7.12.** Apparent stroma specific high intensity peak groups plotted as signal to noise (S/N) maps for patient 44. Haematoxylin and eosin (H&E) stains are provided for patient 44 along with annotation (black dotted outline) showing the stroma regions of the section along with S/N maps for peak groups 1962.0, 2027.0 and 2519.2 in **a**,**b** and **c** respectively. Panel **d** and **e** show S/N maps for peak groups 2869.4 and 2885.4 respectively. Enlarged patient 44 H&E stain with stroma annotation and a flexImaging overlay of ions 2869.4 (red) and 2885.4 (light blue) are shown in panel **f** and **g** respectively. Scale bars are 1 mm and scales for ion and S/N intensity are included.

In addition, the stroma itself demonstrated heterogeneity. Using patient 44 as an example, two S/N maps for peak groups 2869.4 and 2885.4 are presented in **Figure 7.12d** and **7.12e**. In this case these two high intensity peak groups, which were separated by 16 Da (possible oxidation), distinguished between a region of tissue surrounding the cancer (2885.4, **7.12e**) and a region of low density stromal tissue (2869.4, **7.12d**). The two peak groups were overlayed in flexImaging to confirm the distributions. **Figure 7.12f** presents an enlarged annotated (stroma) H&E stain for patient 44. The overlay of the two ions using a red (2869.4) and blue (2885.4) colour scheme appears in **Figure 7.12g**. The two ions overlayed almost perfectly in **7.12g**, confirming heterogeneity within the region annotated as stroma. Similar to the other high abundance features, the widespread detection of the peak groups meant that potentially important features distinguishing cancer from stroma were missed by the PIC filtering process. Despite the heterogeneity it was possible to isolate peak groups which could be used to distinguish cancer and stroma. However, the existence of heterogeneity pointed to the need for application of additional data analysis methods to comprehensively characterize a more complete set of peak groups for molecular annotation of the metastatic tumour boundary.

The peak groups which could distinguish cancer from stroma were subsequently identified using the LC-MS/MS match lists generated in **section 7.2.3**.

### 7.2.9 Identification of cancer and stroma-specific peak groups by LC-coupled mass spectrometry

**The cancer-specific peak groups of interest** from **Table 7.1** and **Table 7.2**, which did not show any apparent overlap with adjacent peak groups, were used to mine the peptide identification match lists described in **section 7.2.3**. **Table 7.4** details the results of the comparison between the high grade cancer peak groups of interest and the identifications made by LC-MS/MS. It should be noted that validation by IHC is required prior to definitive assignment of these identifications. However, validation

was delayed until the importance of the cancer specific peak groups could be confirmed in a larger number of patients.

A total of sixteen tentative protein identifications were made to nine MALDI-IMS peak groups. Six of the nine peak groups had a single protein identification assigned. These peak groups and their corresponding identities were 1609.8 (Heterogeneous nuclear ribonucleoprotein R), 1762.9 (Citrate synthase), 1997.9 (Prohibitin), 2189.9 (Heterogeneous nuclear ribonucleoproteins A2/B1), 2354.1 (Keratin, type II cytoskeletal 7) and 2854.4 (Keratin, type I cytoskeletal 18). Of these identifications, Prohibitin (peak group 1997.9) and Heterogeneous nuclear ribonucleoproteins A2/B1 (peak group 2189.9) existed in patients 44, 540 and 544. The MS/MS spectra (b and y ion annotated) for the peptides matched to peak groups 1997.9 and 2189.9 are shown in Supplementary Figure 7.11. Prohibitin, hnRNP A2/B1 and cytokeratin 7 have potential importance to cancer progression and are therefore presented briefly.

Prohibitin has previously been found to be up-regulated in ovarian carcinomas by two-dimensional gel electrophoresis **[126]**. The protein itself is a negative cell cycle regulator which blocks the G1/S cell cycle transition **[127]**. Up-regulation of this protein within the tumour boundary of ovarian peritoneal metastases could indicate a decrease in cell proliferation, potentially as a result of the advanced differentiation state of the high grade tumour. However, in the absence of further experiments, this was hypothetical. Up-regulated Prohibitin has also been found to provide an anti-apoptotic effect in human B cell and breast cancer cell lines **[128]**. Thus, although this protein is typically a high abundance proteome component, its potential importance to cell cycle arrest and evasion of apoptosis lends weight to its inclusion as part of a panel of MS detectable markers for annotation of ovarian metastases.

In addition to their tentative identification as part of this chapter, heterogeneous nuclear ribonucleoproteins (hnRNPs) were also identified as part of the method development in **chapter five** (see **Table 5.1**). hnRNPs play a potentially important role in various cancer processes. hnRNPs are involved in processing heterogeneous nuclear mRNA into mature mRNA strands and act as trans-

factors for gene expression regulation **[129]**. In this role hnRNP A2 has been implicated as a key splicing factor regulating TP53 splicing and subsequent cancer cell invasion **[130]**. Furthermore, the expression patterns of hnRNPs A1 and A2 have been suggested to be crucial for metabolic changes in cancer cells which drive aerobic glycolysis **[131]**. The characteristics of hnRNP A2/B1 also support the inclusion of this protein as a part of a marker panel for annotating ovarian metastases.

Interestingly, two cytokeratins (7 and 18) could be used to distinguish cancer from surrounding tissue in patients 44 and 540. The identification of cytokeratins as discriminatory for an ovarian cancer metastasis was significant because these proteins are used as IHC markers for ovarian tumours. As an example, cytokeratin 7 is expressed in the majority of primary ovarian carcinomas **[132, 133]**. The implication of the work presented here was that rather than indirectly detecting this marker by IHC, it would be possible to directly detect a peptide of cytokeratin 7 by MALDI-IMS. This would be a more direct and less expensive way to evaluate cytokeratin 7, and other tryptic peptides as known molecular markers that distinguish ovarian cancers from other malignancies **[133]**.

**The stroma-specific peak groups of interest** from **Table 7.3** which did not show any apparent overlap with adjacent peak groups were also used to mine the peptide identification match lists described in **section 7.2.3**. The results of the comparison are detailed in **Table 7.5**. Unfortunately, only one of the peak groups of interest, 2611.3, was identified. The MS/MS spectrum for this peptide (b and y ion annotated) is shown in Supplementary Figure 7.11. The peptide was from the protein Glyceraldehyde-3-phosphate dehydrogenase (GAPDH), which has been shown previously to be a pro-apoptotic factor **[134]**. GAPDH is over expressed in most human cancers, and has been implicated in the promotion of apoptosis through downstream activation of p53 **[134]**. In contrast, GAPDH has also been found to protect against caspase-induced cell death, which is achieved by countering mitochondrial outer membrane permeabilization through increased autophagy to remove damaged mitochondria **[134]**.

**Table 7.4.** Identified cancer specific peptides. MALDI-IMS mono-isotopic mass ([M+H]+), LC-MS/MS [M+H]+, sample source for identification (patient and laser capture micro-dissection sample, LCM), accession number (SwissProt), protein name and peptide sequence for peak groups of interest matched to tryptic peptides identified by laser micro-dissection, tryptic digestion and LC-MS/MS.

| IMS [M+H]+ | LC-MS/MS [M+H]+ | Sample source for ID | | ACCESSION | Protein source name | Peptide sequence |
|---|---|---|---|---|---|---|
| | | Patient | LCM | | | |
| 1406.70 | 1406.674 | 44 | Adipose | ATPB_HUMAN | ATP synthase subunit beta, mitochondrial | AHGGYSVFAGVGER |
| 1406.70 | 1406.674 | | Cancer | | | |
| 1406.70 | 1406.709 | | Adipose | ENOA_HUMAN | Alpha-enolase | GNPTVEVDLFTSK |
| 1406.70 | 1406.709 | | Stroma | | | |
| 1406.70 | 1406.709 | | Cancer | | | |
| 1406.70 | 1406.709 | 540 | Adipose | | | |
| 1406.70 | 1406.709 | 540 | Stroma | | | |
| 1406.70 | 1406.709 | 540 | Cancer | | | |
| 1406.70 | 1406.709 | | Cancer | ENOG_HUMAN | Gamma-enolase | GNPTVEVDLYTAK |
| 1406.70 | 1406.739 | 44 | Cancer | FAS_HUMAN | Fatty acid synthase | VLQGDLVMNVYR |
| 1406.70 | 1406.712 | 540 | Cancer | HXK1_HUMAN | Hexokinase-1 | LSDETLIDIMTR |
| 1406.70 | 1406.663 | | Adipose | SAMP_HUMAN | Serum amyloid P-component | AYSLFSYNTQGR |
| 1609.79 | 1609.792 | 44 | Cancer | HNRPR_HUMAN | Heterogeneous nuclear ribonucleoprotein R | DLYEDELVPLFEK |
| 1609.78 | 1609.792 | 540 | Stroma | | | |
| 1739.91 | 1740.019 | 44 | Stroma | MVP_HUMAN | Major vault protein | KEVEVVEIIQATIIR |

| | | | | | | |
|---|---|---|---|---|---|---|
| 1739.91 | 1740.019 | | Cancer | | | |
| 1739.92 | 1739.925 | 540 | Stroma | COCA1_HUMAN | Collagen alpha-1(XII) chain | NNVILQPLQPDTPYK |
| 1739.92 | 1740.019 | | Cancer | MVP_HUMAN | Major vault protein | KEVEVVEIIQATIIR |
| 1762.85 | 1762.879 | 44 | Cancer | CISY_HUMAN | Citrate synthase, mitochondrial | GLVYETSVLDPDEGIR |
| 1762.86 | 1762.879 | 540 | Stroma | | | |
| 1936.04 | 1936.014 | 44 | Stroma | LBP_HUMAN | Lipopolysaccharide-binding protein | SPVTLLAAVMSLPEEHNK |
| 1936.04 | 1935.920 | | Cancer | TERA_HUMAN | Transitional endoplasmic reticulum ATPase | GPELLTMWFGESEANVR |
| 1997.93 | 1998.079 | 44 | Cancer | PHB_HUMAN | Prohibitin | AAELIANSLATAGDGLIELR |
| 1997.94 | 1998.079 | 540 | Stroma | | | |
| 1997.94 | 1998.079 | | Cancer | | | |
| 1997.90 | 1998.079 | 544 | Cancer | | | |
| 2189.94 | 2189.898 | 44 | Cancer | ROA2_HUMAN | Heterogeneous nuclear ribonucleoproteins A2/B1 | NMGGPYGGGNYGPGGSGGSGGYGGR |
| 2189.92 | 2189.898 | 540 | Cancer | | | |
| 2354.13 | 2354.260 | 44 | Cancer | K2C7_HUMAN | Keratin, type II cytoskeletal 7 | AKLEAAIAEAEERGELALKDAR |
| 2354.14 | 2354.260 | 540 | Cancer | | | |
| 2854.41 | 2854.401 | 44 | Cancer | K1C18_HUMAN | Keratin, type I cytoskeletal 18 | SLGSVQAPSYGARPVSSAASVYAGAGGSGSR |
| 2854.41 | 2854.401 | 540 | Cancer | | | |

**Table 7.5**. Identified STROMA specific peptides. MALDI-IMS mono-isotopic mass ([M+H]+), LC-MS/MS [M+H]+, sample source for identification (patient and laser capture micro-dissection sample, LCM), accession number (SwissProt), protein name and peptide sequence for peak groups of interest matched to tryptic peptides identified by laser micro-dissection, tryptic digestion and LC-MS/MS.

| IMS [M+H]+ | LC-MS/MS [M+H]+ | Sample source for ID | | ACCESSION | Protein source name | Peptide sequence |
|---|---|---|---|---|---|---|
| | | Patient | LCM | | | |
| 2611.31 | 2611.348 | 44 | Adipose | G3P_HUMAN | Glyceraldehyde-3-phosphate dehydrogenase | VIHDNFGIVEGLMTTVHAITATQK |
| | | | Stroma | | | |
| | | | Cancer | | | |
| 2611.31 | 2611.348 | 540 | Stroma | | | |
| | | | Cancer | | | |

