Novel and Emerging Analytical Techniques for the Identification and Quantification of Proteins in Complex Biological Systems

Michael Mariani

This thesis is presented for the degree of
Doctor of Philosophy

Macquarie University
Department of Chemistry and Biomolecular Sciences

Submitted on the 17th day of February 2012

TABLE OF CONTENTS

SUMMARY

The aim of this thesis was to investigate and further develop a series of novel and emerging techniques in the field of proteomics, used for the identification and quantification of proteins in a range of complex biological systems. This involved pursuing three separate projects all linked by this common theme, as outlined below.

The first project focussed on testing the viability of a novel chemistry of self-assembled monolayers orientated in concentric circles on MALDI plates (known as a Biochip) and determining if more information for identification of proteins could be obtained utilising such methodologies. I was able to show that the biochips could concentrate simple peptide samples and afford a practitioner 10-100 fold increases in limit of detection in the attomole/μl range compared to standard MALDI methods. I also developed the first hybrid AnchorChip/4800 plate system so that the AnchorChip^{TM} technology could be used in an Applied Biosystems 4800 TOF/TOF mass spectrometer. The biochip did perform similarly to the novel hybrid AnchorChip^{TM} on single protein digests. The ability of the biochip to remove salt contaminants was shown on spiked peptide samples, though the technique was problematic at best on gel plug digests and needs further investigation before it can be considered a viable and robust method. The ability of the biochip to selectively affinity capture and isolate phosphorylated peptides from a protein digest was shown at the femtomole level. However, the biochip lost the ability to concentrate the sample once the new functional chemistry for affinity capture was applied. These results are an interesting proof of concept, but the method still needs further development before it can be considered a working platform that can achieve both affinity capture and concentration of a biological sample mixture.

The second project was developed to show the potential pitfalls of current bottom-up proteomic methods, namely the misidentification of some proteins in a sample set. The justification for this comes from the protein inference problem and I was aiming to create an argument for the development of better top-down proteomic methods or enhanced bottom-up methods. I developed a novel multidimensional protein fractionation system called PROOF, with a novel graphical interpretation and representation of the peptide data related back to the elution of the proteins from the

PROOF system. This highlighted the proof of concept for the application of PROOF to a complex and important proteome such as human plasma, and brought to light truncated or cleaved elements within this proteome that standard bottom-up proteomic analysis could not identify. Specifically, I identified five protein candidates for which I demonstrated new features. This can serve as a basis for future analysis of their endogenous primary structure, as well as possible tertiary and quaternary structural elements.

The third project involved quantitative proteomics as applied in plant systems. The aim was to develop a sample preparation method that worked in plants for iTRAQ labelling, and compare this with label-free spectral counting methodology in use in our group at the time. The biological aim of this project was to elucidate new information pertaining to the biochemistry of rice under cold stress conditions. I was able to get the iTRAQ labelling to work in a plant system, particularly rice leaf material that had undergone temperature stress. I was also able to show that both quantitative techniques are comparative and identified similar biological insights, while the total number of proteins identified and quantified by spectral counting was proportionately larger, with 236 and 84 proteins for spectral counting and iTRAQ respectively. We were also able to identify two uniquely effected biological pathways for cold stress by spectral counting that iTRAQ did not show; histone production and vitamin B biosynthetic proteins. These results showed that in our hands, spectral counting was more viable than iTRAQ for quantitative proteomic analysis in plant systems.

The body of work presented in this thesis represents a significant contribution to the field of proteomics. I have developed new approaches, validated existing methods, and used some of these to discover new biological insights – which are the ultimate goal of any proteomics experiment.

DECLARATION

I proclaim that the work presented in this thesis entitled “Novel and Emerging Analytical Techniques for the Identification and Quantification of Proteins in Complex Biological Systems” has not been submitted, either in whole or part, for any higher degree to any other university or institution other than Macquarie University. I also affirm that this thesis is an original piece of research and has been written by myself. Any help or assistance received in my research work and the preparation of the thesis has been appropriately acknowledged.

Michael Mariani

Date 17/02/2012

ACKNOWLEDGEMENTS

I wish to acknowledge Macquarie University, the Australian National University, the Australian Proteome Analysis Facility and the Department of Chemistry and Biomolecular Sciences at Macquarie University for the opportunity to conduct my studies.

