

University of Warwick institutional repository: <http://go.warwick.ac.uk/wrap>

A Thesis Submitted for the Degree of PhD at the University of Warwick

<http://go.warwick.ac.uk/wrap/2787>

This thesis is made available online and is protected by original copyright.

Please scroll down to view the document itself.

Please refer to the repository record for this item for information to help you to cite it. Our policy information is available from the repository home page.

Novel Mass Spectrometry-Based Approaches
for the Characterisation of Systems of
Biological Interest

Vibhuti Jitendra Patel B.Sc. (Hons)

A thesis submitted for the degree of Doctor of Philosophy

University of Warwick
Department of Biological Sciences

September 2009

॥ ॐ ॥

या कुंदेदु तुषारहार धवला, या शुभ वस्त्रावृता ।
या वी ावर दण्डमंडितकरा, या श्वेतपद्मासना ॥
या ब्रह्माच्युतशंकरप्रभृतिभिर्देवैः सदा वन्दिता ।
सा मां पातु सरस्वती भगवती निःशेष जाइयापहा ॥

*For my parents, Jitendra and Aruna, without whose encouragement and guidance
none of this would have been possible.*

“If we knew what it was we were doing, it would not be called research, would it?”

- Albert Einstein

Contents

List of figures	vi
List of tables	ix
Acknowledgements	x
Declaration	xii
Summary	xiii
List of abbreviations	xiv
Chapter 1 - Introduction to mass spectrometry	1
1.1 What is mass spectrometry?	2
1.2 Ionisation	3
1.2.1 MALDI	3
1.2.2 ESI	4
1.3 Mass analysers	6
1.3.1 The quadrupole	6
1.3.2 Time-of-flight	6
1.3.3 Ion traps	7
1.3.4 Fourier transform ion cyclotron resonance	8
1.4 Detectors	9
1.5 Tandem mass spectrometry	10
1.6 Current approaches in biological mass spectrometry	12
Chapter 2 - Introduction to ambient ionisation mass spectrometry	20
2.1 Emergence of ambient ionisation techniques	21
2.2 Liquid and gas jet desorption/ionisation	27
2.3 Thermal desorption/ionisation	28
2.4 Laser desorption/ionisation	29
2.5 Applications of ambient ionisation techniques	31
2.6 Aims and objectives	33

Chapter 3 - Materials and Methods - Ambient Ionisation Mass Spectrometry	37
3.1 Overview of ambient ionisation study	38
3.2 Direct analysis in real time	38
3.3 Desorption electrospray ionisation	39
3.4 Desorption atmospheric pressure chemical ionisation	39
3.5 Accurate mass measurement protocol for the Q-TOF I	40
3.6 Collision-induced dissociation pathways	40
3.7 Polarity switching accurate mass measurement	41
Chapter 4 – Results - Ambient Ionisation Techniques	46
4.1 Rapid screening of pharmaceuticals and samples of biological origin ...	47
4.1.1 Analysis of solid tablets	47
4.1.2 Detection of active ingredients in a topical ointment	48
4.1.3 Desorption of a gel formulation from human skin	52
4.1.4 Comparison of positive and negative DAPCI-MS/MS	54
4.1.5 Analysis of a plant alkaloid	57
4.1.6 Identification of ibuprofen metabolites from human urine	59
4.2 Determining collision-induced dissociation pathways	61
4.2.1 Product ion and neutral loss scanning in negative ion DESI ...	61
4.2.2 Product ion and precursor ion scanning of pharmaceuticals.....	63
4.2.3 Product ion and precursor ion scanning of nicotine	66
4.3 Polarity switching accurate mass measurement of pharmaceuticals	69
4.3.1 Preferential ionisation	69
4.3.2 Analysis of multiple active ingredients from a single tablet ...	71
4.3.3 Analysis of a pharmaceutical ointment	71
4.4 Conclusions	73
Chapter 5 - Introduction to metallomics	75
5.1 Metals and proteins	76
5.2 Multi-technique approaches to studying metalloproteins	77
5.3 ICP-MS for metal identification and quantification.....	78
5.3.1 Principles of ICP-MS	78
5.3.2 Overcoming interferences	81

5.3.3 Sample introduction	81
5.4 Organic mass spectrometry methods for the study of metalloproteins ...	85
5.5 Ion mobility	86
5.6 Particulate methane monooxygenase	88
5.7 Hemoglobin and hemoglobin disorders	92
5.8 Aims and objectives	95
Chapter 6 - Materials and Methods - Metallomic Studies	103
6.1 Purification of pMMO.....	104
6.2 EDTA treatment	104
6.3 Gel electrophoresis.....	104
6.4 Inorganic analysis.....	105
6.4.1 Laser ablation	105
6.4.2 Digestion of gel bands.....	106
6.4.3 ICP-MS	106
6.5 Hemoglobin sample preparation	106
6.6 IMMS analysis	107
6.6.1 Data acquisition.....	107
6.6.2 Calibration, modelling and cross-section estimation	108
Chapter 7 – Results – Metallomic Studies	112
7.1 Analysis of gel-resolved proteins by LA-ICP-MS.....	113
7.1.1 Technical aspects of the laser ablation system.....	113
7.1.2 Analysis of dried gels.....	113
7.1.3 Analysis of wet gels	114
7.2 Analysis of gel-resolved proteins by digestion and ICP-MS	118
7.3 In-solution analysis of pMMO by ICP-MS.....	125
7.4 Ion mobility studies of human hemoglobin	130
7.4.1 Instrument acquisition parameters and calibration	130
7.4.2 Hemoglobin tetramer analysis.....	131
7.4.3 Cross-section determination.....	133
7.5 Conclusions	138
7.5.1 Inorganic analysis to determine metal content of pMMO	138

7.5.2 Probing hemoglobin structure	138
Chapter 8 - Introduction to mass spectrometry-based proteomics	143
8.1 Mass spectrometry-based proteomics	144
8.1.1 Top-down proteomics	144
8.1.2 Bottom-up proteomics.....	146
8.2 Quantitative proteomics	152
8.2.1 Gel-based quantification	152
8.2.2 Peptide labelling.....	153
8.2.3 Non-labelling quantification	153
8.3 Computational proteomics	156
8.3.1 Protein identification.....	157
8.3.2 Protein quantification.....	159
8.4 Quality of data.....	161
8.5 Mapping proteomic data to biological pathways	163
8.6 <i>Methylocella silvestris</i>	163
8.7 Aims and objectives	165
Chapter 9 - Materials and Methods - Mass Spectrometry-Based Proteomics .	172
9.1 Sample preparation.....	173
9.1.1 Bacterial growth and sample preparation.....	173
9.1.2 Protein separation by gel electrophoresis.....	173
9.1.3 iTRAQ and strong cation exchange chromatography	174
9.1.4 In-solution tryptic digestion	174
9.2 Mass spectrometry	175
9.2.1 LC-MS/MS acquisition for gel-separated samples	175
9.2.2 LC-MS/MS acquisition for iTRAQ samples.....	174
9.2.3 LC-MS configurations for label-free analysis.....	176
9.3 Data processing and database searching	177
9.3.1 Data processing for DDA acquisitions.....	177
9.3.2 Data processing for label-free acquisitions.....	177
9.3.3 Database searches.....	177
9.4 Protein quantification.....	178

9.4.1 Protein quantification using iTRAQ labelling	178
9.4.2 Protein quantification using label-free system.....	178
9.5 Mapping protein identifications onto biological pathways.....	180
Chapter 10 – Results - Mass Spectrometry-Based Proteomics.....	182
10.1 Protein identifications	183
10.2 Relative quantification of identified proteins.....	185
10.2.1 Gel-based approach.....	185
10.2.2 iTRAQ labelling.....	187
10.2.3 Label-free MS ^E quantification	187
10.2.4 Agreement of quantification approaches	189
10.3 Comparison of experimental approaches	196
10.4 Biological significance of results obtained	198
10.4.1 Placing results in the context of methane oxidation	198
10.4.2 Mapping protein identifications onto biological pathways...	199
10.5 Conclusions	202
Concluding Remarks	205
Appendix A - Supplementary Proteomic Data.....	207
Appendix B - Dissemination of work.....	252

List of figures

1.1	Representation of a modern mass spectrometer	2
1.2	The MALDI ionisation process	5
1.3	The ESI ionisation process	5
1.4	Representation of a TOF instrument equipped with a reflectron	8
1.5	Peptide ion nomenclature	11
2.1	Schematic of a typical DESI experiment	21
2.2	(a) Sample introduction using a DART source	22
	(b) Proposed ionisation mechanism in DART	22
4.1	Anadin Extra tablet by DESI, DAPCI and DART	48
4.2	Positive ion DAPCI MS/MS from protonated caffeine	49
4.3	Solpadeine Max tablet by DESI, DAPCI, DART	49
4.4	DAPCI analyses of proctosedyl ointment	50
4.5	DAPCI MS/MS of cinchocaine in negative and positive ion mode	50
4.6	Proposed fragmentation pathways for cinchocaine	51
4.7	A comparison of DESI, DAPCI without solvent and DAPCI with solvent to analyse proctosedyl ointment	52
4.8	A comparison of DART with DAPCI to analyse proctosedyl ointment	52
4.9	Negative ion DESI of ibuprofen gel desorbed from human skin	53
4.10	Analysis of a metoclopramide tablet by negative ion DAPCI, positive ion DAPCI and positive ion DART	54
4.11	Fragmentation of metoclopramide by negative and positive ion DAPCI	55
4.12	Proposed fragmentation pathways for metoclopramide	56
4.13	Analysis of tobacco from a cigarette by DESI, DAPCI, DART	57
4.14	Proposed fragmentation pathway for nicotine	57
4.15	Negative ion DAPCI analysis of a urine sample, following administration of ibuprofen	59
4.16	(a) negative ion DESI-MS of ibuprofen, naproxen and diclofenac	
	(b) DESI MS/MS spectrum of deprotonated species	
	(c) DESI-MS/MS neutral loss spectrum	61
4.17	DESI-MS of Anadin Extra	63
4.18	Proposed fragmentation pathways of aspirin	64
4.19	Positive ion DESI-MS/MS of caffeine	64
4.20	Proposed fragmentation pathways of caffeine	65
4.21	Positive ion DESI-MS/MS spectrum of protonated nicotine	66
4.22	Precursor ion scans of nicotine fragments	66
4.23	Proposed fragmentation pathways of nicotine	67
4.24	DESI-MS of leucine enkephalin and erythromycin	69