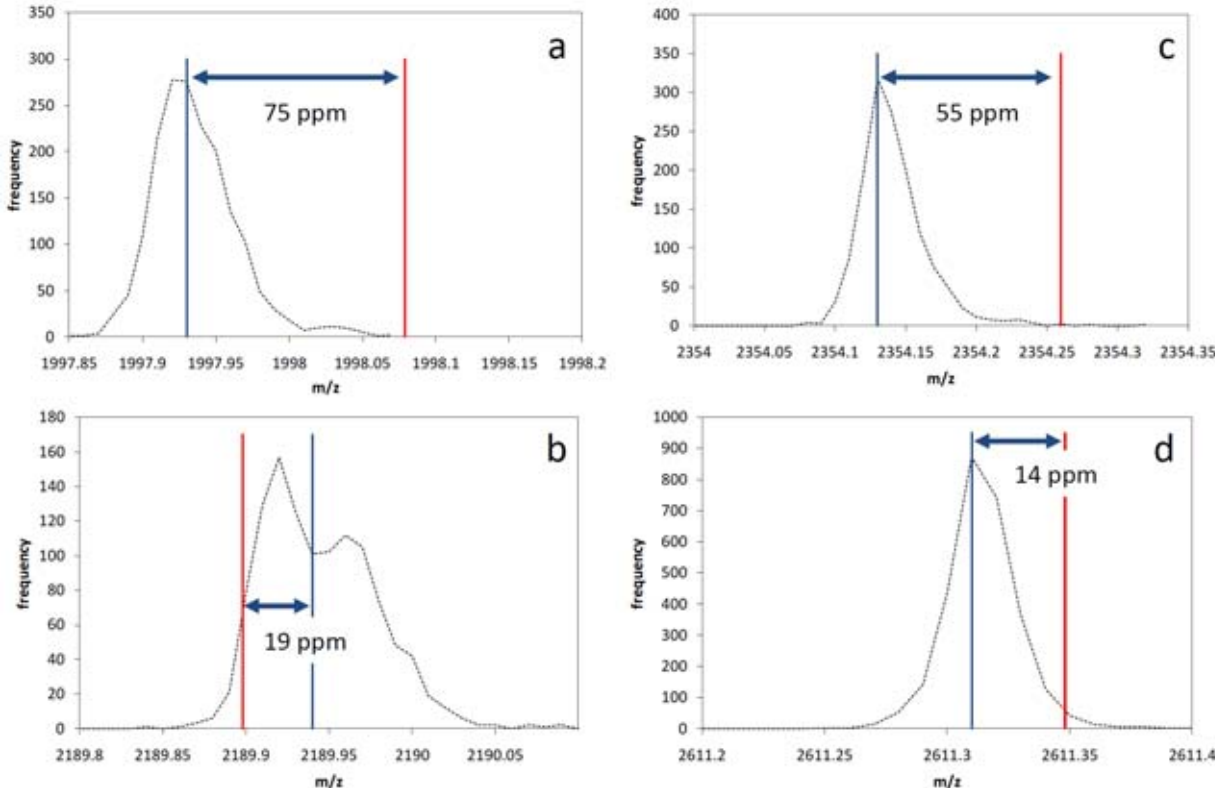



**Figure 7.13.** Patient 44 *m/z* histograms generated from peak groups of interest 1997.9 (**a**), 2189.9 (**b**), 2354.1 (**c**) and 2611.3 (**d**). Each histogram shows the *m/z* deviation within a peak group and is annotated with the abundance weighted mean (AWM) m/z (blue line) and the *m/z* of the peptide identified by LC-MS/MS (red line) and subsequently matched to the peak group. The ppm difference between the AWM *m/z* and identified peptide m/z are shown on each histogram.

The reported roles of GAPDH point to one of several explanations for the observation that this protein was down-regulated in the cancerous tissue for patients 44, 540 and 544. In addition, this discrepancy exemplified important considerations for conclusions drawn from the results presented in this thesis.

Firstly, the correlation between the MALDI-IMS peak group and the LC-MS/MS identified peptide could be incorrect. The assumption made in this thesis was that the peptides identified by LC-MS/MS, from adjacent sections of the same cancer and stroma tissue, could be mass correlated to MALDI-IMS peak groups for tentative identification. However, as already described in **chapter 6**, based on the *Mus musculus* protein database (SwissProt), more than 160 unmodified peptides can exist within a 100 ppm mass window **[114]**. Because this situation would more than likely be mirrored for the data presented in this thesis, the AWM *m/z* and LC-MS/MS peptide *m/z* were compared for the cancer and stroma specific

peptide peak groups of interest discussed above. **Figure 7.13** shows the *m/z* histograms for the peak groups 1997.9 (**7.13a**), 2189.9 (**7.13b**), 2354.1 (**7.13c**) and 2611.3 (**7.13d**). It was noticed from these data that there was variability in the error between the MALDI-IMS peak group AWM *m/z* and the *m/z* of the identified peptide. This variability ranged from 14 ppm for peak group 2611.3 (hnRNP A2/B1) to 75 ppm for peak group 1997.9 (Prohibitin).

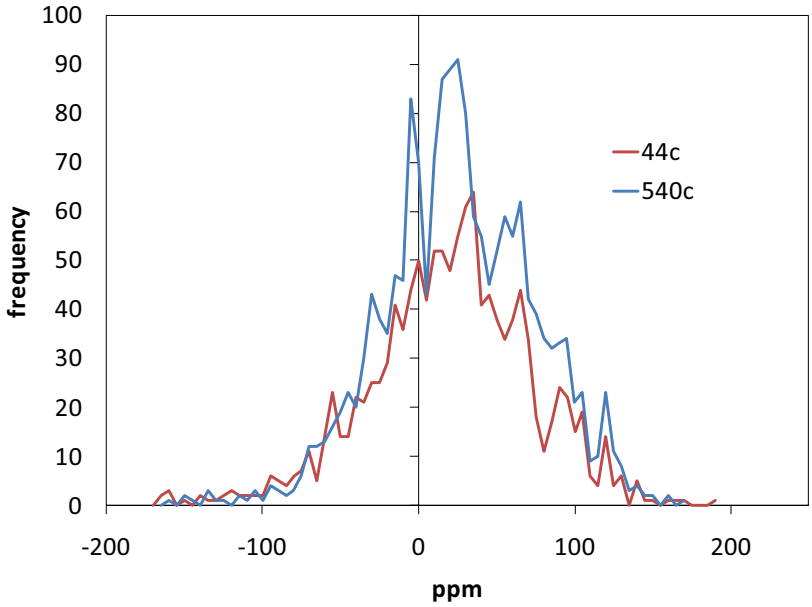



**Figure 7.14.** Example histogram (bin width 5 ppm) showing the error (ppm) between the MALDI-IMS peak group assignments and the LC-MS/MS identified peptides for the LCM tryptic peptide digests. This is shown for the cancerous regions (marked with **c** above) for patient 44 (red line) and 540 (blue line) matched to the peak groups plotted as S/N maps for the same patients.

To ensure that the variability was not a result of poor matching, the frequency of the error (ppm) between the MALDI-IMS peak groups and the matched LC-MS/MS peptides from the cancer regions isolated by LCM were plotted in **Figure 7.14** for patients 44 (red line) and 540 (blue line). The figure shows that the error was roughly normally distributed (using a bin width of 5 ppm), but slightly shifted in the positive direction. The ppm error for the majority of masses was below 100 ppm. Because the error for assignment of the MALDI-IMS peak groups to the identified peptides was variable, the identification of these features was still tentative. But the figure also indicated that all the matched peak groups of interest fell within a normal variation range, which was less than the 0.2 Da (166 ppm) tolerance used

for the IMS-LC-MS/MS matches. In order to confirm the tentative identifications described here, prior to pursuing expensive validation such as IHC, it may be possible to predict the fragmentation pattern of the matched peptides and use this information to annotate *in situ* MS/MS spectra. As described previously, and in contrast to what has been demonstrated in the literature **[63, 135]**, it was difficult to perform MS/MS peptide identification directly from tissue (see **chapter 5**). However, even if the number of internal fragments generated for a peptide is limited by the *in situ* acquisition, the fragmentation pattern should be unique to the peptide being fragmented. These internal fragments could be used to confirm the identifications in this chapter. Proteomics groups have already investigated patterns of peptide cleavage based on their amino acid composition **[136]**. A 2003 manuscript detailed the fragmentation, database annotation and manual validation of 5500 tryptic peptide sequences using an ion trap MS instrument **[137]**. Importantly, the fragmentation patterns for these peptides were analyzed and cleavage abundance ratios were determined for all residue combinations. Although an ion trap, not a MALDI-TOF/TOF, was used for MS/MS **[137]** the use of sequence information as part of peptide scoring functions for MALDI-TOF/TOF MS/MS data has also been demonstrated **[136]**. If the LC-MS/MS match lists described as part of this thesis could be combined with predictions for the abundance of unique internal fragments for the peptides of interest (e.g. peak group 1997.9), the identifications made in this chapter could be confirmed prior to performing IHC validation. This would not only provide more conclusive evidence for the identity of selected tryptic peptide markers, but would reduce the number and therefore cost of subsequent IHC validations.

The second explanation for the apparent down-regulation of GAPDH in the cancer tissue was that either the MS measurement or the SNAP peak picking algorithm did not perform as expected. In the context of the MS measurement, it would be possible for higher abundance peptides present in the cancerous tissue to suppress gas phase protonation of the 2611.3 species by sequestering charge away from this peptide **[138]**. Ionization efficiency could be measured across a tissue section by means of a suitable internal peptide standard or a ubiquitously expressed tissue marker. If this standard could be homogeneously distributed, it could be exploited to map changes in ionization efficiency due to peptide

mixture complexity or tissue composition (e.g. tissue density) **[139]**. Regarding the potential impact of peak picking, it was considered that suppression effects or simply just low S/N of the 2611.3 peak in the cancer region may have resulted in the peptide not being included in the original set of peak lists exported from flexAnalysis. However, manual inspection of randomly selected spectra from cancer and stroma regions suggested that the difference was not due to poor peak picking. The S/N of the 2611.3 peptide was too low to be considered a valid peptide peak feature in these spectra (data not shown).

The final potential explanation was that a real biological difference that contradicted previous findings had been identified in these tissue sections. This hypothesis was subject to further experiments addressing the issues described above, as well as experiments confirming the importance of GAPDH in a larger patient data set. Prior to analyzing a larger patient data set it will also be necessary to make improvements to the MALDI-IMS workflow described in this chapter. These improvements should allow a more complete and robust characterization of ovarian cancer metastasis by MALDI-IMS.

### 7.2.10 Improved workflow for characterization of ovarian peritoneal metastasis

In the previous sections it was demonstrated that an in-house developed method for analyzing tryptic peptides by MALDI-IMS, in combination with IonMapper processing and clustering analysis, could generate putative lists of tissue specific markers. Importantly, these peptide markers were a requirement for the future aims of this project, such as prediction of patient treatment outcomes. Prior to making definitive conclusions about the importance of the tryptic peptides identified here, several steps are needed in order to improve the robustness of the post MS data processing workflow.

Firstly, peak picking accuracy needs to be improved. Both the correct picking of the mono-isotopic peak and discrimination of the correct peptide isotopic profile are required. The SNAP algorithm used by flexAnalysis determines the mono-isotopic peak by fitting the profile spectrum to theoretical isotopic

peptide profiles **[140]**. The SNAP algorithm works well for typical MALDI-TOF measurements on a solid sample support. However, SNAP can fail to pick the mono-isotopic peptide peak for MALDI-IMS spectra when the isotopic profiles of two or more tryptic peptides overlap. A different open source peak picking tool could be evaluated for application to MALDI-IMS, given that several have already been compared in the literature for MALDI-TOF MS spectra **[141]**. An additional potential strategy would be the application of an independent method for predicting the isotopic profile of tryptic peptides. Unfortunately, it is not known what algorithm is used for SNAP peak picking. However, an independent prediction algorithm could utilize the averagine amino acid standard to model the change in isotopic profile that occurs with increasing *m/z* **[142]**. The averagine model could then be used to isolate overlapping peptides and allow visualization of potential contributing isotopic peptide patterns. With a better understanding of the complete peptide profile present in a data set it will be possible to determine the relative abundance of more peptide features as well as the suitability of the overlapping features for tissue annotation.

Secondly, the spectra from a single tissue section either need to be aligned to each other or a list of landmark peptide masses (see **chapter 6**). This will ensure that *m/z* deviation within a data set is as low as possible. As described in the previous section, and exemplified in **Figure 7.13b**, the internal calibrants can not re-calibrate every spectrum perfectly because of the inherent difficulty of ensuring even distribution of the internal calibrants across the whole tissue section. This resulted in a bimodal distribution for the peak group 2189.9 *m/z* histogram (**Figure 7.13b**). The calibration would also be affected by ionization suppression as a result of the complex peptide mixtures present *in situ*. The suppression would be greatest for Angiotensin I and [Glu$^1$]-Fibrinopeptide B as these calibrants were in the most peptide rich *m/z* range for the MALDI-IMS experiments described in this thesis (*m/z* 1200-2000). In **chapter 6** it was demonstrated that use of the internal calibrants could reduce *m/z* deviation at FWHM (*m/z* histogram for peak group) from 54 ppm to 20 ppm for an example peak group. It would be of great benefit to further reduce the *m/z* deviation in a typical MALDI-IMS measurement to below 10 ppm and consequently increase peptide identification confidence through a reduction in the number of

peptide database matches. This could be achieved by the alignment of spectra and re-calibration of the whole data set to the internal calibrants.

The improved peak picking and spectral alignment algorithm will increase analytical robustness and should be coupled to MALDI-IMS analysis of more patients. This will address the aforementioned intra and inter-patient tissue heterogeneity. Similar to previously published work, it is envisioned that at least fifty patients with stage III high grade peritoneal metastases would be needed to begin addressing heterogeneity **[62]**. Ultimately, tissue micro-arrays are also needed as they will speed up analysis times by reducing the amount of MS data generated per patient **[62, 143]**.

The fourth and final step that needs to be achieved is the application of additional statistical methods to further address tissue heterogeneity and robustly assign peak groups of interest across multiple patients. K-means clustering was used in the current project as an exploratory step to demonstrate the existence of key features discriminatory for cancer and stroma. Given that clustering has previously been used to discern tissues with different molecular fingerprints, it was expected that cancer and stroma could be clearly distinguished in this project. It was shown that cancer and stroma could be clearly distinguished using clustering for three of the five patients analyzed (see **Figure 7.6**). However, the data presented in this chapter indicated that intra-cluster variability may not have been as low as possible (see **Figure 7.12**). In essence, more clusters existed than were allocated by a k-value of four. In order to account for this using k-means clustering, it would be necessary to test several k-values for each clustering distance and each patient. This approach is less than ideal because ultimately clusters from different patients will not necessarily contain the same key peptide peak groups. As such, operator based interpretation and data analysis would always be required, and would be complicated by the need for a choice between k values. The logical next step for this project lies either in the evaluation and combination of other types of multivariate statistics, as was recently proposed for MALDI-IMS **[144]**, or the generation of a tissue spectra classifier **[145, 146]**.