I acknowledge the funding and support of Lumicyte and Qiagen, particularly Dr. Christopher M. Belisle, Dr. John A. Walker II, Dr. Mark J. Levy and Dr. Douglas P. Greiner for the biochip work.

I acknowledge Applied Biosystems (AB SCIEX) for access to their instruments and support staff both in Australia and the United States of America.

I want to extend my deep gratitude and sincere thanks to Prof. Paul A. Haynes for his continued and unwavering support, for being a truly great supervisor (advisor), scientist and person at large. I feel privileged to have known you.

I would like to recognise Karlie Nielson as my ‘partner in crime’ for the rice study and being both an impeccable scientist and person, with intelligence and caring that will never be forgotten.

I acknowledge the collaboration and friendship of my ‘brother from another mother’ Dr. (Dr.) Amit Kapur, for the PROOF work and all the rest that comes with being a life friend.

I wish to acknowledge the collaboration and friendship of Dr Brian Hood for the PROOF work and being of unwavering and impeccable character.

I wish to recognise Dana Pascovici, Tim Keighley and Jürgen Busam for bioinformatic and statistical assistance in the PROOF and Rice work, and for being available for abstract conversations.

I would like to thank Brett Cooke for his technical assistance, friendship and wisdom.

I would like to thank Michael Otero for proof reading this document and being an unexpected jewel of friend, with intelligence, insight, wisdom and moral fortitude that has been and I hope will be a continued positive influence on my life.

Lastly, I would like to thank my family and friends of whom there are too many to name, they know who they are, for their support and influence throughout this time – Thank you!

PUBLICATIONS ARISING FROM THIS THESIS

Peer reviewed journal articles (see appendices):

1. Neilson, K. A., Ali, N. A., Muralidharan, S., Mirzaei, M., Mariani, M., Assadourian, G., Lee, A., van Sluyter, S. C., Haynes, P. A., Less label, more free: Approaches in label-free quantitative mass spectrometry. *Proteomics* **2011**, 11, (4), 535-553.

2. Neilson, K. A., Mariani, M., Haynes, P. A., Quantitative proteomic analysis of cold-responsive proteins in rice. *Proteomics* **2011**, 11, (9), 1696-706.

CONFERENCE PRESENTATIONS

Oral presentations:

1. Mariani, M., Ali, N., Ashman, K., Baker, M. S. Kapur, A., Lee, A., An Orthogonal Approach for the Discovery of Biomarkers in Plasma. 12th Lorne Proteomics Symposium, Lorne, Australia, **February 2007**

Poster presentations:

1. Mariani, M., Baker, M. S., Evaluation and Development of Methodologies for a New High-Throughput Affinity MALDI Mass Spectrometry Biochip Platform. 10th Lorne Proteomics Symposium, Lorne, Australia, **February 2005**

2. Mariani, M., Adler, L., Aristoteli, L., Baker, M. S., Belisle, C. M., Chen, I. Y., Connolly, A, M., Kapur, A., Kohler, J. D., Levy, M. J., Molloy., Walker II, J. A., Applications of a Multi-Zone SAM-Biochip for the Affinity Capture, Purification and Concentration of Phosphorylated Peptides and Engineered Proteins with Detection by MALDI-MS. 53rd American Society for Mass Spectrometry Conference (ASMS), San Antonio, Texas, USA, **June 2005**

3. Mariani, M., Belisle, C. M., Kapur, A., Walker II, J. A., Baker, M. S., Concentric Self-Assembled Monolayer MALDI Targets Result in Affinity Peptide Capture and Enhanced Sensitivity. 11th Lorne Proteomics Symposium, Lorne, Australia, **February 2006**

4. Mariani, M., Ali, N., Ashman, K., Baker, M. S. Kapur, A., Lee, A., An Orthogonal Approach for the Discovery of Biomarkers in Plasma. 3rd Asia Oceania Human Proteome Organization Conference (AOHUPO), Singapore **December 2006**

5. Mariani, M., Ashman, K., Belisle, C. M., Kapur, A., Milburn, P. J., Walker II, J. A., Baker. M. S., To Concentrate or Not To Concentrate on a MALDI Plate. 12th Lorne Proteomics Symposium, Lorne, Australia, **February 2007**