4.25	DESI analysis of erythromycin	70
4.26	Polarity switching acquisition of Anadin Extra tablet	71
4.27	Isotope distribution in chloramphenicol	71
5.1	Involvement of metals in biological processes	76
5.2	Current hyphenated techniques for the study of metalloproteomes	78
5.3	An overview of the processes involved in ICP-MS	79
5.4	Schematic of an ICP-MS instrument	79
5.5	The DRC cell in an ICP MS instrument	82
5.6	Pneumatic nebulisers for liquid sample introduction in ICP-MS	83
5.7	A laser ablation inductively coupled plasma source	83
5.8	A schematic of the Synapt HDMS instrument	87
5.9	The methane oxidation pathway of methanotrophic bacteria	88
5.10	Crystal structures of pMMO	90
5.11	Three-dimensional structure of pMMO as determined by EM	91
5.12	The tetrameric structure of human hemoglobin	92
5.13	Approaches to the study of metalloproteins	94
7.1	Gel visualisation in the laser ablation chamber	113
7.2	Ablation of dried polyacrylamide gels	114
7.3	LA-ICP-MS analysis of gels spiked with iron and zinc	115
7.4	Analysis of protein bands in wet gels	116
7.5	SDS-PAGE of purified pMMO	118
7.6	Analysis of digested gel bands from SDS-PAGE of purified pMMO	119
7.7	Analysis of digested gel bands from SDS-PAGE of pMMO treated with EDTA	120
7.8	Repeat analysis of digested gel bands from SDS-PAGE of purified pMMO	121-122
7.9	Blue Native gel of purified pMMO	123
7.10	Analysis of digested bands from BN-PAGE of purified pMMO	123
7.11	In-solution ICP-MS analysis of pMMO	125
7.12	In-solution ICP-MS analysis of untreated pMMO	127
7.13	ICP-OES analysis of untreated pMMO	128
7.14	Calibration curve for estimation of cross sections	129
7.15	Mass spectra of normal and sickle hemoglobin	131
7.16	Average estimated cross-sections for hemoglobin monomers	133
7.17	Average estimated cross-sections for hemoglobin monomers, dimer and tetramer	134
7.18	Estimated cross-sections for HbA and HbS tetramers	134
7.19	Evolution of the structure of a globular protein following ESI	136
8.1	Typical approaches in top-down and bottom-up proteomics	148
8.2	An overview of the iTRAQ labelling quantification system	153
8.3	Typical workflow for a profiling and quantitative proteomics experiment	155
8.4	Whole cell hybridisation in a culture of <i>Methylocella silvestris</i>	163
9.1	Overview of the data processing and database searching workflow from an MS ^E acquisition	178

10.1	Number of proteins identified by each experimental approach	184
10.2	1D SDS-PAGE separation of the <i>M. silvestris</i> proteome	185
10.3	Differential expression as determined by iTRAQ labelling	187
10.4	Automated protein-level quantification of regulated proteins using a label-free system	188
10.5	Correlation of quantification data from the iTRAQ and label-free methods	190
10.6	Peptide sequences for precursors and corresponding isobaric reporter ions for outlying protein identification citrate synthase	191-194
10.7	Pathway of methane oxidation in methanotrophic bacteria	197
10.8	iTRAQ and label-free quantification data for key metabolic enzymes	198
10.9	Extracted pathway analysis of the citric acid cycle from the KEGG database	200

List of tables

1.1	Instrumental and experimental developments in MS	14
2.1	Summary of currently available ambient ionisation methods	23
3.1	Pharmaceutical formulations analysed using ambient ionisation techniques	42
3.2	Active ingredients of the pharmaceutical formulations investigated	43
4.1	Fragments from DAPCI MS/MS of codeine	47
5.1	Proposed compositions of the metal sites in pMMO	89
7.1	Natural abundances of copper and iron isotopes	128
8.1	Commonly used mass spectrometers for proteomics	149
10.1	Total protein identifications from proteomics experiments	183
10.2	A comparison of the experimental requirements for each of the approaches employed	196

Acknowledgements

*"I count myself in nothing else so happy
As in a soul rememb'ring my good friends."*

- William Shakespeare, King Richard II, Act II, Scene 3

Although I am the sole author of this thesis, I am by no means the sole contributor. Many people have been there for me, and it is their support and belief which have enabled me to complete my research.

I would like to thank my supervisor Jim Scrivens for the opportunity to undertake this Ph.D. His counsel, backing, unfaltering encouragement and constant supply of coffee have been invaluable over the past four years. Thank you to Dr. Jeff Franks for his supervision of my CASE Award project in the Measurement Science Group at ICI Wilton (now Intertek MSG), for making time to assist my experimental work and for his continual advice.

Thanks to all members of the Scrivens Group, past and present, for helpful discussions and for providing an enjoyable working environment: Charlie, Elle, Fran, Georgios, Gill, Jonathon, Matt, Narinder, Nisha, Richard, Sarah, and last but by no means least, Sue. Thank you to the Measurement Science Group at ICI Wilton for accommodating me, and to Tony and Julia for welcoming me into their home on so many occasions as I carried out work for my CASE project.

A special thank you goes to Kostas for being a wonderful post-doc, a trusted advisor and, above all, a true friend. I will always be grateful for the endless affection and generosity that he and Jo have shown me; I sincerely consider them as family.

The work undertaken has been supported by many colleagues and collaborators, who I would also like to thank: Dr. Nataliia Myronova and the late Prof. Sir Howard Dalton for the investigation of pMMO, Prof. Colin Murrell and Andrew Crombie for the *M. silvestris* project, Joanne Connolly, Richard Lock, Brian Green, Kevin Giles, Mark McDowall and all the other people at Waters for their expertise, and Prof. Keith Jennings for his inexhaustible wisdom.

I have been blessed with a greater number of friends than I am able to list, but I am grateful to every one of them for the love, laughter and many cups of tea which have

helped me more than they know. Particular mention must go to Andy, Paul, Helen and Anne for the domestic delights of living at Number 131.

Thank you to my brother, Himesh, who has been my confidante and my unexpected best friend.

Finally, the biggest thank you goes to my parents, whose patience, understanding and friendship have kept me afloat. Everything I have achieved, I owe to their unconditional love.

Declaration

I hereby declare that this thesis, submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy and entitled *Novel Mass Spectrometry-Based Approaches for the Characterisation of Systems of Biological Interest* represents my own work and has not previously been submitted to this, or any other, institution for any degree, diploma or other qualification. Work undertaken by my collaborators is explicitly stated where appropriate.

Vibhuti J. Patel

September 2009

Summary

Originally established as an analytical technique in the fields of physics and chemistry, mass spectrometry has now also become an essential tool in biology. Advances in ionisation methods and novel types of instrumentation have led to the development of mass spectrometry for the analysis of a wide variety of biological samples. The work presented here describes the use mass spectrometry for the study of a number of biological systems.

A new family of techniques has been developed allowing ions to be created under ambient conditions. Three of these ambient ionisation techniques, coupled to different mass analysers, were employed for the rapid screening of pharmaceutical formulations. Active ingredients were identified and subjected to collisionally-induced dissociation, enabling the elucidation of potential fragmentation pathways. Drug metabolites were also successfully identified from biological samples.

Inorganic mass spectrometry was employed to probe the metal centres of the enzyme, particulate methane monooxygenase, a methane-oxidising complex found in certain bacteria. This protein has been extensively studied, but questions remain regarding its catalytic mechanism, particularly the involvement of indigenous metal ions. Inductively-coupled plasma mass spectrometry experiments have indicated the presence of copper and iron within the enzyme.

Protein cross-sections, obtained using ion mobility mass spectrometry, can be used to probe the conformation of molecules in the gas phase. A commercial instrument was used to investigate human hemoglobin from clinical samples. A complex assembly mechanism was deduced, resolving previous disputes in the literature, and conformational differences were observed between healthy and sickle molecules.

The field of proteomics is rapidly evolving; as described, techniques are constantly being developed and improved to deal with the enormous complexity that proteomes present. Three proteomics approaches were used to study a recently identified bacterium under two growth conditions. Differences in protein expression were observed and correlated to relevant biological pathways.