The development of a classifier for ovarian cancer metastasis would be preferable given that the application of multiple multivariate statistics presents the same problem as k-means clustering. In effect the operator is responsible for selecting the best analysis method and interpreting the results. Furthermore, an appropriate classifier will identify key tryptic peptides that span multiple patients. As demonstrated in this chapter, individual clustering analyses may segregate anatomically similar regions, but the underlying features contributing to the segregation do not have to be the same. Thus, by employing an analysis that uses MALDI-IMS data from multiple patients, common peak groups which distinguish cancer and stroma in a majority of the patients could be isolated.

To create a classifier, the patient samples would be split into a teaching and validation set. The crucial information for the training set is the meta data (e.g. cancer spectrum or stroma spectrum), which groups the MS spectra from the patients into classes. A classification algorithm (e.g. SVM) determines the key features which are different between these classes. As an example, the classifier could identify the tryptic peptides which distinguish cancer spectra from stroma spectra. An initial input of information, in this case whether the spectra are from cancer or stroma (determined by a pathologist), is thus required in order to assign a cancer class and stroma class. In order to ensure robust validation, the classifier would then be applied to an independent set of cancer and stroma spectra. Importantly, the source of these spectra would not be known. This ensures that the classifiier, which is based upon the initial classes used for the teaching set (e.g. cancer/stroma), can correctly classify unknown patient spectra.

In the context of classification, the tryptic peptides of interest discovered in this chapter were all sourced from proteins that are high abundance relative to the rest of the proteome. While this initially seemed like a disadvantage, molecular features indicative of cancer type, grade and patient outcome do not have to be proteins that are fundamental to the underlying metastatic process. In fact, the identity of discriminatory features need not be known at all, as demonstrated by identification of different NSCLC sub-types **[46]** and grading of STS **[47]**. Both studies classified tissues based on their MS profile at the protein level. As the principle was sound it has also been applied at the peptide level. It was

demonstrated in a 2008 manuscript that a tryptic peptide MALDI-IMS method could be used for classification of NSCLC using a support vector machine (SVM) based algorithm **[62]**. Adenocarcinoma and squamous cell carcinoma regions within a tissue micro-array of needle core biopsies were successfully discriminated in this work with accuracies (relative to pathologist annotation) of 97.9% and 98.6% respectively. The SVM classification for this study was performed in ClinProTools software (Bruker Daltonics), which was originally written for protein profiling MS data that almost invariably exhibits Gaussian peak shapes where the apex of the peak represents the average mass. Conversely, peptide data always contains multiple isotope peaks for each peptide feature, including the mono-isotopic mass. This means the researchers in the study were forced to manually select the mono-isotopic peaks for the entire average spectrum which was generated by ClinProTools **[62]**. Despite this disadvantage it was still possible to develop a classifier for adenocarcinoma and carcinoma.

A peak group type approach as described in this thesis has the advantage of presenting a set of key features rather than entire profile spectra. This negated the use of ClinProTools. However, the potential advantages include the use of the various free software platforms for classification which are available **[145]**. Furthermore, the clustering in this thesis was performed on peak lists extracted from proprietary formats to text files (ASCII) which could be converted for use by analytical programs such as Matlab. Therefore, the possibility exists for creation of analytical tools (multivariate or classification) that incorporate existing Matlab packages and algorithms **[144]**. The aim would be to use the peak list text files, along with meta data, to first generate a classifier composed of peptide features that could distinguish cancer and stroma across the 50 patient data set suggested above. Following validation in an independent set of patients the classifier could be developed further by incorporating more patients or changed to search for peptides that distinguish cancer spectra from patients who responded well to chemotherapy versus those who did not. Importantly, because the original peak list data would be in a text file, it could be analyzed by anyone with access to Matlab, making it relatively easy to produce data, analysis procedures as well as interpreted results and make them available to the scientific community for independent evaluation **[144]**.

The ultimate aim of future work is to further improve and add to the methods developed in this chapter. This includes making improvements to peak picking and spectral alignment as well as increasing the patient number analyzed. The peak groups resulting from this more exhaustive analysis will then be used to generate a classifier capable of discriminating peritoneal metastases from surrounding tissue.

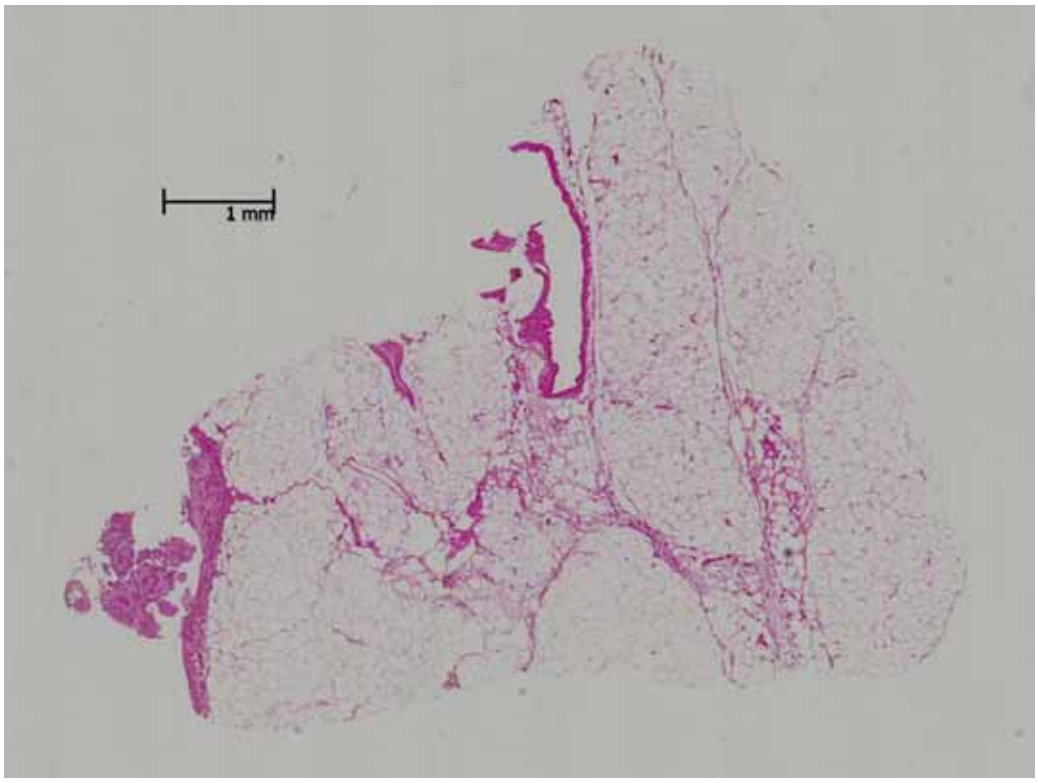
## 7.3 Supplementary Information




Supplementary Figure 7.1. High resolution haematoxylin and eosin (H&E) stain for patient 44.



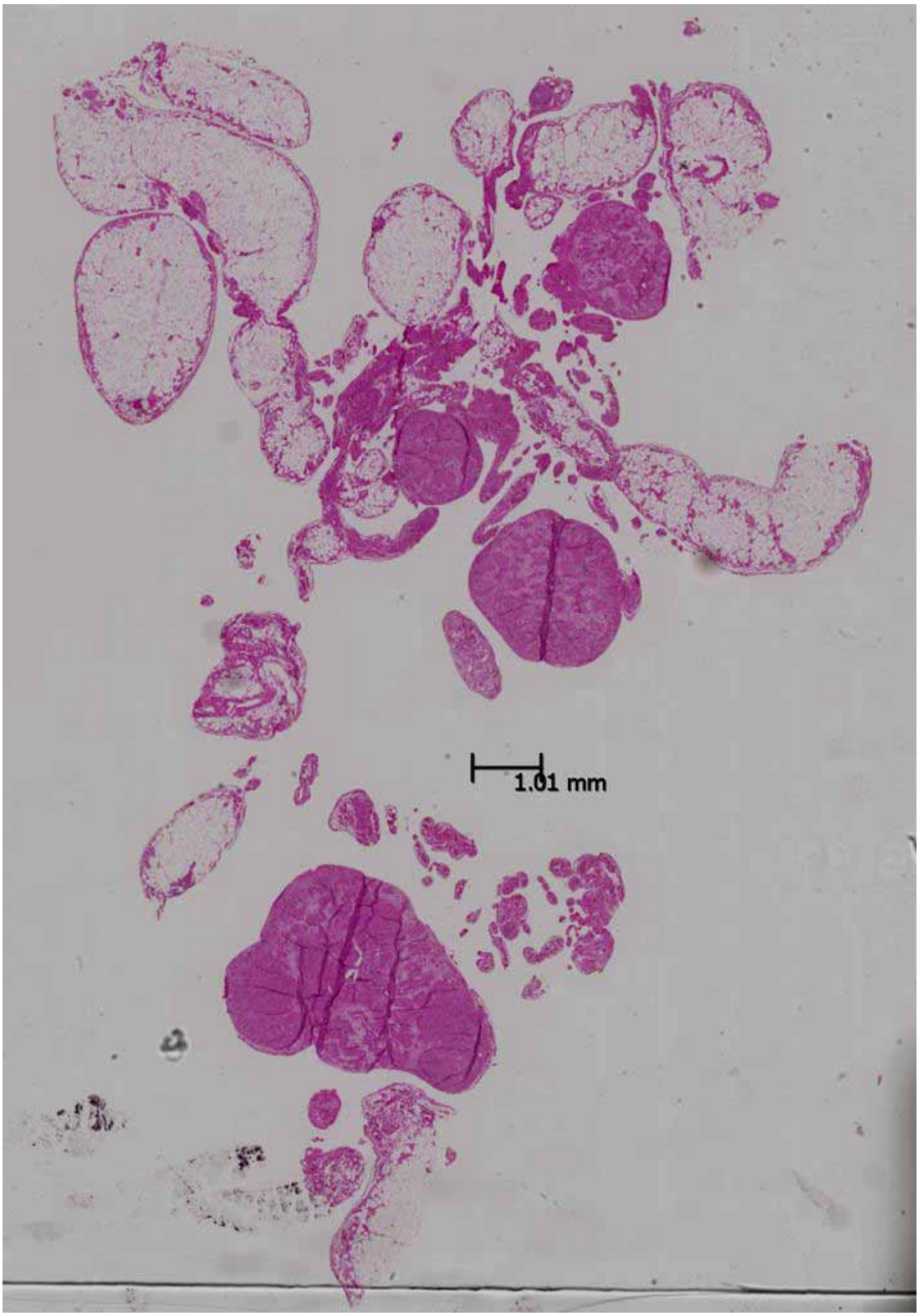

Supplementary Figure 7.2. High resolution haematoxylin and eosin (H&E) stain for patient 173.



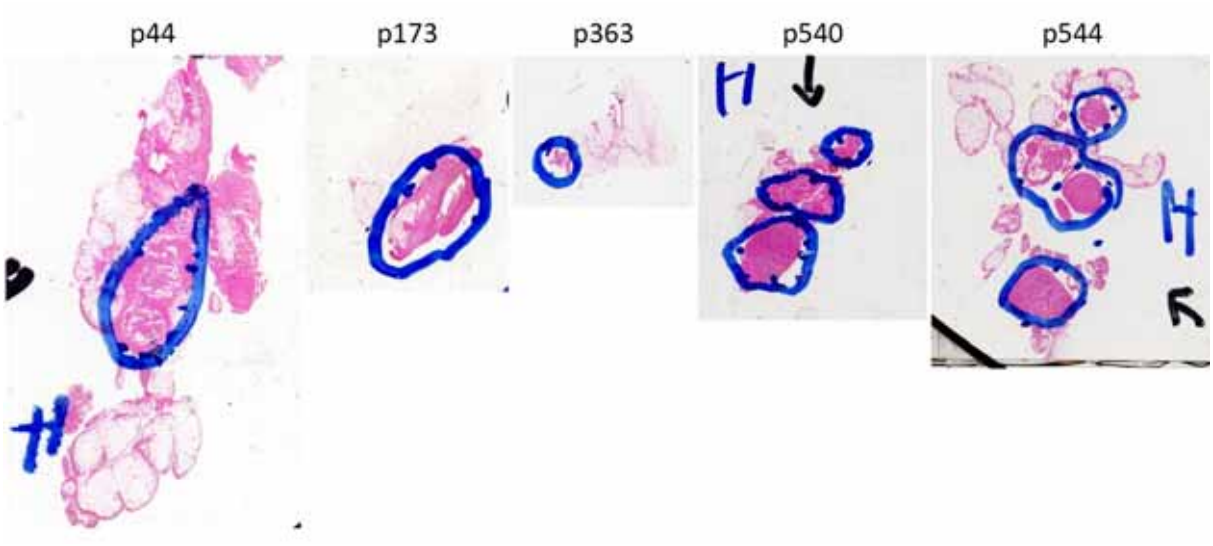

Supplementary Figure 7.3. High resolution haematoxylin and eosin (H&E) stain for patient 363.