6. Mariani, M., Ashman, K., Baker. M. S., Clark, C., Jay, N., Molloy, M., Sherman, J., Song, X., Quantification of Proteins in Plasma Using iTRAQ a Comparison of LC-MALDI and ESI. 12th Lorne Proteomics Symposium, Lorne, Australia, **February 2007**

7. Mariani, M., Ali, N., Ashman, K., Baker, M. S. Kapur, A., Lee, A., An Orthogonal Approach for the Discovery of Biomarkers in Plasma. 12th Lorne Proteomics Symposium, Lorne, Australia, **February 2007**

8. Mariani, M., Neilson, K. A., Haynes, P. A., Label Free and Labelled Quantitative Proteomic Analysis of Cold Stress in Rice. 3rd European Proteomics Association Conference (EuPA), Stockholm, Sweden, **June 2009**

9. Mariani, M., Neilson, K. A., Haynes, P. A., Label Free and Labelled Quantitative Proteomic Analysis of Cold Stress in Rice. 18th International Mass Spectrometry Conference (IMSC), Bremen, Germany, **August 2009**

10. Neilson, K. A., Mariani, M., Keighley, T., Pascovici, D., Haynes, P. A. Label-free Quantitative Proteomic Analysis of Cold-responsive Proteins in Rice. 9th Human Proteome Organisation World Congress (HUPO), Sydney Australia, **September 2010**

ABBREVIATIONS

1D	One Dimensional
1D-GE	One Dimensional Gel Electrophoresis
2D	Two Dimensional
2D-GE	Two Dimensional Gel Electrophoresis
3D	Three Dimensional
Ab-NTA	N-(5-Amino-1-carboxypentyl)iminodiacetic acid
ABI	Applied Biosystems Incorporated
ACTH	Adrenocorticotropic hormone
ACN	Acetonitrile
AC/RP-MS/MS	Affinity Chromatography coupled to Reverse Phase Chromatography Tandem Mass Spectrometry
AFP	Alpha-fetoprotein
AHC	Ammonium Hydrogen Citrate
APAF	Australian Proteomics Analysis Facility
AUC	Area Under the Curve / Peak Area
BIRD	blackbody infrared radiative dissociation
C^{18}	Reverse Phase
CD	Circular Dichroism
CEA	Carcinoembryonic Antigen
CID	Collision Induced Dissociation (collision activated dissociation)
CHCA	α-cyano-4-hydroxy-cinniamic acid
CRP	C-reactive Protein
CSF	Cerebrospinal Fluid
DC	Direct Current
DCM	Dichloromethane
DDA	Data Dependant Acquisition
DHB	2,5-dihydroxy benzoic acid
DNA	Deoxyribonucleic Acid
ECD	Electron Capture Dissociation
EDC	Ethyl-3-(3-dimethylaminopropyl)
ELISA	Enzyme-linked Immunosorbent Assay
ESI	Electrospray Ionisation
ETD	Electron Transfer Dissociation
EtOH	Ethanol
F	Linker Region for iTRAQ
FASP	Filter-Aided Sample Preparation
FDA	Federal Drug Administration of the United States of America
FPLC	Fast Protein Liquid Chromatography
FTICR	Fourier Transform Ion Cyclotron Resonance
GELFrEE	Gel-eluted Liquid Fraction Entrapment Electrophoresis
HeLa cells	Human Cell Line derived from Cervical Cancer
HIC	Hydrophobic interaction chromatography
HPLC	High-Pressure Liquid Chromatography
HAS	Human Serum Albumin
HUPO	Human Proteome Organisation
ICAT	Isotope-coded Affinity Tags
ICR	Ion Cyclotron Resonance
IEF	Isoelectric Focusing
IEX	Ion Exchange
IEX/RP-MS/MS	Ion Exchange Chromatography coupled to Reverse Phase MS/MS
IgG	Immunoglobulin G
IgY	Immunoglobulin Y
IM	Ion Mobility
IMAC	Immobilised Metal Affinity Chromatography
IPG-IEF	Immobilised pH Gradient Isoelectric Focusing
IRMPD	Infrared multiphoton dissociation
iTRAQ	isobaric Tags for Relative and Absolute Quantification
IUPAC	International Union of Pure and Applied Chemistry
IVD	*in vitro* Diagnostic