List of abbreviations

1D	one-dimensional
2D	two-dimensional
A	
ACN	acetonitrile
ADC	analog-to-digital converter
ADH	alcohol dehydrogenase
AFIMS	alternating-field ion mobility spectrometry
AMO	ammonia monooxygenase
AP	atmospheric pressure
APCI	atmospheric pressure chemical ionisation
APGDDI	atmospheric pressure glow discharge desorption ionisation
APIMS	ambient pressure ion mobility spectrometry
AP-LD/SI	atmospheric pressure laser desorption with secondary ionisation
APTDI	atmospheric pressure thermal desorption ionisation
APTI	atmospheric pressure thermal ionisation
ASAP	atmospheric pressure solids analysis probe
B	
BN	blue native

C

CHCA	α -cyano-4-hydroxycinnamic acid
CI	chemical ionisation
CID	collisionally-induced dissociation
cm	centimetre

D

Da	Dalton
DAPCI	desorption atmospheric pressure chemical ionisation
DAPPI	desorption atmospheric pressure photoionisation
DART	direct analysis in real time
DBDI	dielectric barrier discharge ionisation
DBT	dynamic bandpass tuning
DC	direct current
DDA	data dependent acquisition
DESI	desorption electrospray ionisation
DeSSI	desorption sonic spray ionisation
DiGE	difference gel electrophoresis
DMS	differential mobility spectrometry
DNA	deoxyribonucleic acid
DRC	dynamic reaction cell
DRE	dynamic range extension
DTIMS	drift-time ion mobility spectrometry

E

EASI	easy ambient sonic spray ionisation
EBI	European Bioinformatics Institute

ECD	electron capture dissociation
EDTA	ethylenediamine tetraacetic acid
EESI	extractive electrospray ionisation
EHSS	exact hard sphere scattering
EI	electron impact
ELDI	electrospray-assisted laser desorption ionisation
EM	electron microscopy
emPAI	exponentially modified protein abundance index
ESI	electrospray ionisation
ETD	electron transfer dissociation

F

FAB	fast atom bombardment
FAIMS	field-asymmetric waveform ion mobility spectrometry
FIS	field-ion spectrometry
FT-ICR	Fourier transform ion cyclotron resonance
FWHM	full width at half maximum

G

g	gram
GC	gas chromatography
GFP	[Glu ¹]-fibrinopeptide B
GUI	graphical user interface

H

Hb	hemoglobin
HbA	normal hemoglobin

HbS	sickle hemoglobin
HPLC	high performance liquid chromatography
Hz	Hertz

I

ICAT	isotope-coded affinity tags
ICP	inductively coupled plasma
ICP-AES	inductively coupled plasma atomic emission spectroscopy
ICP-OES	inductively coupled plasma optical emission spectrometry
IMMS	ion mobility mass spectrometry
IMS	ion mobility spectrometry
iTRAQ	isobaric tags for relative and absolute quantification

J

JeDI	jet desorption ionisation
------	---------------------------

K

KAAS	KEGG Automatic Annotation Server
kDa	kilodalton
KEGG	Kyoto Encyclopedia of Genes and Genomes
kV	kilovolts

L

L/hr	litres per hour
LA	laser ablation
LAESI	laser ablation with electrospray ionisation
LC	liquid chromatography

LDTD	laser diode thermal desorption
LIMS	laboratory information management system
LIT	linear ion trap
M	
M	molar
m/s	metres per second
<i>m/z</i>	mass-to-charge
MALDESI	matrix-assisted laser desorption electrospray ionisation
MALDI	matrix-assisted laser desorption/ionisation
mbar	millibar
MCP	microchannel plate detector
mg	milligram
mL	millilitre
mM	millimolar
MRC	Medical Research Council
MS	mass spectrometry
ms	millisecond
MS/MS	tandem mass spectrometry
MS ^E	alternating high/low collision energy tandem mass spectrometry
MudPIT	multi-dimensional protein identification techniques
µg	microgram
µL	microlitres
µm	micrometre
µM	micromolar
µs	microsecond

N

Nd:YAG	neodymium:yttrium
NDEESI	neutral desorption extractive electrospray ionisation
ng	nanogram
nm	nanometre
NMR	nuclear magnetic resonance
NMS	nitrate mineral salts
NSAID	non-steroidal anti-inflammatory drug

O

oa	orthogonal acceleration
OD	optical density

P

PA	projection approximation
PADI	plasma-assisted desorption/ionisation
PAGE	polyacrylamide gel electrophoresis
PAI	protein abundance index
pg	picogram
PIPES	piperazine-1,4-bis(2-ethanesulfonic acid)
PIR	Protein Information Resource
PLGS	ProteinLynx Global Server
pMMO	particulate methane monooxygenase
PMSS	peptide match score summation
PTFE	polytetrafluoroethylene

Q

QIT	quadrupole ion trap
Q-TOF	quadrupole time-of-flight

R

RF	radio frequency
RNA	ribonucleic acid
RPIMS	reduced pressure ion mobility spectrometry

S

SCX	strong cation-exchange
SDS	sodium dodecyl sulphate
SIB	Swiss Institute of Bioinformatics
SILAC	stable isotope labelling in cell culture
SIMS	secondary ion mass spectrometry
sMMO	soluble methane monooxygenase
SpS	spectrum sampling

T

TD	thermal desorption
TDC	time-to-digital converter
TLC	thin layer chromatography
TM	trajectory method
TNT	trinitrotoluene
TOF	time-of-flight
TWIMS	travelling-wave ion mobility spectrometry

U

UV ultraviolet

V

V volts

W

WHO World Health Organisation

Chapter 1

Introduction to mass spectrometry

1.1 What is mass spectrometry?

Mass is a fundamental yet easily understood characteristic of any chemical species, and for many analytical problems, mass is also highly specific. The use of mass in scientific measurements began with the pioneering work of Sir Joseph John Thomson, who was awarded the Nobel Prize in Physics in 1906 for the discovery of the electron (Thomson, 1899). In 1911, Sir Joseph built the first mass spectrometer, then called the parabola spectrograph, and used it to separate isotopes of neon by mass (Thomson, 1911). This technology was subsequently developed by his student, Francis W. Aston, who designed an instrument with improved resolving power that allowed him to study isotopes of many other elements (Aston, 1919). Aston was awarded the Nobel Prize for Chemistry in 1922.

The modern definition of mass spectrometry (MS) is the measurement of mass-to-charge ratios (m/z) of gas-phase ions. The major components of a mass spectrometer are represented in Figure 1.1. An ion source is required to produce gaseous ions from the analyte; the formed ions are resolved by a mass analyser according to their mass and charge. A detector registers the ions at each defined m/z value, and a computer system then stores the acquired data and presents it in an accessible format.

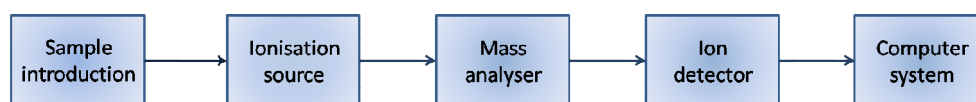


Figure 1.1 – Schematic representation of a modern mass spectrometer

Mass spectrometers are in daily use worldwide for the identification and analysis of a wide range of molecules, including flavours, natural products, pollutants, drugs, and metabolites. MS-based techniques are being broadly used for probing the molecular complexity found in many fields, such as agriculture, atmospheric chemistry, medicine, food, forensics and geochemistry. Many of the improvements in biochemical MS are a direct consequence of the introduction of ‘soft ionisation’ methods that permit the ionisation of large, polar and thermally labile biomolecules.

1.2 Ionisation

Until the 1980s, to analyse a sample by MS, ionisation was usually performed by electron impact (EI) or chemical ionisation (CI) methods (Munson and Field, 2002). These require a vaporised sample, which is of no great concern when studying small organic molecules or those amenable to gas chromatography. Biomolecules of interest, however, are generally polar thermally labile, and as such cannot be analysed by EI- or CI-MS without prior derivatisation (Griffiths *et al.*, 2001). The introduction of fast atom bombardment (FAB) (Barber *et al.*, 1981) allowed, for the first time, the routine MS analysis of polar thermally labile molecules of masses up to a few thousand Daltons (Da). FAB proved most suitable for analysis of samples which exist as pre-formed ions in solution, i.e. protonated/deprotonated or sodiated molecules. Whilst molecules as large as insulin have been studied, this approach works best for species of masses below 1000 Da. Another limitation is the formation of matrix-associated chemical noise, which may mask the presence of sample ions. The most significant step forward in the ionisation of biological samples for MS analysis was the development of matrix-assisted laser desorption/ionisation (MALDI) and electrospray ionisation (ESI) around two decades ago.

1.2.1 MALDI

Koichi Tanaka was awarded the Nobel Prize in Chemistry in 2002 for his work on laser desorption (Tanaka *et al.*, 1988). It is Michael Karas and Franz Hillenkamp, however, whose competitive work became more commercially accepted, and which has led to the place of MALDI as an established technique (Karas and Hillenkamp, 1988).

In MALDI, ions are desorbed from the solid phase. A sample is first dissolved in a suitable solvent and mixed with an excessive amount of an appropriate matrix. For peptide analysis, the most commonly used matrix is α -cyano-4-hydroxycinnamic acid (CHCA); for intact proteins sinapinic acid (3,5-dimethoxy-4-hydroxycinnamic acid) is usually employed. Subsequently, the mixture is spotted on a MALDI plate (commonly made of stainless steel) and air-dried. The sample co-crystallises with the matrix. The components in the sample/matrix mixture are brought into the gas phase via a laser beam (usually a nitrogen laser at a wavelength of 337 nm) that hits

the sample-matrix crystal, leading to absorption of the laser energy by the matrix and subsequent desorption and ionisation of the analytes in the sample. Figure 1.2 represents the process of ion formation in MALDI.