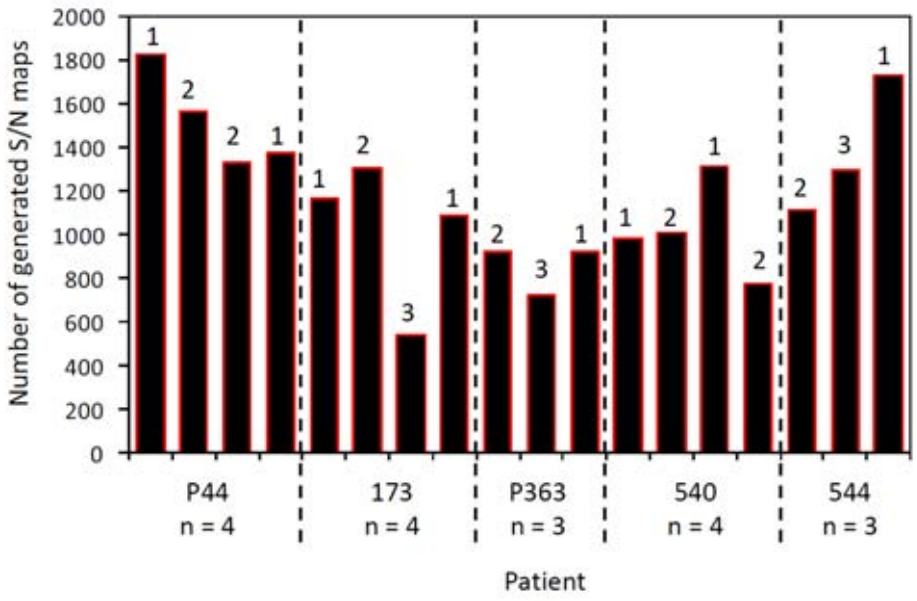

Supplementary Figure 7.4. High resolution haematoxylin and eosin (H&E) stain for patient 540.



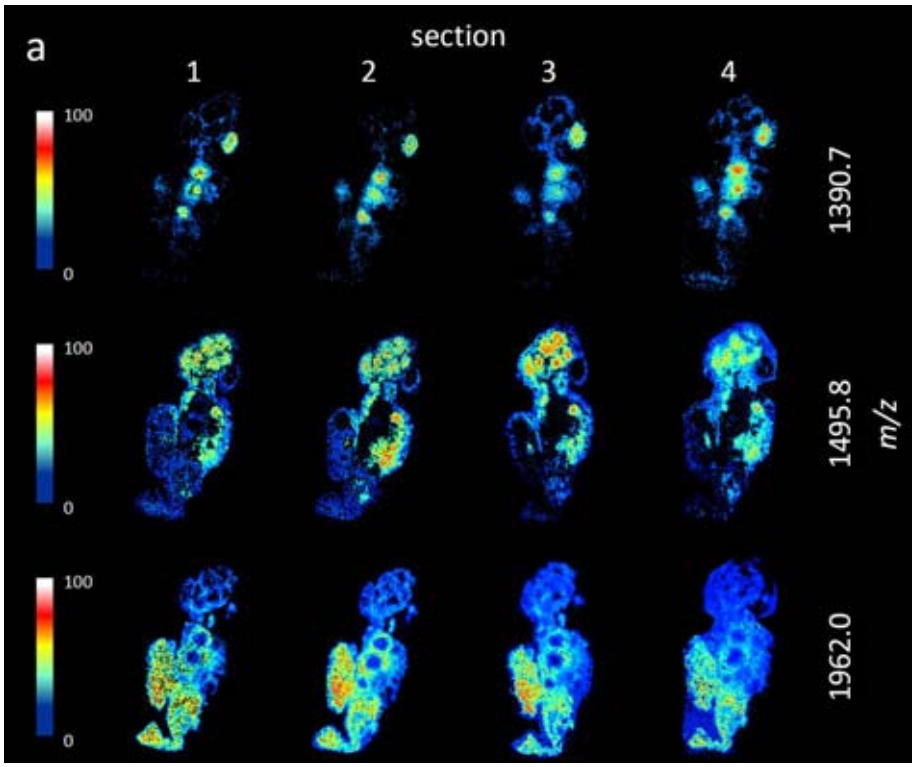

Supplementary Figure 7.5. High resolution haematoxylin and eosin (H&E) stain for patient 544.

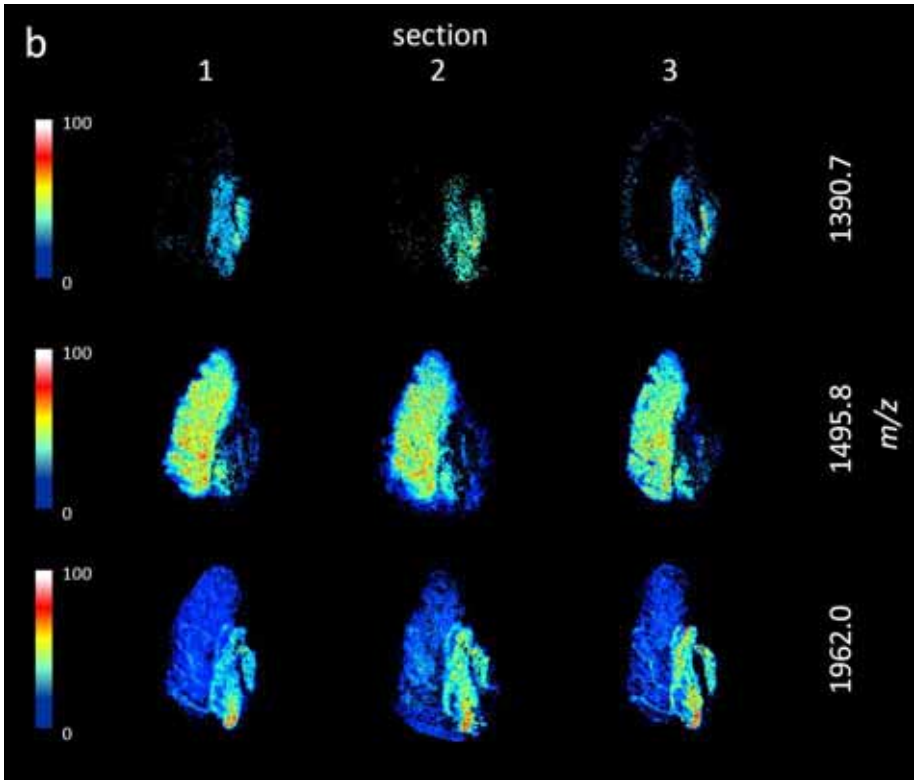



Supplementary Figure 7.6. Haematoxylin and eosin (H&E) stains for the five patient group (patient numbers 44, 173, 363, 540 and 544) analyzed by MALDI-IMS are shown above. The H&E images are overlayed with annotations made by a pathologist to indicate the high grade cancer component of the tissue.

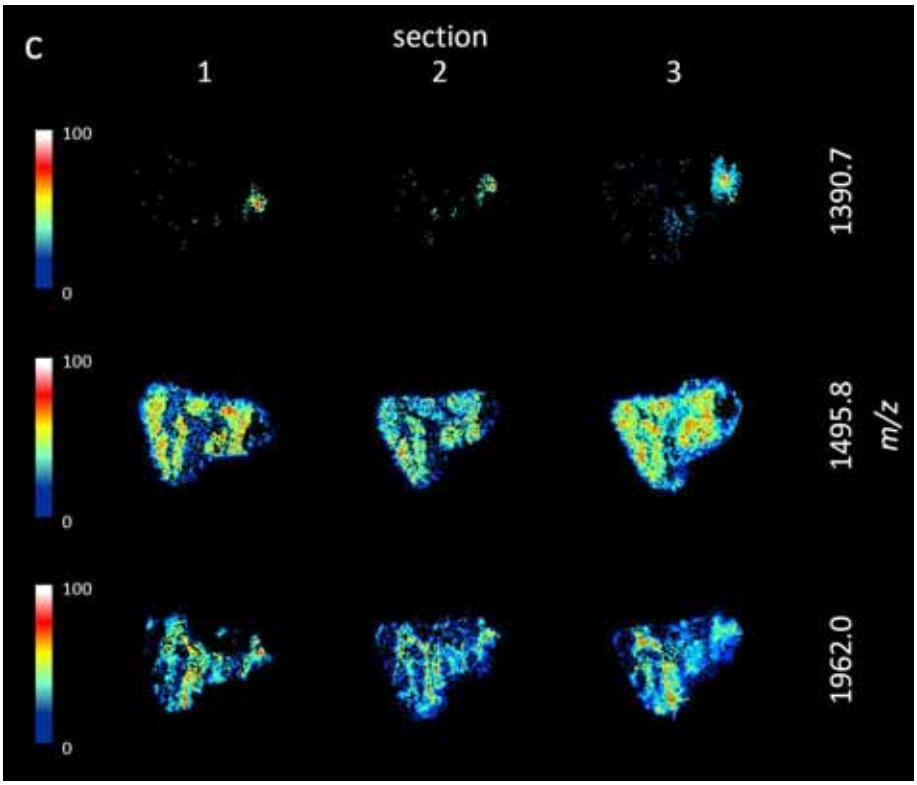


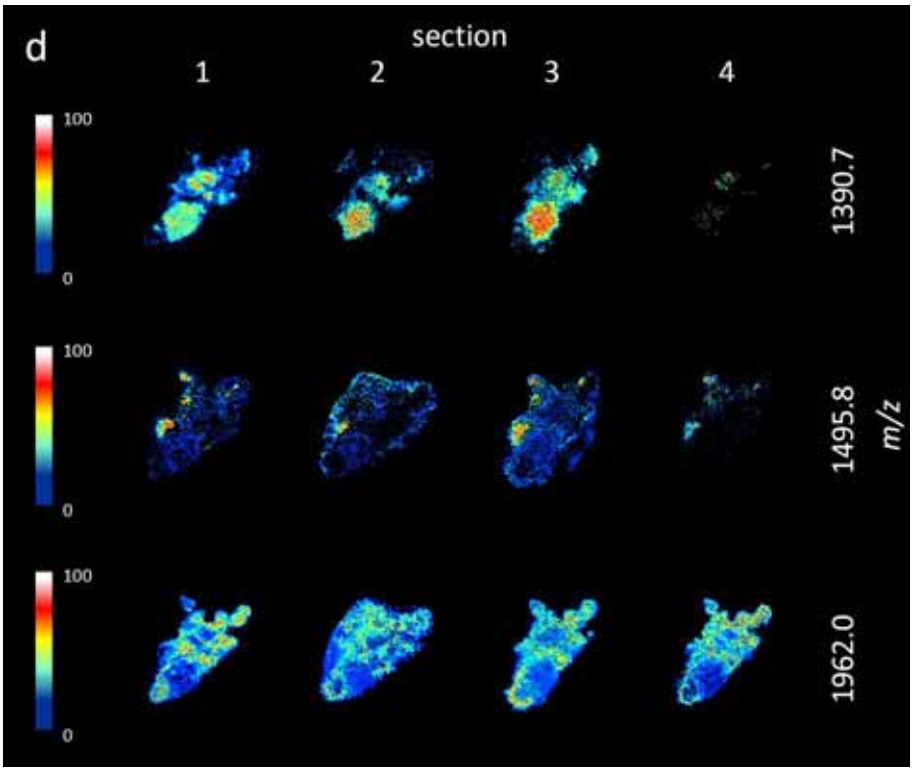

Supplementary Figure 7.7. The number of S/N maps created for each of the ovarian cancer patient sections. The number above each column indicates the section position on the slide. Patient replicate numbers are given on the x-axis.



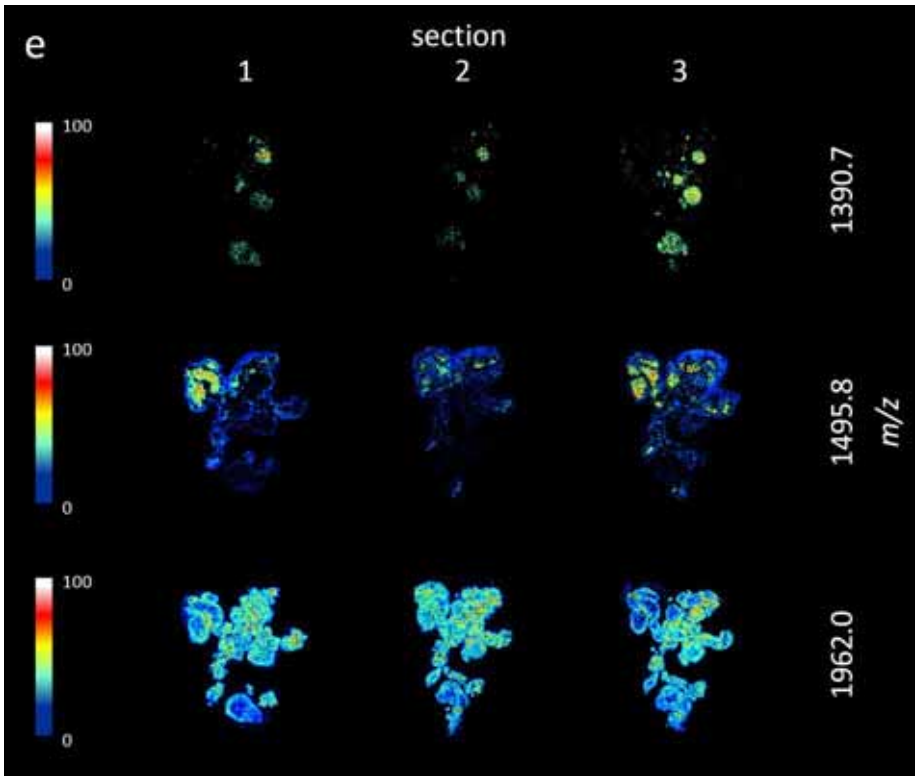




Supplementary Figure 7.8. Tryptic peptide MALDI-IMS data was used to generate signal to noise maps (IonMapper) for peak groups 1390.7, 1495.8 and 1962.0 across the replicate sections analyzed by MALDI-IMS for patients 44 (a), 173 (b), 363, (c), 540 (d) and 544 (e). Scale bars are 2 mm and intensity scales are included.