LC	Liquid Chromatography
LC-MALDI	Liquid Chromatography MALDI
$LCMS^E$	MS of both precursor and product ions in a single analytical run
LC-MS/MS	Liquid Chromatography linked to MS/MS
LOD	Limit of Detection
LOQ	Limit of Quantification
LMWM	Low Molecular Weight Marker
m/z	Mass to Charge ratio (mass divided by charge ratio)
M	Reported Region for iTRAQ
M/A	Matrix to Analyte ratio
MALDI	Matrix Assisted Laser Desorption Ionisation
MALDI TOF	MALDI linked to a TOF analyser
MALDI TOF/TOF	MALDI linked to a tandem TOF analyser
MASCOT	Proteomics search engine of peptide MS data by Matrix Sciences
MEMS	Micro-electro-mechanical systems
METS	Macquarie University Engineering and Technical Services
MQ	Macquarie University
M_r	Molecular Weight
MRI	Magnetic Resonance Imaging
mRNA	messenger Ribonucleic Acid
MS^3	Mass Spectrometry three times
MS^n	Mass Spectrometry to the n^{th} degree of times
MS/MS	Tandem Mass Spectrometry – MS followed by MS (MS^2)
MTP	Micro-titre Plate
MudPIT	Multidimensional Protein Identification Technology
N	Normalisation Region for iTRAQ
nano-ESI	Nano-litre ESI
nano-RP-LC-MS/MS	Nano-litre flow rate Reversed Phase Liquid Chromatography MS/MS
Nd:YAG	Neodymium-doped yttrium aluminium garnet; $Nd:Y_3Al_5O_{12}$
NMR	Nuclear Magnetic Resonance
NSAF	Normalised Spectral Abundance Factors
NTA3 (Biochip)	Nitrilotriacetic Acid Biochip
NTA	Nitrilotriacetic Acid
OMSSA	Open Mass Spectrometry Search Algorithm
OsSALT	encoding a 15-kDa mannose-binding lectin protein
OsNac6	encoding an apical meristem transcription factor
OVA1	Probability blood test of five proteins for Ovarian Cancer Mass
PCR	Polymerase Chain Reaction
pI	Isoelectric Point
PLRP	Polymeric Reverse Phase
PMF	Peptide Mass Fingerprint
pre-mRNA	precursor-messenger Ribonucleic Acid
PROOF	protein repetitive orthogonal off-line fractionation
pS	Phosphorylated Serine
PSD	Post Source Decay
pY	Phosphorylated Tyrosine
PTM	Post Translational Modification
QIT	Quadrupole Ion Trap
QQQ	Triple Quadrupole
Q-Sepharose	Quaternary ammonium anion-exchanger
R	Reactive Group for iTRAQ
RF	Radio Frequency
RP3 (Biochip)	Reverse Phase Biochip
RP	Reverse Phase / C^{18}
RP-LC-MS/MS	Reverse Phase Liquid Chromatography MS/MS
RP/RP-MS/MS	Two Reverse Phase Columns ‘in-line’ to MS/MS
RNA	Ribonucleic Acid
RT	Room Temperature
SAM	Self Assembled Monolayer
SAX	Strong Anionic Exchange
SC	Spectral Counting
SC-MALDI	Surface Chemistry MALDI
SCSC-MALDI	Spherically Concentric Surface Chemistry MALDI

SCX	Strong Cation Exchange
SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
SELDI	Surface Enhanced Laser Desorption Ionisation
SEQUEST	Proteomics search algorithm for tandem MS data
SILAC	Stable Isotope Labels with Amino Acids
sIEF	Solution Phase Isoelectric Focusing
SLA	Soft Laser Desorption
SpC	Spectral Counts
SP-Sepharose	Chromatography media Sulphopropyl cation-exchanger
SRM	Selected Reaction Monitoring
Std-MALDI	Standard MALDI
TFA	Trifluoroacetic Acid
TOF	Time of Flight
UPLC	Ultra-performance Liquid Chromatography
VEMS	Program for analysis of MS quantitative proteomics data
X3 (Biochip)	Mass Spec Focus Chip / Concentration Biochip
X!Tandem	Proteomics search algorithm for tandem MS data