Several hypotheses have been outlined to explain the ionisation process, but details of the mechanism remain unclear. The most commonly accepted theory is that protons are transferred from the matrix to the analyte molecules, either during passage into, or just subsequent to entering, the gas phase (Zenobi and Knochenmuss, 1998). Singly-charged species resulting from the attachment of a single proton are the most commonly formed type of ion in MALDI.

1.2.2 ESI

Alongside Tanaka, John Fenn was also awarded the 2002 Nobel Prize in Chemistry for his work on electrospray ionisation (Fenn *et al.*, 1989). In contrast to the pulsed nature of MALDI, ESI is a continuous ionisation process. The sample is solubilised in aqueous solution with an appropriate concentration of organic solvent, typically acetonitrile (ACN) or methanol. For the more common approach of positive ion ESI, an acid is also added, such as formic or acetic. The particular solvent make-up will depend upon the compound class being studied. The solution is sprayed through a capillary needle, which is held at a high electrical potential with respect to the entrance of the mass spectrometer. The most commonly accepted mechanism for ion formation in ESI is that the charged liquid at the end of the needle forms a cone, known as a Taylor cone, that minimises the charge-to-surface ratio. Droplets are released, the formation of which is facilitated by a nebulising gas (usually nitrogen) flowing around the outside of the needle directing the emerging spray towards the mass spectrometer. Once airborne, the charge repulsion in the droplets overcomes the surface tension (at a point referred to as the Rayleigh limit) and solvent molecules are lost, leading to an increased charge density at the surface of the droplets. Smaller and smaller droplets are formed via this process, and charged, multiply-protonated ions enter the mass spectrometer, as represented in Figure 1.3. It is common for ESI sources to be coupled to high performance liquid chromatography (HPLC) columns, particularly in biochemical analysis.

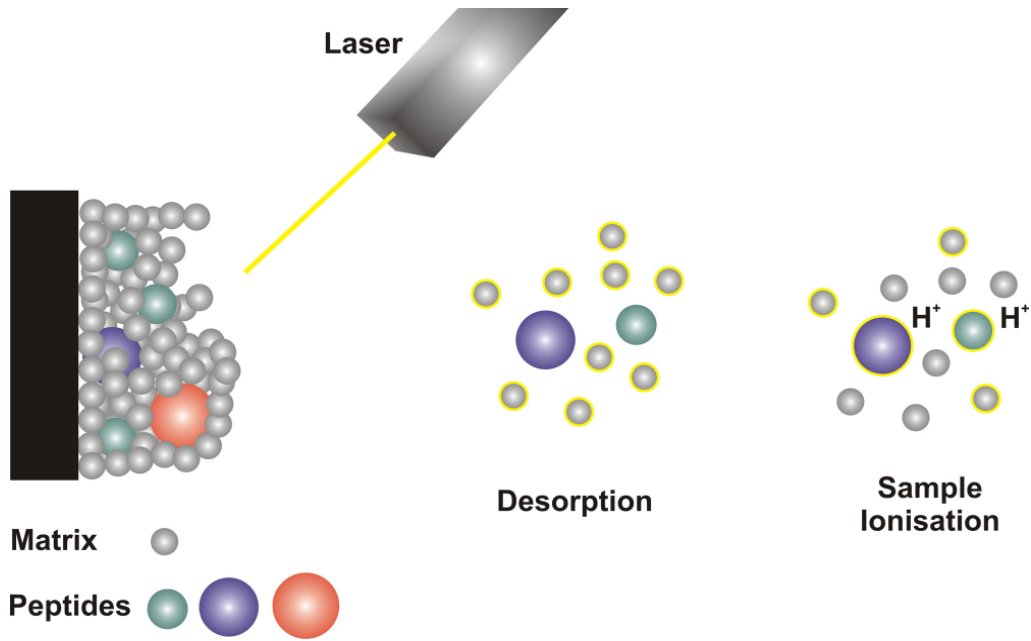


Figure 1.2 – The MALDI ionisation process; energy from the UV laser is absorbed by matrix molecules, which transfer the energy to the co-crystallised peptides.

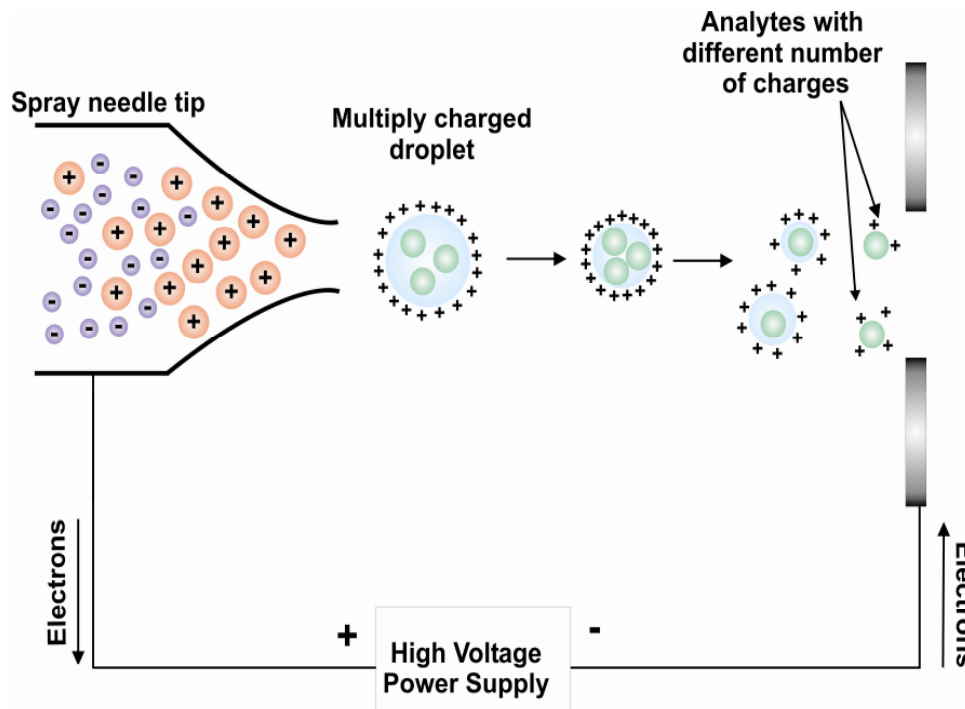


Figure 1.3 – The ESI ionisation process (modified from (de Hoffmann and Stroobant, 2007))

1.3 Mass analysers

Once ions have been produced, they are separated by m/z ratio in the mass analyser, of which there are several types. All mass analysers have a number of main characteristics, amongst which the most important are: upper mass limit, transmission efficiency and resolving power. The upper mass limit defines the highest value of m/z which can be measured, the transmission efficiency is the ratio of number of ions detected to the number of ions produced, and the resolving power represents the ability to separate two adjacent masses with a small mass difference.

1.3.1 The quadrupole

The quadrupole mass analyser consists of four round or hyperbolic rods, arranged as two sets of two electrically connected rods. A combination of radio frequency (RF) and direct current (DC) voltages are applied to each pair, producing a complex oscillating movement of the ions as they move along the mass analyser. The motion of the ions will depend on the electric fields, so that only ions of a particular m/z will have a trajectory stable enough to allow passage through to the detector. The remaining ions will collide with the sides of the rods, or will be lost at the walls of the analyser (Paul and Steinwendel, 1953).

Depending how the voltages are applied, quadrupoles can be used to selectively pass ions of a particular m/z , or can be operated in RF-only mode to allow through a much broader ion packet. This latter form often serves as a collision cell for collision-induced dissociation (CID), which will be discussed in further detail later. Variations on RF-only quadrupoles include multipoles with higher numbers of rods, e.g. hexapoles and octopoles. The multipole mass filter can be found in a large variety of mass spectrometers, often in series with other mass analysers.

1.3.2 Time-of-flight

First described by Wiley and McLaren (Wiley and McLaren, 1955), the time-of-flight (TOF) mass analyser measures the m/z of an ion based on the time it takes for that ion to traverse a field-free region. Ions are pulsed and given a fixed amount of kinetic energy to provide initial acceleration. They subsequently enter the field-free

region, known as the flight tube, where they travel at a velocity inversely proportional to the square root of their mass. The general principle is that the larger the ion, the longer the time needed for its flight.

A disadvantage of linear TOF analysers is the poor resolution. Factors such as the length of the ion formation pulse, the volume of the space where the ions are formed and the variation in initial kinetic energy between ions will all affect resolution. Differences in initial kinetic energy will cause ions of the same m/z to reach the detector at slightly different times, resulting in peak broadening. This can be overcome by using delayed pulse extraction (also known as pulsed ion extraction), which compensates for the variation (Vestal *et al.*, 1995). This reduces the kinetic energy spread among ions with the same m/z by introducing a time delay between ion formation and extraction. The ions are allowed to expand into a field-free region in the source and after a lag (hundreds of nanoseconds to several microseconds) a voltage pulse is applied to extract the ions. As mass resolution is proportional to flight time, increasing the length of the flight tube will achieve higher resolution. Reflectron instruments (Mamyryn *et al.*, 1973) accomplish this by reflecting ions back down the flight tube using a series of ring electrodes which act as an ion mirror, as shown in Figure 1.4. Ions with higher initial kinetic energies travel further into the reflectron compared to those with less initial kinetic energy. The addition of an extra pass (also known as W optics) effectively quadruples the flight path and further improves resolution.