Supplementary Figure 7.8. Tryptic peptide MALDI-IMS data was used to generate signal to noise maps (IonMapper) for peak groups 1390.7, 1495.8 and 1962.0 across the replicate sections analyzed by MALDI-IMS for patients 44 (a), 173 (b), 363, (c), 540 (d) and 544 (e). Scale bars are 2 mm and intensity scales are included.

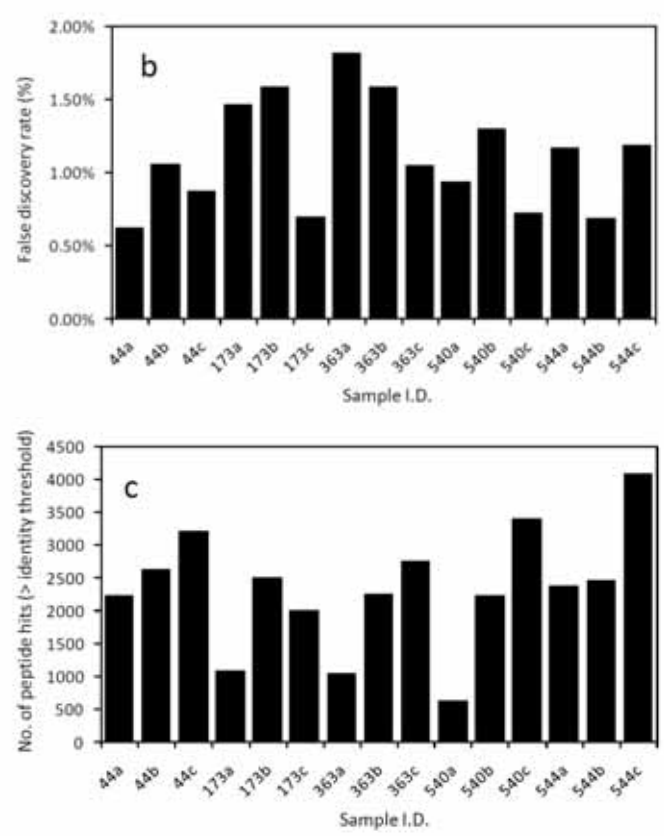



Supplementary Figure 7.8. Tryptic peptide MALDI-IMS data was used to generate signal to noise maps (IonMapper) for peak groups 1390.7, 1495.8 and 1962.0 across the replicate sections analyzed by MALDI-IMS for patients 44 (a), 173 (b), 363, (c), 540 (d) and 544 (e). Scale bars are 2 mm and intensity scales are included.

.

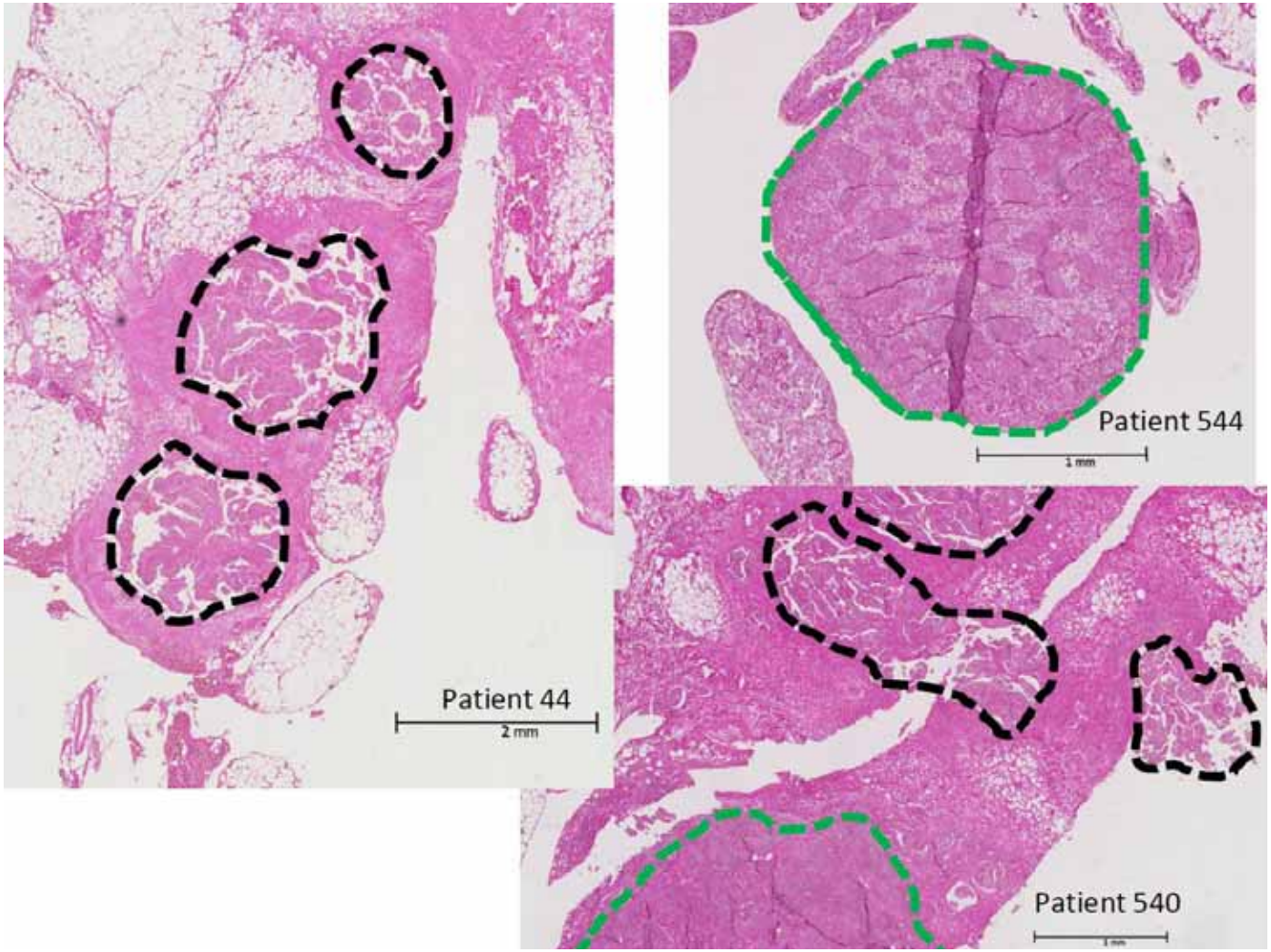
a

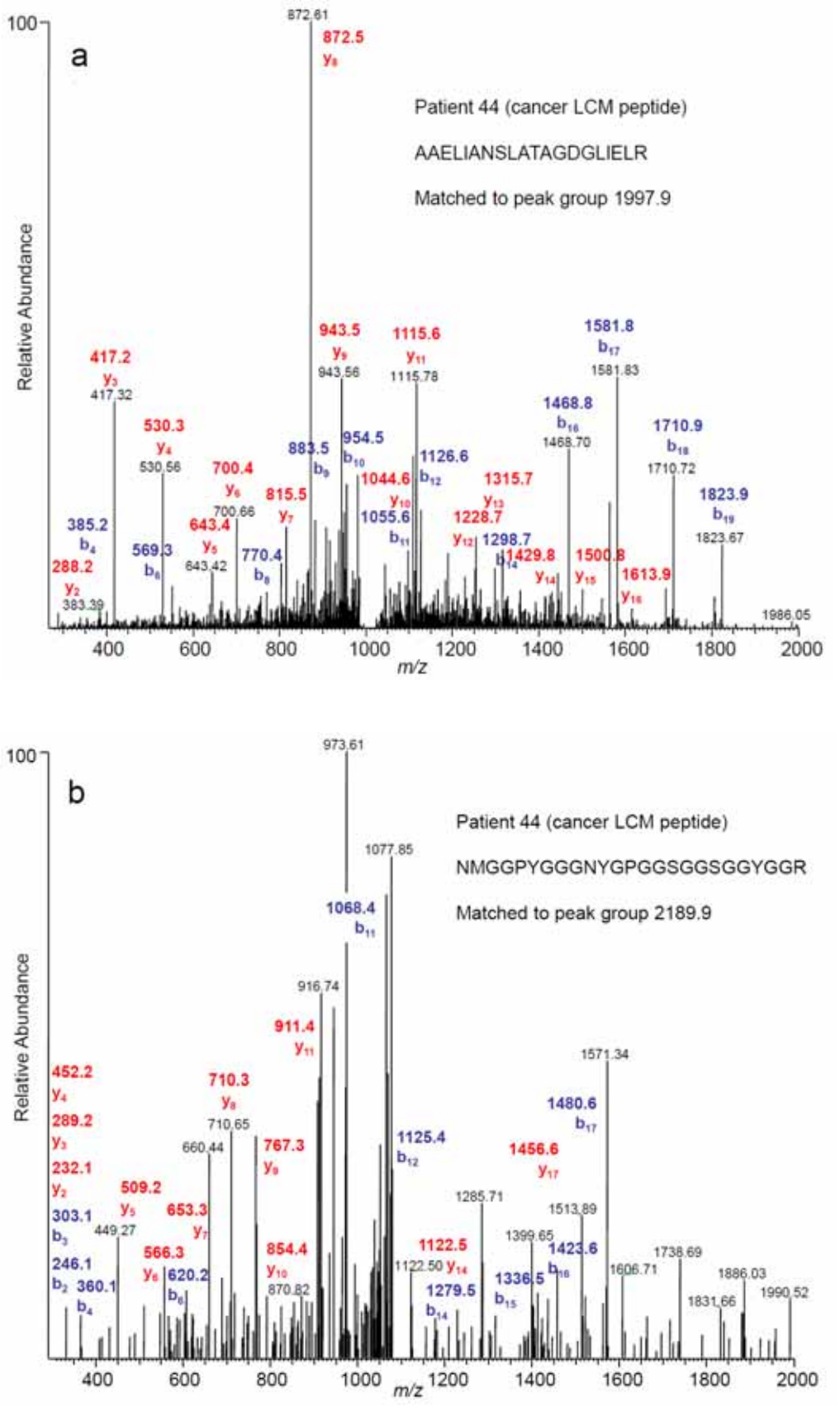
| Patient | Tissue region<br>a - adipose<br>b - stroma<br>c - cancer | False discovery rate | No. hits > Identity threshold | No. hits > Homology threshold |
|---|---|---|---|---|
| 44 | a | 0.62% | 2246 | 3200 |
| | b | 1.06% | 2637 | 3597 |
| | c | 0.87% | 3224 | 4496 |
| 173 | a | 1.47% | 1089 | 1571 |
| | b | 1.59% | 2512 | 3467 |
| | c | 0.70% | 2005 | 2600 |
| 363 | a | 1.82% | 1045 | 1487 |
| | b | 1.59% | 2264 | 3058 |
| | c | 1.05% | 2765 | 3922 |
| 540 | a | 0.94% | 635 | 960 |
| | b | 1.30% | 2238 | 3152 |
| | c | 0.73% | 3418 | 4776 |
| 544 | a | 1.17% | 2383 | 3351 |
| | b | 0.69% | 2475 | 3551 |
| | c | 1.19% | 4111 | 5758 |



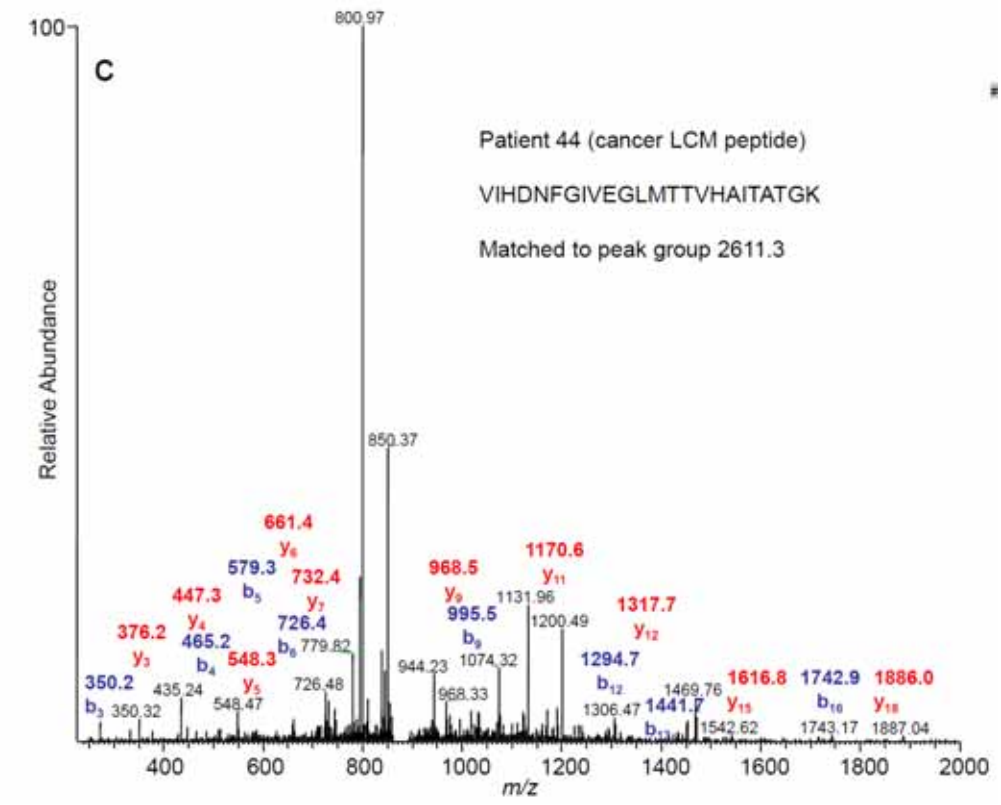

Supplementary Figure 7.9. The table above in panel **a** contains a table of the false discovery rate (%) and number of peptide hits above the identity and homology thresholds for each patient sample MASCOT search result. Panel **b** and **c** contain graphs of the false discovery rate and number of peptide hits above the identity threshold respectively.