1.3.3 Ion traps

As the name suggests, ion trap mass analysers operate on the principle of trapping ions rather than passing them on. The quadrupole ion trap (QIT) can be imagined as a quadrupole bent in on itself; it uses a combination of RF and DC voltages to select ions of a particular m/z and then trap them in three dimensions (Stafford *et al.*, 1984). A mass spectrum is generated by trapping a broad range of m/z values, then scanning them out of the trap to the detector using a ramping of RF voltages. The linear, or 2D, ion trap (LIT) resembles the quadrupole, but additional DC potentials allow for the trapping of ions along the long axis. Ions can be ejected either radially or axially.

These analysers offer the advantage of greater trapping volumes, thereby allowing the analysis of more ions per cycle (Douglas *et al.*, 2005).

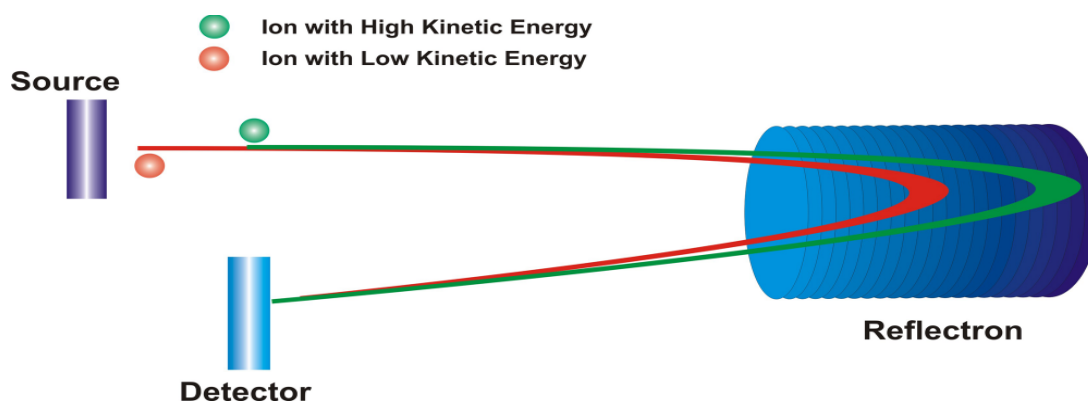


Figure 1.4 – Representation of a TOF instrument equipped with a reflectron. Both ions have the same mass, but different initial kinetic energy. The ion with the higher energy will travel further into the reflectron compared to the one with less energy; both ions will reach the detector simultaneously.

1.3.4 Fourier transform ion cyclotron resonance

Fourier transform ion cyclotron resonance (FT-ICR) instruments measure mass indirectly by oscillating ions in a strong magnetic field (Henry *et al.*, 1989). The oscillations occur as a function of m/z , meaning their frequency can be used to infer m/z using a Fourier transform. FT-ICR mass spectrometers offer the highest mass accuracy and resolution of all current instrumentation. The novel Orbitrap, introduced by Makarov in 2000, also uses an FT-based strategy to measure m/z (Makarov, 2000). The ion trapping is performed electrostatically, compared to magnetically in other FT MS approaches; this is attractive as it negates the need for a large superconducting magnet and concomitant requirements for liquid helium and nitrogen (Qizhi *et al.*, 2005).

1.4 Detectors

Once an ion beam passes through the mass analyser it is recorded and transformed into a usable signal by a detector. Various types of detector have been employed. The first mass spectrometers used photographic plates located behind the analyser; ions of the same m/z reach the plate at the same time. A calibration scale allows determination of m/z values, and darkness of spots enables approximation of beam intensity. Also used were Faraday plates. Ions reach the cylinder where they give up their charge, and the discharge current is then amplified and measured. These types of detector provide a direct measurement of the charges recorded.

Electron or photon multiplier detectors and array detectors increase the intensity of the signal. Photon multipliers comprise a phosphorescent screen and a photomultiplier, allowing the detection of both positive and negative ions with an amplification of 10^4 to 10^5 . In electron multipliers, ions which reach the plate cause the emission of secondary particles, those of interest are typically negative ions and electrons. These secondary particles are accelerated into a continuous dynode electron multiplier, causing electrons to be dislodged. A cascade of electrons is therefore generated, which results in a measurable current. An amplification of up to 10^7 can be achieved (de Hoffmann and Stroobant, 2007).

An array detector consists of a plate where parallel channels have been drilled. A widely used detector in modern mass spectrometry is the microchannel plate detector (MCP) (Wiza, 1979). MCP plates contain an array of miniature electron multiplier channels, typically placed at an 8° angle to the surface of the plate. A single ion reaching the MCP detector will affect only a few of the channels, making it possible to detect many different ions at the same time. This is useful for the analysis of complex biological samples, where hundreds of ions can be created simultaneously. Great care needs to be taken not to saturate the MCP with ion signals, as the channels must have time to recover (dead time) before they can detect new signals. If the MCP becomes saturated with a large signal, it will not be able to detect any smaller signal which may follow directly afterwards.

Modern mass spectrometers utilise analog-to-digital (ADC) or time-to-digital (TDC) converters as their detector system. ADCs register the ion current produced, amplify the signal and filter to remove high frequency noise. The current is then plotted on

an m/z scale by comparing to previously acquired calibration data; these are employed by certain Agilent (Agilent Technologies, CA, USA) and JEOL (JEOL, Tokyo, Japan) instruments. TDCs are most commonly used in ESI-MS instrumentation, and are used on all Waters (Waters, Manchester, UK) instruments. TDCs record the time at which each ion strikes the detector, individual acquisitions are then summed together to produce the observed spectrum; these offer the advantages of speed, efficiency and low noise (de Hoffmann and Stroobant, 2007).

1.5 Tandem mass spectrometry

Tandem mass spectrometry (commonly abbreviated as MS/MS) is any method involving at least two stages of mass analysis. In the most common approach, the first mass analyser is used to isolate a precursor ion, which then undergoes some sort of fragmentation to yield product ions and neutral fragments. A second mass analyser then studies the product ions. The number of steps can be increased to yield an MS^n experiment, where n refers to the number of generations of ions being analysed. In protein analysis, interpretation of the product ion spectrum can provide a peptide sequence (either partially or in full), depending on the type of fragmentation observed. The nomenclature used for describing peptide product ions was proposed by Roepstorff, Fohlmann and Biemann, and is represented in Figure 1.5 (Biemann, 1992; Roepstorff and Fohlmann, 1984).

Two types of tandem mass spectrometry can be performed: in-space and in-time. In-space experiments utilise a number of spatially separate mass analysers, such as TOF/TOF, Q-TOF and multiple quadrupoles. Four main scan modes are used, although many others are possible:

Product ion scan: A precursor ion of a certain m/z is selected and all resulting product ions are determined. The first analyser is focused and the second scans all masses.

Precursor ion scan: A product ion is chosen, and the associated precursor ions are determined. The second analyser is focused and the first scans all masses.

Neutral loss scan: A neutral fragment is selected, and all fragmentations leading to the loss of that fragment are determined. Both mass analysers operate in scanning mode, but with a constant mass offset.

Selected reaction monitoring: A fragmentation reaction is selected, and both mass analysers are focused on selected masses. Ions selected by the first analyser are only detected if they produce a given fragment by the selected reaction.

In-time experiments are performed using instruments which are capable of trapping ions, such as ion traps and FT mass spectrometers, where RF and DC fields are applied to select the ion to undergo dissociation. The peptide and resulting fragments are scanned by the same analyser.

One of the most commonly used fragmentation methods employed in MS/MS experiments is collisionally-induced dissociation (CID), pioneered by Jennings and McLafferty in the 1960s (Haddon and McLafferty, 1968; Jennings, 1968). In CID, precursor ions are isolated and subjected to collision with an inert gas such as argon.

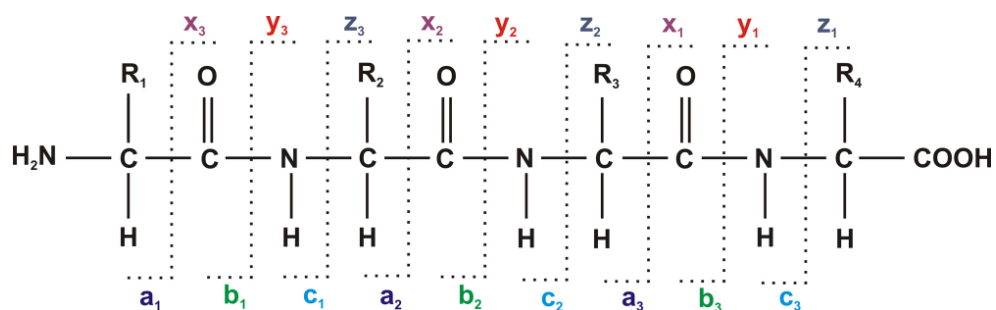


Figure 1.5 – Peptide ion nomenclature, as proposed by Roepstorff, Fohlmann and Biemann.

The energy transferred to the precursor ion upon impact with the gas molecule causes fragmentation. In the case of peptides, this occurs primarily at amide bonds to give *b* and *y* fragment ions. Initial observations of CID products were dismissed as a curiosity or nuisance, and CID was viewed as a ‘technique waiting for a problem’ until developments at the end of the 1970s. The construction of commercially available triple quadrupole instrumentation and the introduction of FAB meant that

CID became an integral part of methodology for structural studies of biological molecules. This remains the case in modern biological mass spectrometry, with CID employed in quadrupole, TOF, QIT, LIT and FT-ICR MS (Jennings, 2000).