Supplementary Figure 7.10. Haematoxylin and eosin stains for patients 44, 540 and 544 showing regions with similar morphology. Regions outlined in black are similar in patients 44 and 540, while regions outlined in green are similar in patients 540 and 544.






Supplementary Figure 7.11. MS/MS spectra for the peptides matched to peak groups 1997.9 (**a**), 2189.9 (**b**) and 2611.3 (**c**). Each spectrum is annotated with the matched b and y ions, the patient and laser capture micro-dissection tissue source as well as the peptide sequence.




Supplementary Figure 7.11. MS/MS spectra for the peptides matched to peak groups 1997.9 (**a**), 2189.9 (**b**) and 2611.3 (**c**). Each spectrum is annotated with the matched b and y ions, the patient and laser capture micro-dissection tissue source as well as the peptide sequence.

# Chapter 8 Concluding remarks

The original aims of this thesis, as outlined in **chapter 1**, were as follows:

1. Develop MALDI-IMS into a platform for reliably tracking protein ion changes across sections derived from frozen tissue.
2. Apply developed MALDI-IMS methods to formalin-fixed paraffin-embedded (FFPE) tissue and optimize to achieve a reliable platform for imaging FFPE tissues.
3. Apply LC-MS/MS for the identification of protein/peptide ions of interest.
4. Application of developed MALDI-IMS methods to the characterization of peritoneal metastases of serous epithelial ovarian cancer.

## 8.1 Aim one

In **chapter 3**, the proof-of-principle application of MALDI-IMS technology showed that proteins could be tracked across murine brain sections sourced from frozen tissue blocks. This was significant as it confirmed that similar to previous studies, protein ion distributions could be used to annotate tissue sections. This work was subsequently published as part of a special issue in the Journal of Proteomics and Bioinformatics **[40]**. As a follow on from this proof-of-principle chapter, **chapter 4** described method developments made to available MALDI-IMS workflows to demonstrate that meaningful protein ion distributions could be generated on frozen murine brain sections. It was subsequently shown that experiments at the tryptic peptide level for MALDI-IMS were beneficial for numerous reasons, including improved ionization efficiencies and easier downstream identification of the peptides analyzed. Aim one was completed in **chapter 4** following demonstration that tryptic peptide MALDI-IMS could be used to

generate reproducible sum spectra and ion intensity maps for consecutive murine brain sections and ovarian tumour sections.

## 8.2 Aim two

Completion of aim one for tryptic peptide MALDI-IMS was achieved in parallel to aim two; the development of methods for analyzing FFPE tissue sections. A previously described CAAR method for partially reversing formalin induced protein-protein cross-links was modified for application to tryptic peptide MALDI-IMS in **chapter 5**. Using printed tryptic digest methods (300 μm centre to centre) developed as part of **chapter 4**, it was demonstrated that the CAAR method was superior to published Tris/EDTA methods of antigen retrieval in a 17 year old ovarian cancer specimen. Furthermore, it was shown that by employing the ImagePrep station, higher resolution (100 μm centre to centre) tryptic peptide MALDI-IMS could be performed reproducibly across FFPE sections of ovarian cancer. Aim two was therefore achieved in **chapter 5** and subsequently published in the Journal of Proteome Research **[64]**.

## 8.3 Aim three

Aim three was addressed in chapters **5**, **6** and **7**. Initial matching of tryptic peptides mapped using MALDI-IMS to extracted peptides identified by LC-MS/MS was achieved in **chapter 5**. However, because the *m/z* deviation within a tissue section for a given peptide species was not known, **chapter 6** described the development of in-house scripts for extracting peak lists containing *m/z* and intensity of mono-isotopic peptide peaks using flexAnalysis (Bruker Daltonics) were written. Peak groups were

assigned from these peak lists and it was possible to rapidly analyze the *m/z* deviation for a single peptide species across a tissue. With the ability to analyze peptide MALDI-IMS data in this way it was shown that the deviation within a data set could be reduced by more than half (from 54 to 20 ppm in the example given in **chapter 6**) for a typical peptide peak group following re-calibration using a set of optimized internal calibrants. The improved peptide peak group AWM *m/z* value following re-calibration resulted in increased confidence for subsequent matches between tryptic peptide MALDI-IMS data and LC-MS/MS data in **chapter 7**. Aim three was therefore achieved in conjunction with aim four, which involved application of the methods developed in **chapters 5** and **6** to metastatic peritoneal tumours of ovarian origin.

## 8.4 Aim four

Because aims one to three were achieved, it was possible to move directly to application of the developed methods for tryptic peptide MALDI-IMS. **Chapter 7** outlined the application of these methods to ovarian peritoneal metastases in an attempt to identify peptide peak groups discriminatory for cancer and stroma. K-means clustering was used as an exploratory method for identifying spectra with similar peak group features. Those features contributing greater than 80% of their intensity to either the cancer or stroma cluster spectra were designated as peak groups of interest. These peak groups were matched to LC-MS/MS identified peptides sourced from LCM isolated tissue. A short list of peptides belonging to prohibitin, hnRNP A2/B1 and cytokeratin 7 were found to be discriminatory for cancer tissue in a subset of the five patients analyzed. One peptide from Glyceraldehyde-3-phosphate dehydrogenase was found to be discriminatory for stroma. Using the FWHM *m/z* deviation (see **chapter 6**), the *m/z* error between the MALDI-IMS peak groups and the LC-MS/MS data was extracted. It was found that this error varied between 14 and 74 ppm depending on the peak group, with hnRNP and Glyceraldehyde-3-phosphate

dehydrogenase showing the most convincing match with errors of 19 and 14 ppm respectively. This observation demonstrated the need to further develop the data processing tools presented here to not only align spectra to each other, but re-calibrate the whole data set based on the presence of internal calibrants. Only then could users of this method be confident that the matches between LC-MS/MS identified peptides and MALDI-IMS are as robust as possible.

Aim four was thus also achieved, given successful identification of peptides which could be used for discriminating cancer from stroma in the patients analyzed. Importantly, this was achieved using a complete pre-IHC validation MALDI-IMS workflow, including tryptic peptide MALDI-IMS, spectral processing, statistical analyses (k-means clustering) and downstream matching of peptide peak groups of interest to LC-MS/MS identified peptides.

## 8.5 Conclusions and future prospects

The sum of the developed methods and data presented in this thesis demonstrated the advantages of tryptic peptide MALDI-IMS over protein MALDI-IMS, including the ability to access large FFPE sample archives through the use of CAAR. Crucially, it was demonstrated that efficient data analysis for tryptic peptide MALDI-IMS data could be achieved firstly by extracting peak lists from the MS data using proprietary software. In the absence of suitable proprietary or freeware software platforms for downstream analysis, these peak lists were processed by an in-house developed software tool, IonMapper, which could be used to automatically generate thousands of S/N maps. Because of the simple file formats used (ASCII text), the data contained in these files could be used directly by collaborating groups to perform k-means clustering. This analysis provided the opportunity to identify tryptic peptides which were discriminatory for cancer and stroma in peritoneal ovarian metastases.

The complexities of developing a complete method for characterizing a human cancer by spatial proteomics meant that the methods developed here are the initial framework for a more robust workflow which can utilize tryptic peptides for prediction of patient prognosis or response to chemotherapy.

In future it is envisioned that continued development of the IonMapper tool will yield a complete software platform for MALDI-IMS data analysis. This platform would combine features such as the flexImaging sum spectrum with a complete interactive interface to allow, for example, single click generation of histograms for *m/z* deviation and grouping of S/N maps with similar distributions. It is also expected that collaborations with other departments will assist in the application of spectral alignment algorithms that can be used to align data sets in combination with calibration. This step will be instrumental in providing confident assignment of tryptic peptide identity based on LC-MS/MS data. Furthermore, these alignments will be crucial in the development of classifiers, which will rely on the similarity of a given peptide feature across many patient tumour samples. Finally, it is expected that the findings of this project will be used to continually improve upon the sample preparation and MS acquisition procedures used for MALDI-IMS. In doing so, it will be possible to provide partial sequence information for peptides of interest by *in situ* MS/MS. This information would be used in conjunction with the alignment and calibration of spectra to confirm peptide identifications. Furthermore, the application of these methods to larger patient data sets, and ultimately tissue micro-arrays, will ensure a large number of biological replicates, greater confidence in the importance of identified peak groups of interest and will compensate for the tissue heterogeneity discussed in the previous chapters.

Ultimately, these future aims would be impossible without the developments described in this thesis. The complete tryptic peptide MALDI-IMS workflow described here can in principle be applied to any human cancer, and as such will form the framework for future MALDI-IMS research at the Adelaide Proteomics Centre for years to come.

# References


1. Hennessy, B.T., R.L. Coleman, and M. Markman, *Ovarian cancer.* Lancet, 2009. **374**(9698): p. 1371-82.
2. Menon, U., et al., *Sensitivity and specificity of multimodal and ultrasound screening for ovarian cancer, and stage distribution of detected cancers: results of the prevalence screen of the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS).* Lancet Oncol, 2009. **10**(4): p. 327-40.
3. Moore, R.G., S. MacLaughlan, and R.C. Bast, Jr., *Current state of biomarker development for clinical application in epithelial ovarian cancer.* Gynecol Oncol, 2010. **116**(2): p. 240-5.
4. Mor, G., et al., *Serum protein markers for early detection of ovarian cancer.* Proc Natl Acad Sci U S A, 2005. **102**(21): p. 7677-82.
5. Muller, C.Y., *Doctor, should I get this new ovarian cancer test-OVA1?* Obstet Gynecol, 2010. **116**(2 Pt 1): p. 246-7.
6. Anderson, N.L. and N.G. Anderson, *The human plasma proteome: history, character, and diagnostic prospects.* Mol Cell Proteomics, 2002. **1**(11): p. 845-67.
7. Poschmann, G., et al., *Cell-based proteome analysis: the first stage in the pipeline for biomarker discovery.* Biochim Biophys Acta, 2009. **1794**(9): p. 1309-16.
8. Eriksson, J. and D. Fenyo, *Improving the success rate of proteome analysis by modeling protein-abundance distributions and experimental designs.* Nat Biotechnol, 2007. **25**(6): p. 651-5.
9. Benedet, J.L., et al., *FIGO staging classifications and clinical practice guidelines in the management of gynecologic cancers. FIGO Committee on Gynecologic Oncology.* Int J Gynaecol Obstet, 2000. **70**(2): p. 209-62.
10. SurveillanceEpidemiologyandEndResults. *Cancer of the Ovary - SEER Stat Fact Sheets.* 2010; Available from: http://seer.cancer.gov/statfacts/html/ovary.html.
11. Kosary, C.L., *FIGO stage, histology, histologic grade, age and race as prognostic factors in determining survival for cancers of the female gynecological system: an analysis of 1973-87 SEER cases of cancers of the endometrium, cervix, ovary, vulva, and vagina.* Semin Surg Oncol, 1994. **10**(1): p. 31-46.
12. Shimizu, Y., et al., *Toward the development of a universal grading system for ovarian epithelial carcinoma: testing of a proposed system in a series of 461 patients with uniform treatment and follow-up.* Cancer, 1998. **82**(5): p. 893-901.

13. Sagae, S., et al., *The reproducibility of a binary tumor grading system for uterine endometrial endometrioid carcinoma, compared with FIGO system and nuclear grading.* Oncology, 2004. **67**(5-6): p. 344-50.
14. Vang, R., M. Shih Ie, and R.J. Kurman, *Ovarian low-grade and high-grade serous carcinoma: pathogenesis, clinicopathologic and molecular biologic features, and diagnostic problems.* Adv Anat Pathol, 2009. **16**(5): p. 267-82.
15. Ali-Fehmi, R., et al., *Molecular typing of epithelial ovarian carcinomas using inflammatory markers.* Cancer, 2010.
16. Cho, K.R., *Ovarian cancer update: lessons from morphology, molecules, and mice.* Arch Pathol Lab Med, 2009. **133**(11): p. 1775-81.
17. Lu, P., et al., *Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation.* Nat Biotechnol, 2007. **25**(1): p. 117-24.
18. Gygi, S.P., et al., *Correlation between protein and mRNA abundance in yeast.* Mol Cell Biol, 1999. **19**(3): p. 1720-30.
19. Kim, H., et al., *Comparative proteomic analysis of low stage and high stage endometrioid ovarian adenocarcinomas.* Proteomics Clin Appl, 2008. **2**(4): p. 571-584.
20. Washburn, M.P., D. Wolters, and J.R. Yates, 3rd, *Large-scale analysis of the yeast proteome by multidimensional protein identification technology.* Nat Biotechnol, 2001. **19**(3): p. 242-7.
21. Wilkins, M.R., et al., *From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis.* Biotechnology (N Y), 1996. **14**(1): p. 61-5.
22. Mitulovic, G. and K. Mechtler, *HPLC techniques for proteomics analysis--a short overview of latest developments.* Brief Funct Genomic Proteomic, 2006. **5**(4): p. 249-60.
23. Condina, M.R., et al., *EZYprep LC-coupled MALDI-TOF/TOF MS: an improved matrix spray application for phosphopeptide characterisation.* Proteomics, 2010. **10**(13): p. 2516-30.
24. Zhu, Y., et al., *Classifications of ovarian cancer tissues by proteomic patterns.* Proteomics, 2006. **6**(21): p. 5846-56.
25. Caprioli, R.M., T.B. Farmer, and J. Gile, *Molecular imaging of biological samples: localization of peptides and proteins using MALDI-TOF MS.* Anal Chem, 1997. **69**(23): p. 4751-60.
26. Amstalden van Hove, E.R., D.F. Smith, and R.M. Heeren, *A concise review of mass spectrometry imaging.* J Chromatogr A, 2010. **1217**(25): p. 3946-54.
27. Meriaux, C., et al., *Liquid ionic matrixes for MALDI mass spectrometry imaging of lipids.* J Proteomics, 2010. **73**(6): p. 1204-18.

28. Lemaire, R., et al., *Specific MALDI imaging and profiling for biomarker hunting and validation: fragment of the 11S proteasome activator complex, Reg alpha fragment, is a new potential ovary cancer biomarker.* J Proteome Res, 2007. **6**(11): p. 4127-34.
29. Kang, S., et al., *Molecular proteomics imaging of tumor interfaces by mass spectrometry.* J Proteome Res, 2010. **9**(2): p. 1157-64.
30. El Ayed, M., et al., *MALDI imaging mass spectrometry in ovarian cancer for tracking, identifying, and validating biomarkers.* Med Sci Monit, 2010. **16**(8): p. BR233-45.
31. Franck, J., et al., *MALDI mass spectrometry imaging of proteins exceeding 30,000 daltons.* Med Sci Monit, 2010. **16**(9): p. BR293-9.
32. Karas, M., M. Gluckmann, and J. Schafer, *Ionization in matrix-assisted laser desorption/ionization: singly charged molecular ions are the lucky survivors.* J Mass Spectrom, 2000. **35**(1): p. 1-12.
33. Knochenmuss, R. and R. Zenobi, *MALDI ionization: the role of in-plume processes.* Chem Rev, 2003. **103**(2): p. 441-52.
34. Knochenmuss, R., *Ion formation mechanisms in UV-MALDI.* Analyst, 2006. **131**(9): p. 966-86.
35. Daltonics, B., *MALDI Theory Mass Spectrometry*. 2004, Bremen: Bruker Daltonics. 18.
36. Reyzer, M.L., et al., *Direct analysis of drug candidates in tissue by matrix-assisted laser desorption/ionization mass spectrometry.* J Mass Spectrom, 2003. **38**(10): p. 1081-92.
37. Chen, R., et al., *Three Dimensional Mapping of Neuropeptides and Lipids in Crustacean Brain by Mass Spectral Imaging.* J Am Soc Mass Spectrom, 2009.
38. Taban, I.M., et al., *Imaging of peptides in the rat brain using MALDI-FTICR mass spectrometry.* J Am Soc Mass Spectrom, 2007. **18**(1): p. 145-51.
39. Rauser, S., et al., *Classification of HER2 receptor status in breast cancer tissues by MALDI imaging mass spectrometry.* J Proteome Res, 2010. **9**(4): p. 1854-63.
40. Gustafsson, J.O.R., S.R. McColl, and P. Hoffmann, *Imaging mass spectrometry and its methodological application to Murine tissue.* Journal of Proteomics and Bioinformatics, 2008. **1**(9): p. 458-463.
41. Lemaire, R., et al., *MALDI-MS direct tissue analysis of proteins: Improving signal sensitivity using organic treatments.* Anal Chem, 2006. **78**(20): p. 7145-53.
42. Seeley, E.H., et al., *Enhancement of Protein Sensitivity for MALDI Imaging Mass Spectrometry After Chemical Treatment of Tissue Sections.* J Am Soc Mass Spectrom, 2008.
43. Schwartz, S.A., M.L. Reyzer, and R.M. Caprioli, *Direct tissue analysis using matrix-assisted laser desorption/ionization mass spectrometry: practical aspects of sample preparation.* J Mass Spectrom, 2003. **38**(7): p. 699-708.

44. Chaurand, P., et al., *New developments in profiling and imaging of proteins from tissue sections by MALDI mass spectrometry.* J Proteome Res, 2006. **5**(11): p. 2889-900.
45. Ilina, E.N., et al., *Direct bacterial profiling by matrix-assisted laser desorption-ionization time-of-flight mass spectrometry for identification of pathogenic Neisseria.* J Mol Diagn, 2009. **11**(1): p. 75-86.
46. Yanagisawa, K., et al., *Proteomic patterns of tumour subsets in non-small-cell lung cancer.* Lancet, 2003. **362**(9382): p. 433-9.
47. Caldwell, R.L., G.E. Holt, and R.M. Caprioli, *Tissue Profiling by MALDI Mass Spectrometry Distinguishes Clinical Grades of Soft Tissue Sarcomas.* Cancer Genomics and Proteomics, 2005. **2**: p. 333-346.
48. Chaurand, P., et al., *Integrating histology and imaging mass spectrometry.* Anal Chem, 2004. **76**(4): p. 1145-55.
49. Samsi, S.S., et al., *Imaging mass spectrometry analysis for follicular lymphoma grading.* Conf Proc IEEE Eng Med Biol Soc, 2009. **2009**: p. 6969-72.
50. Rahman, S.M., et al., *Proteomic patterns of preinvasive bronchial lesions.* Am J Respir Crit Care Med, 2005. **172**(12): p. 1556-62.
51. Schwamborn, K., et al., *Identifying prostate carcinoma by MALDI-Imaging.* Int J Mol Med, 2007. **20**(2): p. 155-9.
52. Cazares, L.H., et al., *Imaging mass spectrometry of a specific fragment of mitogen-activated protein kinase/extracellular signal-regulated kinase kinase kinase 2 discriminates cancer from uninvolved prostate tissue.* Clin Cancer Res, 2009. **15**(17): p. 5541-51.
53. Agar, N.Y., et al., *Imaging of meningioma progression by matrix-assisted laser desorption ionization time-of-flight mass spectrometry.* Anal Chem, 2010. **82**(7): p. 2621-5.
54. Schwartz, S.A., et al., *Proteomic-based prognosis of brain tumor patients using direct-tissue matrix-assisted laser desorption ionization mass spectrometry.* Cancer Res, 2005. **65**(17): p. 7674-81.
55. Rompp, A., et al., *Histology by mass spectrometry: label-free tissue characterization obtained from high-accuracy bioanalytical imaging.* Angew Chem Int Ed Engl, 2010. **49**(22): p. 3834-8.
56. Jardin-Mathe, O., et al., *MITICS (MALDI Imaging Team Imaging Computing System): a new open source mass spectrometry imaging software.* J Proteomics, 2008. **71**(3): p. 332-45.
57. Martens, L., et al., *mzML - a Community Standard for Mass Spectrometry Data.* Mol Cell Proteomics, 2010.
58. Rompp, A., et al., *imzML: Imaging Mass Spectrometry Markup Language: A Common Data Format for Mass Spectrometry Imaging.* Methods Mol Biol, 2011. **696**: p. 205-24.

59. Aerni, H.R., D.S. Cornett, and R.M. Caprioli, *Automated acoustic matrix deposition for MALDI sample preparation.* Anal Chem, 2006. **78**(3): p. 827-34.
60. Franck, J., et al., *Improving tissue preparation for matrix-assisted laser desorption ionization mass spectrometry imaging. Part 1: using microspotting.* Anal Chem, 2009. **81**(19): p. 8193-202.
61. Lemaire, R., et al., *Solid ionic matrixes for direct tissue analysis and MALDI imaging.* Anal Chem, 2006. **78**(3): p. 809-19.
62. Groseclose, M.R., et al., *High-throughput proteomic analysis of formalin-fixed paraffin-embedded tissue microarrays using MALDI imaging mass spectrometry.* Proteomics, 2008. **8**(18): p. 3715-24.
63. Groseclose, M.R., et al., *Identification of proteins directly from tissue: in situ tryptic digestions coupled with imaging mass spectrometry.* J Mass Spectrom, 2007. **42**(2): p. 254-62.
64. Gustafsson, J.O., et al., *Citric Acid Antigen Retrieval (CAAR) for Tryptic Peptide Imaging Directly on Archived Formalin-Fixed Paraffin-Embedded Tissue.* J Proteome Res, 2010. **9**(9): p. 4315-28.
65. Alexandrov, T., et al., *Spatial Segmentation of Imaging Mass Spectrometry Data with Edge-Preserving Image Denoising and Clustering.* J Proteome Res, 2010.
66. Oppenheimer, S.R., et al., *Molecular analysis of tumor margins by MALDI mass spectrometry in renal carcinoma.* J Proteome Res, 2010. **9**(5): p. 2182-90.
67. Woo, M.M., et al., *Gain of OGP, an estrogen-regulated oviduct-specific glycoprotein, is associated with the development of endometrial hyperplasia and endometrial cancer.* Clin Cancer Res, 2004. **10**(23): p. 7958-64.
68. Elg, S.A., et al., *Alpha-1 acid glycoprotein is an immunosuppressive factor found in ascites from ovaria carcinoma.* Cancer, 1997. **80**(8): p. 1448-56.
69. Addis, M.F., et al., *Generation of high-quality protein extracts from formalin-fixed, paraffin-embedded tissues.* Proteomics, 2009. **9**(15): p. 3815-23.
70. Addis, M.F., et al., *2-D PAGE and MS analysis of proteins from formalin-fixed, paraffin-embedded tissues.* Proteomics, 2009. **9**(18): p. 4329-39.
71. Grantzdorffer, I., et al., *Comparison of different tissue sampling methods for protein extraction from formalin-fixed and paraffin-embedded tissue specimens.* Exp Mol Pathol, 2010. **88**(1): p. 190-6.
72. Hood, B.L., T.P. Conrads, and T.D. Veenstra, *Mass spectrometric analysis of formalin-fixed paraffin-embedded tissue: unlocking the proteome within.* Proteomics, 2006. **6**(14): p. 4106-14.

73. Aoki, Y., et al., *A novel method for analyzing formalin-fixed paraffin embedded (FFPE) tissue sections by mass spectrometry imaging.* Proc Jpn Acad Ser, 2007. **83**: p. 205-214.
74. Lemaire, R., et al., *Direct analysis and MALDI imaging of formalin-fixed, paraffin-embedded tissue sections.* J Proteome Res, 2007. **6**(4): p. 1295-305.
75. Stauber, J., et al., *MALDI Imaging of Formalin-Fixed Paraffin-Embedded Tissues: Application to Model Animals of Parkinson Disease for Biomarker Hunting.* J Proteome Res, 2008.
76. R Development Core Team, *R: A language environment for statistical computing.* 2010: Vienna, Austria.
77. Shi, S.R., et al., *Antigen retrieval technique utilizing citrate buffer or urea solution for immunohistochemical demonstration of androgen receptor in formalin-fixed paraffin sections.* J Histochem Cytochem, 1993. **41**(11): p. 1599-604.
78. Ricciardelli, C., et al., *Androgen receptor levels in prostate cancer epithelial and peritumoral stromal cells identify non-organ confined disease.* Prostate, 2005. **63**(1): p. 19-28.
79. Zhang, X., et al., *An improved method of sample preparation on AnchorChip targets for MALDI-MS and MS/MS and its application in the liver proteome project.* Proteomics, 2007. **7**(14): p. 2340-9.
80. Koch, I., *Analysis of Multivariate and High-Dimensional Data, Theory and Practice, Monograph*. 2012, Cambridge University Press.
81. Schmidt, A., J. Kellermann, and F. Lottspeich, *A novel strategy for quantitative proteomics using isotope-coded protein labels.* Proteomics, 2005. **5**(1): p. 4-15.
82. Vanrobaeys, F., et al., *Profiling of myelin proteins by 2D-gel electrophoresis and multidimensional liquid chromatography coupled to MALDI TOF-TOF mass spectrometry.* J Proteome Res, 2005. **4**(6): p. 2283-93.
83. Baluya, D.L., T.J. Garrett, and R.A. Yost, *Automated MALDI matrix deposition method with inkjet printing for imaging mass spectrometry.* Anal Chem, 2007. **79**(17): p. 6862-7.
84. Dekeyser, S.S., et al., *Imaging mass spectrometry of neuropeptides in decapod crustacean neuronal tissues.* J Proteome Res, 2007. **6**(5): p. 1782-91.
85. Monroe, E.B., et al., *SIMS and MALDI MS imaging of the spinal cord.* Proteomics, 2008. **8**: p. 1-9.
86. Herring, K.D., S.R. Oppenheimer, and R.M. Caprioli, *Direct tissue analysis by matrix-assisted laser desorption ionization mass spectrometry: application to kidney biology.* Semin Nephrol, 2007. **27**(6): p. 597-608.