Complementary to CID are electron capture dissociation (ECD) and electron transfer dissociation (ETD). Developed by Zubarev, ECD is based on ion-electron reactions in which the capture of electrons by a gaseous positive ion leads to fragmentation and neutralisation of the positive ion (Zubarev *et al.*, 1998). When applied to peptides, ECD leads to extensive cleavage of the backbone at N-C α bonds to yield *c* and *z* ions. ECD is an available option on commercial FT-ICR instruments. Since it is not amenable to ion trap instrumentation, the analogous ETD was developed by Hunt and colleagues. ETD is an ion-ion reaction between singly-charged anions and multiply-charged peptide cations (Syka *et al.*, 2004). Reagent anions are created by a CI source and serve as a source of electrons. The interaction of the anions with the peptides results in proton transfer without dissociation and electron transfer with or without dissociation. Proton transfer leads to charge reduction, and dissociation generates *c* and *z* ions with fragmentation characteristic of ECD.

1.6 Current approaches in biological mass spectrometry

Biological MS could consider its beginning to be when the coupling of gas chromatography (GC) first occurred in the 1950s. This allowed, for the first time, the chromatographic separation of complex biological mixtures for their identification by MS (Gohlke and McLafferty, 1993). Since then, via a number of key milestones, MS has become firmly established as an essential tool in biochemical analysis.

The advent of MALDI and ESI saw the birth of a whole new era in MS, supported by the commercial introduction of instruments such as triple quadrupoles, ion traps and updated TOFs, FT-ICR, as well as hybrid systems and the novel Orbitrap. Tandem MS methods and the coupling of MS to high performance LC were other key developments, and more recently, the use of MALDI to map biological tissue sections, and coupling MS to ion mobility separations to add the dimension of shape.

The analysis of peptides, proteins, nucleic acids, carbohydrates, lipids and metabolites by MS is now commonplace, with their spectral complexity offset by improvements in resolution and mass accuracy. Diverse MS methodology is now available, with unique capabilities for speed, sensitivity, specificity and automation. Table 1.1 highlights some of the key landmarks in the development of mass spectrometry as a tool for analysing biological systems, with particular focus on developments which will be touched upon later in this thesis. The research presented here focuses on three aspects of modern mass spectrometry for the study of biological problems:

1. The use of recently developed ambient ionisation techniques to study pharmaceutical formulations and their resulting metabolites.
2. The study of metal-containing protein complexes by inorganic mass spectrometry and by organic mass spectrometry incorporating an ion mobility separation aspect.
3. The characterisation of the proteome of a recently identified prokaryote using quantitative mass spectrometry-based proteomic techniques.

The work has been separated into three discrete sections, to reflect the self-contained nature of the projects.

1911	Sir JJ Thomson builds his parabola spectrograph and separates isotopes of neon.
1919	Francis Aston develops Thomson's design and separates isotopes of other elements.
1922	Aston awarded Nobel Prize for Chemistry
1955	TOF first described by Wiley and McLaren.
1962	Biemann and McCloskey analyse free amino acids by TOF-MS.
1965	Researchers at MIT describe the GC/MS interface.
1966	CI developed. Researchers at MIT, Manchester and Purdue use a computer to interpret mass spectra to sequence oligopeptides.
1968	Precursor work on CID (Jennings) and MS/MS in mixtures (Haddon and McLafferty)
1969	First report on MS peptide sequencing from protein hydrolysates (Lucas <i>et al.</i> , 1969)
1972	Amino acid sequencing by computer-assisted GC/MS of oligopeptide mixtures (Nau <i>et al.</i> , 1972)
1973	Development of the reflectron TOF (Mamyrin <i>et al.</i> , 1973) Peptide sequencing by enzymatic hydrolysis and GC/MS of resulting peptides
1974	Introduction of FT-ICR MS
1975	High resolution separation of proteins by two-dimensional electrophoresis (O'Farrell, 1975)
1978	Introduction of triple-quadrupole tandem MS (Yost and Enke, 1978)
1980	Inductively coupled plasma MS (ICP-MS) developed.
1981	Introduction of FAB.
1983	First commercial ion trap analyser (Finnigan). First commercial ICP-MS system (SCIEX).
1984	John Fenn and colleagues use electrospray to ionise biomolecules.
1985	Karas and Hillenkamp coin the term MALDI.
1986	First protein sequence database (SWISS-PROT).
1988	MALDI of proteins over 10,000 Da (Karas and Hillenkamp).
1989	Computer deconvolution of ESI spectrum (Mann <i>et al.</i> , 1989) First combination of ESI and FT-ICR (Henry <i>et al.</i> , 1989)
1990	First use of ESI-MS to monitor the conformation of a protein in solution (Chowdhury <i>et al.</i> , 1990)
1991	Non-covalent complexes analysed by MS (Ganem <i>et al.</i> , 1991) Development of LC-MS/MS for complex peptide mixtures (Covey <i>et al.</i> , 1991)
1993	Peptide mass fingerprinting described (James <i>et al.</i> , 1993)
1994	Direct correlation of peptide MS/MS spectra with sequences in databases (Eng <i>et al.</i> , 1994) First use of the term 'proteome' at the Sienna 2D electrophoresis meeting. (Wilm <i>et al.</i> , 1996)
1995	2D gel electrophoresis used for profiling proteomics (Klose and Kobalz, 1995)
1996	Combination of 2D gels with MS and database searching for protein identification (Wilm <i>et al.</i> , 1996). Commercial introduction of a Q-TOF instrument (Micromass) and the LCQ ion trap (Finnigan).

1997	Development of 2D differential gel electrophoresis (DIGE) (Ünlü <i>et al.</i> , 1997)
1998	Application of ECD to proteins (Zubarev <i>et al.</i> , 1998)
1999	Isotope coded affinity tags (ICAT) for protein quantification (Gygi <i>et al.</i> , 1999) Makarov presents the Orbitrap.
2001	'Gel-free' approach to profiling proteomics (Washburn <i>et al.</i> , 2001) Automated method for shotgun proteomics (MudPIT) (Wolters <i>et al.</i> , 2001) Introduction of imaging MALDI MS (Stoeckli <i>et al.</i> , 2001)
2004	Desorption electrospray ionisation (DESI) (Cooks <i>et al.</i> , 2006). Development of ETD (Syka <i>et al.</i> , 2004)
2005	Commercial introduction of the Orbitrap.
2006	Introduction of a commercial ion mobility MS instrument (Synapt from Waters).
2009	Introduction of an improved commercial ion mobility MS instrument (Waters) (Giles <i>et al.</i> , 2009)

Table 1.1 – Instrumental and experimental developments that have contributed to the study of biological systems by MS (adapted from (Gelpí, 2008) and (Gelpí, 2009))

References

- Aston, F. W.** (1919). A positive ray spectrograph. *Philos. Mag.* **38**: 707 - 714.
- Barber, M., Bordoli, R. S., Sedgwick, R. D., Tyler, A. N.** (1981). Fast atom bombardment of solids (FAB): a new ion source for mass spectrometry. *J. Chem. Soc., Chem. Comm.* **7**: 325-327.
- Biemann, K.** (1992). Mass Spectrometry of Peptides and Proteins. *Annu. Rev. Biochem.* **61**: 977-1010
- Chowdhury, S. K., Katta, V., Chait, B. T.** (1990). Probing conformational changes in proteins by mass spectrometry. *J. Am. Chem. Soc.* **112**: 9012-9013.
- Cooks, R. G., Ouyang, Z., Takats, Z., Wiseman, J. M.** (2006). Ambient Mass Spectrometry. *Science* **311**: 1566-1570.
- Covey, T. R., Huang, E. C., Henion, J. D.** (1991). Structural characterization of protein tryptic peptides via liquid chromatography/mass spectrometry and collision-induced dissociation of their doubly charged molecular ions. *Anal. Chem.* **63**: 1193-1200.
- de Hoffmann, E., Stroobant, V.** (2007). *Mass Spectrometry. Principles and Applications.* Wiley.
- Douglas, D., J., Frank, A. J., Mao, D.** (2005). Linear ion traps in mass spectrometry. *Mass Spectrom. Rev.* **24**: 1-29.
- Eng, J. K., McCormack, A. L., Yates, J. R.** (1994). An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **5**: 976-989.
- Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., Whitehouse, C. M.** (1989). Electrospray Ionization for Mass Spectrometry of Large Biomolecules. *Science* **246**: 64-71.
- Ganem, B., Li, Y. T., Henion, J. D.** (1991). Detection of noncovalent receptor-ligand complexes by mass spectrometry. *J. Am. Chem. Soc.* **113**: 6294-6296.
- Gelpí, E.** (2008). From large analogical instruments to small digital black boxes: 40 years of progress in mass spectrometry and its role in proteomics. Part I 1965-1984. *J. Mass Spectrom.* **43**: 419-435.
- Gelpí, E.** (2009). From large analogical instruments to small digital black boxes: 40 years of progress in mass spectrometry and its role in proteomics. Part II 1985-2000. *J. Mass Spectrom.* **44**: 1137-1161.
- Giles, K., Gilbert, T., Green, M., Wildgoose, J., Scott, G.** 2009. Enhancements to the performance of a hybrid quadrupole/travelling wave ion mobility/oa-TOF instrument. *18th International Mass Spectrometry Conference.* Bremen.