87. Bunch, J., M.R. Clench, and D.S. Richards, *Determination of pharmaceutical compounds in skin by imaging matrix-assisted laser desorption/ionisation mass spectrometry.* Rapid Commun Mass Spectrom, 2004. **18**(24): p. 3051-60.

88. Pierson, J., et al., *Molecular profiling of experimental Parkinson's disease: direct analysis of peptides and proteins on brain tissue sections by MALDI mass spectrometry.* J Proteome Res, 2004. **3**(2): p. 289-95.

89. McDonnell, L.A., et al., *Subcellular imaging mass spectrometry of brain tissue.* J Mass Spectrom, 2005. **40**(2): p. 160-8.

90. Meistermann, H., et al., *Biomarker discovery by imaging mass spectrometry: transthyretin is a biomarker for gentamicin-induced nephrotoxicity in rat.* Mol Cell Proteomics, 2006. **5**(10): p. 1876-86.

91. Chaurand, P., et al., *Profiling and imaging proteins in the mouse epididymis by imaging mass spectrometry*. Proteomics, 2003. **3**(11): p. 2221-39.

92. Deutskens, F., J. Yang, and R.M. Caprioli, *High spatial resolution imaging mass spectrometry and classical histology on a single tissue section.* J Mass Spectrom, 2011. **46**(6): p. 568-71.

93. Gustafsson, J.O., et al., *MALDI Imaging Mass Spectrometry (MALDI-IMS)-Application of Spatial Proteomics for Ovarian Cancer Classification and Diagnosis.* Int J Mol Sci, 2011. **12**(1): p. 773-94.

94. Gustafsson, J.O.R., *Tracking protein expression changes in experimental autoimmune encephalomyelitis*, in *School of Molecular and Biomedical Science*. 2007, The University of Adelaide: Adelaide.

95. Daltonics, B., *ImagePrep 2.0 User Manual*. Vol. 3. 2010, Bremen: Bruker Daltonics, Bremen, Germany. 68.

96. Sidman, R.L., et al. *High Resolution Mouse Brain Atlas*. 1999; Available from: http://www.hms.harvard.edu/research/brain.

97. Strader, M.B., et al., *Efficient and specific trypsin digestion of microgram to nanogram quantities of proteins in organic-aqueous solvent systems.* Anal Chem, 2006. **78**(1): p. 125-34.

98. Ahmed, N., et al., *Proteomic tracking of serum protein isoforms as screening biomarkers of ovarian cancer.* Proteomics, 2005. **5**(17): p. 4625-36.

99. Hood, B.L., et al., *Proteomic analysis of formalin-fixed prostate cancer tissue.* Mol Cell Proteomics, 2005. **4**(11): p. 1741-53.

100. Palmer-Toy, D.E., et al., *Efficient method for the proteomic analysis of fixed and embedded tissues.* J Proteome Res, 2005. **4**(6): p. 2404-11.

101. Dekker, L.J., et al., *A mass spectrometry based imaging method developed for the intracellular detection of HIV protease inhibitors.* Rapid Commun Mass Spectrom, 2009. **23**(8): p. 1183-8.
102. Trim, P.J., et al., *Matrix-assisted laser desorption/ionisation mass spectrometry imaging of lipids in rat brain tissue with integrated unsupervised and supervised multivariant statistical analysis.* Rapid Commun Mass Spectrom, 2008. **22**(10): p. 1503-9.
103. McDonnell, L.A. and R.M. Heeren, *Imaging mass spectrometry.* Mass Spectrom Rev, 2007. **26**(4): p. 606-43.
104. Ronci, M., et al., *Protein unlocking procedures of formalin-fixed paraffin-embedded tissues: application to MALDI-TOF imaging MS investigations.* Proteomics, 2008. **8**(18): p. 3702-14.
105. Ricciardelli, C., et al., *Elevated levels of HER-2/neu and androgen receptor in clinically localized prostate cancer identifies metastatic potential.* Prostate, 2008. **68**(8): p. 830-8.
106. Hazelbag, H.M., et al., *Immunostaining of chain-specific keratins on formalin-fixed, paraffin-embedded tissues: a comparison of various antigen retrieval systems using microwave heating and proteolytic pre-treatments.* J Histochem Cytochem, 1995. **43**(4): p. 429-37.
107. Yamashita, S. and Y. Okada, *Mechanisms of heat-induced antigen retrieval: analyses in vitro employing SDS-PAGE and immunohistochemistry.* J Histochem Cytochem, 2005. **53**(1): p. 13-21.
108. Condina, M.R., et al., *A Sensitive Magnetic Bead Approach for the Detection and Identification of Tyrosine Phosphorylation in proteins by MALDI-TOF/TOF mass spectrometry.* Proteomics, 2009. **Accepted 02-2009**.
109. Ricciardelli, C. and R.J. Rodgers, *Extracellular matrix of ovarian tumors.* Semin Reprod Med, 2006. **24**(4): p. 270-82.
110. Becker, M., et al., *Application Note #MT-97 MALDI Imaging of Proteins at 20µm Resolution on the ultrafleXtreme.* 2009, Bruker Daltonics: Bremen, Germany. p. pp 1-6.
111. Mange, A., et al., *Liquid chromatography-tandem and MALDI imaging mass spectrometry analyses of RCL2/CS100-fixed, paraffin-embedded tissues: proteomics evaluation of an alternate fixative for biomarker discovery.* J Proteome Res, 2009. **8**(12): p. 5619-28.
112. McDonnell, L.A., et al., *Imaging mass spectrometry data reduction: automated feature identification and extraction.* J Am Soc Mass Spectrom, 2010. **21**(12): p. 1969-78.
113. Kaneko, Y., et al., *Imaging mass spectrometry analysis reveals an altered lipid distribution pattern in the tubular areas of hyper-IgA murine kidneys.* Exp Mol Pathol, 2011. **91**(2): p. 614-21.

114. Schober, Y., et al., *Protein identification by accurate mass matrix-assisted laser desorption/ionization imaging of tryptic peptides.* Rapid Commun Mass Spectrom, 2011. **25**(17): p. 2475-83.
115. Kong, X. and C. Reilly, *A Bayesian approach to the alignment of mass spectra.* Bioinformatics, 2009. **25**(24): p. 3213-20.
116. Norris, J.L., et al., *Processing MALDI Mass Spectra to Improve Mass Spectral Direct Tissue Analysis.* Int J Mass Spectrom, 2007. **260**(2-3): p. 212-221.
117. Wong, J.W., G. Cagney, and H.M. Cartwright, *SpecAlign--processing and alignment of mass spectra datasets.* Bioinformatics, 2005. **21**(9): p. 2088-90.
118. Feki, A., et al., *Dissemination of intraperitoneal ovarian cancer: Discussion of mechanisms and demonstration of lymphatic spreading in ovarian cancer model.* Crit Rev Oncol Hematol, 2009. **72**(1): p. 1-9.
119. Lengyel, E., *Ovarian cancer development and metastasis.* Am J Pathol, 2010. **177**(3): p. 1053-64.
120. Ricciardelli, C. and M.K. Oehler, *Diverse molecular pathways in ovarian cancer and their clinical significance.* Maturitas, 2009. **62**(3): p. 270-5.
121. Pomel, C., et al., *Cytoreductive surgery in ovarian cancer.* Cancer Imaging, 2007. **7**: p. 210-5.
122. Kolomainen, D.F. and D.P. Barton, *Surgical management of bowel obstruction in gynaecological malignancies.* Curr Opin Support Palliat Care, 2011. **5**(1): p. 55-9.
123. Kolomainen, D.F., et al., *Outcomes of surgical management of bowel obstruction in relapsed epithelial ovarian cancer (EOC).* Gynecol Oncol, 2011.
124. Hardesty, W.M., et al., *Protein signatures for survival and recurrence in metastatic melanoma.* J Proteomics, 2011. **74**(7): p. 1002-14.
125. Deininger, S.O., et al., *MALDI Imaging Combined with Hierarchical Clustering as a New Tool for the Interpretation of Complex Human Cancers.* J Proteome Res, 2008.
126. Lim, R., et al., *2D-PAGE of ovarian cancer: analysis of soluble and insoluble fractions using medium-range immobilized pH gradients.* Biochem Biophys Res Commun, 2011. **406**(3): p. 408-13.
127. Gregory-Bass, R.C., et al., *Prohibitin silencing reverses stabilization of mitochondrial integrity and chemoresistance in ovarian cancer cells by increasing their sensitivity to apoptosis.* Int J Cancer, 2008. **122**(9): p. 1923-30.
128. Fraser, M., et al., *Chemoresistance in human ovarian cancer: the role of apoptotic regulators.* Reprod Biol Endocrinol, 2003. **1**: p. 66.

129. Chaudhury, A., P. Chander, and P.H. Howe, *Heterogeneous nuclear ribonucleoproteins (hnRNPs) in cellular processes: Focus on hnRNP E1's multifunctional regulatory roles.* RNA, 2010. **16**(8): p. 1449-62.

130. Moran-Jones, K., et al., *hnRNP A2 regulates alternative mRNA splicing of TP53INP2 to control invasive cell migration.* Cancer Res, 2009. **69**(24): p. 9219-27.

131. Chen, M., J. Zhang, and J.L. Manley, *Turning on a fuel switch of cancer: hnRNP proteins regulate alternative splicing of pyruvate kinase mRNA.* Cancer Res, 2010. **70**(22): p. 8977-80.

132. Heatley, M.K., *Immunohistochemical biomarkers of value in distinguishing primary ovarian carcinoma from gastric carcinoma: a systematic review with statistical meta-analysis.* Histopathology, 2008. **52**(3): p. 267-76.

133. Tot, T., *Cytokeratins 20 and 7 as biomarkers: usefulness in discriminating primary from metastatic adenocarcinoma.* Eur J Cancer, 2002. **38**(6): p. 758-63.

134. Colell, A., D.R. Green, and J.E. Ricci, *Novel roles for GAPDH in cell death and carcinogenesis.* Cell Death Differ, 2009. **16**(12): p. 1573-81.

135. Casadonte, R. and R.M. Caprioli, *Proteomic analysis of formalin-fixed paraffin-embedded tissue by MALDI imaging mass spectrometry.* Nat Protoc, 2011. **6**(11): p. 1695-709.

136. Khatun, J., E. Hamlett, and M.C. Giddings, *Incorporating sequence information into the scoring function: a hidden Markov model for improved peptide identification.* Bioinformatics, 2008. **24**(5): p. 674-81.

137. Kapp, E.A., et al., *Mining a tandem mass spectrometry database to determine the trends and global factors influencing peptide fragmentation.* Anal Chem, 2003. **75**(22): p. 6251-64.

138. Kratzer, R., et al., *Suppression effects in enzymatic peptide ladder sequencing using ultraviolet - matrix assisted laser desorption/ionization - mass spectormetry.* Electrophoresis, 1998. **19**(11): p. 1910-9.

139. Koeniger, S.L., et al., *A quantitation method for mass spectrometry imaging.* Rapid Commun Mass Spectrom, 2011. **25**(4): p. 503-10.

140. Daltonics, B., *Application Note #MT-75 Improved SNAP Algorithm: Application for Quantitation of $^{18}$O-Labeled Peptides*. 2005, Bruker Daltonics: Bremen.

141. Yang, C., Z. He, and W. Yu, *Comparison of public peak detection algorithms for MALDI mass spectrometry data analysis.* BMC Bioinformatics, 2009. **10**: p. 4.

142. Valkenborg, D., I. Jansen, and T. Burzykowski, *A model-based method for the prediction of the isotopic distribution of peptides.* J Am Soc Mass Spectrom, 2008. **19**(5): p. 703-12.

143. Morita, Y., et al., *Imaging mass spectrometry of gastric carcinoma in formalin-fixed paraffin-embedded tissue microarray.* Cancer Sci, 2010. **101**(1): p. 267-73.

144. Jones, E.A., et al., *Multiple statistical analysis techniques corroborate intratumor heterogeneity in imaging mass spectrometry datasets of myxofibrosarcoma.* PLoS One, 2011. **6**(9): p. e24913.
145. Liu, Q., et al., *Comparison of feature selection and classification for MALDI-MS data.* BMC Genomics, 2009. **10 Suppl 1**: p. S3.
146. Le Faouder, J., et al., *Imaging mass spectrometry provides fingerprints for distinguishing hepatocellular carcinoma from cirrhosis.* J Proteome Res, 2011. **10**(8): p. 3755-65.