- Gohlke, R. S., McLafferty, F. W.** (1993). Early gas chromatography/mass spectrometry. *J. Am. Soc. Mass Spectrom.* **4**: 367-371.
- Griffiths, W. J., Jonsson, A. P., Liu, S., Rai, D. K., Wang, Y.** (2001). Electrospray and tandem mass spectrometry in biochemistry. *Biochem. J.* **355**: 545-561.
- Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H., Aebersold, R.** (1999). Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotech* **17**: 994-999.
- Haddon, W. F., McLafferty, F. W.** (1968). Metastable ion characteristics. VII. Collision-induced metastables. *J. Am. Chem. Soc.* **90**: 4745-4746.
- Henry, K. D., Williams, E. R., Wang, B. H., McLafferty, F. W., Shabanowitz, J., Hunt, D. F.** (1989). Fourier-transform mass spectrometry of large molecules by electrospray ionization. *Proc. Natl. Acad. Sci.* **86**: 9075-9078.
- James, P., Quadroni, M., Carafoli, E., Gonnet, G.** (1993). Protein Identification by Mass Profile Fingerprinting. *Biochem. Biophys. Res. Commun.* **195**: 58-64.
- Jennings, K. R.** (1968). Collision-induced decompositions of aromatic molecular ions. *Int. J. Mass Spectrom. Ion Phys.* **1**: 227-235.
- Jennings, K. R.** (2000). The changing impact of the collision-induced decomposition of ions on mass spectrometry. *Int. J. Mass Spectrom.* **200**: 479-493.
- Karas, M., Hillenkamp, F.** (1988). Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal. Chem.* **60**: 2299-2301.
- Klose, J., Kobalz, U.** (1995). Two-dimensional electrophoresis of proteins: An updated protocol and implications for a functional analysis of the genome. *Electrophoresis* **16**: 1034-1059.
- Lucas, F., Barber, M., Wolstenholme, W. A., Geddes, A. J., Graham, G. N., Morris, H. R.** (1969). Mass-spectrometric determination of the amino acid sequences in peptides isolated from the protein silk fibroin of *Bombyx mori*. *Biochem. J.* **114**: 695-702.
- Makarov, A.** (2000). Electrostatic Axially Harmonic Orbital Trapping: A High-Performance Technique of Mass Analysis. *Anal. Chem.* **72**: 1156-1162.
- Mamyrin, B. A., Karataev, V. L., Schmikk, D. V.** (1973). The mass reflectron, a new non-magnetic time-of-flight spectrometer with high resolution. *Sov. Phys. JETP* **37**.
- Mann, M., Meng, C. K., Fenn, J. B.** (1989). Interpreting mass spectra of multiply charged ions. *Anal. Chem.* **61**: 1702-1708.
- Munson, M. S. B., Field, F. H.** (2002). Chemical Ionization Mass Spectrometry. I. General Introduction. *J. Am. Chem. Soc.* **88**: 2621-2630.

- Nau, H., Kelley, J. A., Biemann, K.** (1972). Determination of the amino acid sequence of the C-terminal cyanogen bromide fragment of actin by computer-assisted gas chromatography-mass spectrometry. *J. Am. Chem. Soc.* **95**: 7162-7164.
- O'Farrell, P. H.** (1975). High resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.* **250**: 4007-4021.
- Paul, W., Steinwendel, H.** (1953). Ein neues massenspektrometer ohne magnetfeld (A new mass spectrometer without a magnetic field). *Z. Naturforsch., A: Phys. Sci.* **8**: 448-450.
- Qizhi, H., Robert, J. N., Hongyan, L., Alexander, M., Mark, H., Cooks, R. G.** (2005). The Orbitrap: a new mass spectrometer. *J. Mass Spectrom.* **40**: 430-443.
- Roepstorff, P., Fohlmann, J.** (1984). Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed, Mass Spectrom.* **11**: 601.
- Stafford, G. C., Kelley, P. E., Syka, J. E. P., Reynolds, W. E., Todd, J. F. J.** (1984). Recent improvements in and analytical applications of advanced ion trap technology. *Int. J. Mass Spectrom. Ion Processes* **60**: 85-98.
- Stoeckli, M., Chaurand, P., Hallahan, D. E., Caprioli, R. M.** (2001). Imaging mass spectrometry: A new technology for the analysis of protein expression in mammalian tissues. *Nat Med* **7**: 493-496.
- Syka, J. E. P., Coon, J. J., Schroeder, M. J., Shabanowitz, J., Hunt, D. F.** (2004). Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc. Natl. Acad. Sci.* **101**: 9528-9533.
- Tanaka, K., Waki, H., Ido, Y., Akita, S., Yoshida, Y., Yoshida, T.** (1988). Protein and polymer analyses up to m/z 100 000 by laser ionization time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* **2**: 151-153.
- Thomson, J. J.** (1899). On the masses of the ions in gases at low pressures. *Philos. Mag.* **48**: 547 - 567.
- Thomson, J. J.** (1911). Rays of positive electricity. *Philos. Mag.* **6**: 752-767.
- Ünlü, M., Morgan, M. E., Minden, J. S.** (1997). Difference gel electrophoresis. A single gel method for detecting changes in protein extracts. *Electrophoresis* **18**: 2071-2077.
- Vestal, M. L., Juhasz, P., Martin, S. A.** (1995). Delayed extraction matrix-assisted laser desorption time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* **9**: 1044-1050.
- Washburn, M. P., Wolters, D., Yates, J. R.** (2001). Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotech* **19**: 242-247.

- Wiley, W. C., McLaren, I. H.** (1955). Time-of-Flight Mass Spectrometer with Improved Resolution. *Rev. Sci. Instrum.* **26**: 1150-1157.
- Wilm, M., Shevchenko, A., Houthaeve, T., Breit, S., Schweigerer, L., Fotsis, T., Mann, M.** (1996). Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry. *Nature* **379**: 466-469.
- Wiza, J. L.** (1979). Microchannel plate detectors. *Nucl. Instrum. Methods* **162**: 587-601.
- Wolters, D. A., Washburn, M. P., Yates, J. R.** (2001). An Automated Multidimensional Protein Identification Technology for Shotgun Proteomics. *Anal. Chem.* **73**: 5683-5690.
- Yost, R. A., Enke, C. G.** (1978). Selected ion fragmentation with a tandem quadrupole mass spectrometer. *J. Am. Chem. Soc.* **100**: 2274-2275.
- Zenobi, R., Knochenmuss, R.** (1998). Ion formation in MALDI mass spectrometry. *Mass Spectrom. Rev.* **17**: 337-366.
- Zubarev, R. A., Kelleher, N. L., McLafferty, F. W.** (1998). Electron Capture Dissociation of Multiply Charged Protein Cations. A Nonergodic Process. *J. Am. Chem. Soc.* **120**: 3265-3266.

Chapter 2

Introduction to ambient ionisation mass spectrometry

2.1 Emergence of ambient ionisation techniques

One of the major limitations of mass spectrometry is the ability to transfer the sample of interest into the vacuum environment of the instrument, in the form of ions which are suitable for analysis. This problem has been partially tackled with the advent of electrospray ionisation (ESI) and matrix-assisted laser desorption ionisation (MALDI), as previously described. A more recent development is the advent of a new family of techniques which allow ions to be created under ambient conditions.

Described by Takáts *et al* in 2004 (Takáts *et al.*, 2004), desorption electrospray ionisation (DESI) is related to spray ionisation methods such as ESI and to desorption ionisation methods such as secondary ion mass spectrometry (SIMS) and MALDI. A fine spray of charged droplets hits a solid surface of interest, from which it picks up small organic molecules and large biomolecules, ionises them and delivers them, as desolvated ions, into the mass spectrometer (Cooks *et al.*, 2006). In the original experiments, an aqueous spray was directed at an insulating sample or an analyte deposited on an insulating surface such as polytetrafluoroethylene (PTFE), and the desorbed ions were sampled with an ion trap mass spectrometer equipped with an atmospheric interface (Takáts *et al.*, 2004). Advances since have seen the interfacing of other types of instrument with an ambient ionisation source in order to perform DESI experiments. An overview of the system can be seen in Figure 2.1.

Ion formation in DESI has been proposed to occur by one of two different mechanisms. One is the formation of charged solvent species in the electrospray source, followed by ionisation of the analyte molecule on the surface by charge transfer. Desorption of analyte ions from the surface is a type of chemical sputtering (Cooks *et al.*, 2004). The other mechanism involves the impact of electrosprayed droplets on the surface, dissolution of the analyte in the droplet, and subsequent evaporation by mechanisms as in ESI; this has been referred to as ‘droplet pick-up’. A modelling study investigating droplet collisions with a liquid thin film in DESI has indicated that the droplet pick-up mechanism is the dominant process, showing that the secondary droplets leaving the surface of the analyte after primary droplet impact contain both liquid from the thin film and liquid from the primary droplet (Costa and Cooks, 2007).

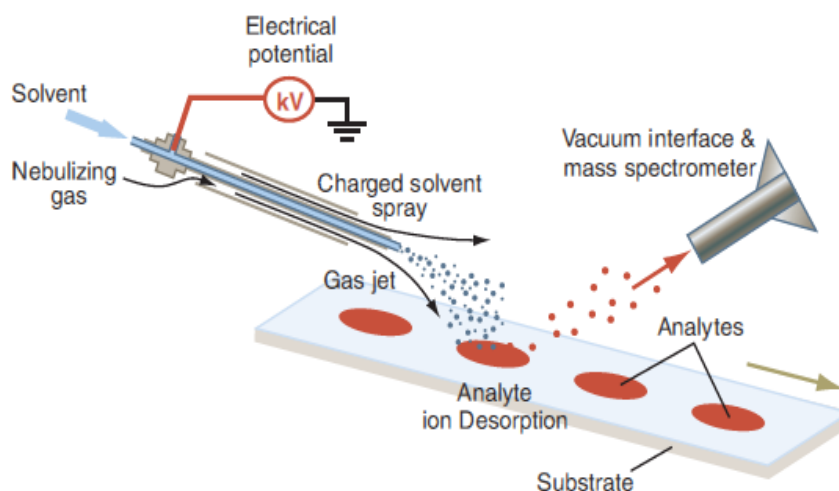


Figure 2.1 – Schematic of a typical DESI experiment. The sample is deposited on a solid surface and a spray of charged solvent is directed towards it. The desorbed analyte is then carried into the mass spectrometer. From (Cooks *et al.*, 2006).

A variant of DESI is desorption atmospheric pressure chemical ionisation (DAPCI). The electrospray emitter in a DESI source is replaced by a stainless steel needle, and ions are generated by an atmospheric pressure corona discharge in the vapour of a solvent which is mixed into the gas flow. As droplets are not formed, the ions are produced by the charge transfer mechanism (Takáts *et al.*, 2005a).

Another ambient ionisation technique which has recently emerged is the direct analysis in real time (DART) method (Cody *et al.*, 2005). An electrical potential is applied to a gas (usually nitrogen or helium) to form a plasma of excited-state electrons, ions, and metastable species. These species then desorb low-molecular weight molecules from the surface of a sample. A representation of the DART process can be seen in Figure 2.2(a). The ionisation mechanism involves the reaction of excited-state gas molecules with water in the atmosphere to produce protonated water clusters followed by proton transfer to the analyte. Other potential ionisation mechanisms that have been reported involve Penning ionisation, in which a metastable species ionises a neutral to produce a radical cation and an electron, and the interaction of a metastable species with a surface to produce electrons (Van Berkel *et al.*, 2008); these reactions are shown schematically in Figure 2.2(b). In the

negative ion mode, negative ions of analytes have been proposed to form from reactions with negatively charged oxygen/water cluster ions (Cody et al., 2005).

The emergence of DESI and DART could be described as milestones in the field of mass spectrometry, as they spearheaded the development of a whole new family of related ionisation methods. These techniques, which require little or no sample preparation, are advancing the analytical sciences by enabling complex systems to be probed without chemical separations with significant speed and sensitivity. A summary of current ambient ionisation techniques is presented in Table 2.1, giving an outline of the method and any particular advantages or issues.

The methods can be broadly grouped into three categories: liquid and gas jet desorption/ionisation, thermal desorption/ionisation, and laser desorption/ionisation.

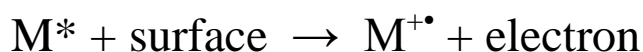
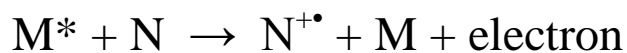
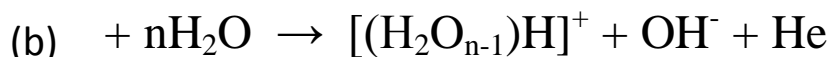
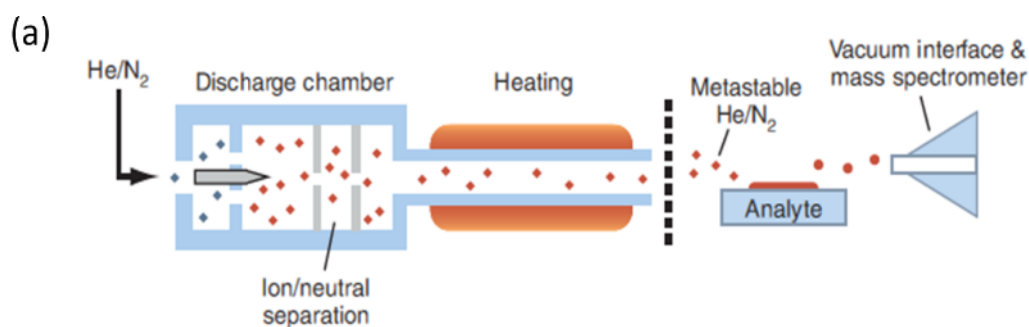


Figure 2.2 – (a) sample introduction using a DART source, from (Cooks et al., 2006), (b) the interaction of helium with water, which precedes the protonation of the analyte, and the two alternative mechanisms, which produce charged radicals and electrons, from (Cody et al., 2005).

Technique		Overview	Features
Atmospheric pressure glow discharge desorption ionisation	APGDDI	Atmospheric pressure glow discharge source used for surface sampling; thermal desorption is component from plasma heating.	
Atmospheric pressure matrix-assisted desorption ionisation	AP-MALDI	As MALDI, but performed at atmospheric pressure	Less fragmentation than vacuum MALDI due to rapid collisional cooling.
Atmospheric solids analysis probe	ASAP	Vaporisation of materials in hot nitrogen gas stream from ESI or APCI probe. Ionisation of thermally induced vapours occurs by corona discharge.	Can be installed on any commercial atmospheric pressure instrument; faster than vacuum solids probe analysis.
Atmospheric pressure thermal desorption ionisation	APTI	Heat used to liberate sample (mainly organic salts) from condensed phase directly to gas phase as ions.	
Desorption atmospheric pressure chemical ionisation	DAPCI	Gaseous solvent vapours ionised by corona discharge from stainless steel tip in DESI source.	Known charge transfer mechanism.
Desorption atmospheric pressure photoionisation	DAPPI	Heated nebuliser microchip delivers heated jet of vaporised solvent towards sample surface to desorb sample. Photons emitted from photoionisation lamp ionise analytes.	Effective for non-polar and neutral compounds.
Desorption electrospray ionisation	DESI	Aqueous spray at analyte on solid surface.	Wide range of analytes.

Desorption sonic spray ionisation	DeSSI	Polar solutions of analyte sprayed from a capillary with a supersonic nebulising gas flow.	Avoiding high voltages may prove beneficial for tissue analysis, especially <i>in situ</i> .
Dielectric barrier discharge ionisation	DBDI	Plasma generated which interacts with surface of analyte.	
Direct analysis in real time	DART	Plasma of charged species created by application of electric potential to a gas; molecules desorbed from sample surface.	
Electrospray-assisted laser desorption ionisation	ELDI	Nitrogen laser pulse desorbs intact molecules from matrix-containing sample droplets, followed by ESI post-ionisation.	Produces ESI-like charge distributions in tandem MS experiments, which provides potential for sequencing peptides and proteins.
Extractive electrospray ionisation	EESI	One sprayer nebulises the sample solution, another produces charged microdroplets of solvent. Liquid-liquid extraction occurs between the colliding droplets.	Longer-term stability of signal compared with alternative methods.
Jet desorption ionisation	JeDI	High-velocity solvent jet erodes sample surface and generates gas-phase ions.	Depth profiling capabilities.
Laser ablation with electrospray ionisation	LAESI	Combination of laser ablation and ESI. Interactions between ablation plume and ESI spray result in ESI-like ionisation.	No matrix required; laser enables depth analysis; potential imaging capabilities.
Laser desorption atmospheric pressure chemical ionisation	LD-APCI	Laser desorption to sample surface, followed by secondary ionisation via an APCI process.	

Laser desorption electrospray ionisation	LD-ESI	Much the same as LD-APCI, but desorbed material ionised through reaction with charged solvent droplets.	High secondary ionisation efficiency; ability to form multiply charged species from macromolecules.
Laser diode thermal desorption	LDTD	IR laser thermally desorbs samples deposited onto stainless steel sample wells.	Has been used for high-throughput assays.
Natural desorption extractive electrospray ionisation	NDEESI	Room temperature gas jet desorbs sample; ionisation via secondary ESI process	Predominantly used for small, volatile compounds.
Plasma-assisted desorption ionisation	PADI	Non-thermal plasma generated which interacts with surface of analyte.	Close to ambient temperature; tolerant of contaminants.
Thermal desorption atmospheric pressure chemical ionisation	TD-APCI	Heat used to liberate sample from condensed phase; once in gas phase, ionisation occurs by APCI.	

Table 2.1 – Summary of currently available ambient ionisation methods.

2.2 Liquid and gas jet desorption/ionisation

The development of the liquid and gas desorption/ionisation group of techniques was triggered by the advent of DESI in 2004. Any solvent appropriate for use in ESI can be used for DESI. Solvent additives such as volatile acids and bases, non-volatile salts or other chemical reagents can be used to enhance detection, a process referred to as reactive DESI. By switching off the variable high voltage in DESI and providing sufficient nebulising gas velocity, the desorption sonic spray ionisation (DeSSI) source is created (Haddad *et al.*, 2006). SSI uses only the force of a high-velocity nebulising gas to generate gas-phase ions from species in solution (Hirabayashi *et al.*, 2002). As no voltages are used, the spray plume is nearly neutral, with a distribution of positively- and negatively-charged droplets. This allows the analysis of both positive and negative ions without the need to rapidly switch high voltages in the source. Somewhat confusingly, the technique has also been referred to as easy ambient sonic spray ionisation (EASI) (Haddad *et al.*, 2008). This method is of interest as biological relevant charge states have been absorbed, but the poor sensitivity achieved has prevented its wider use.

At the extreme of liquid flow is jet desorption ionisation (JeDI), where a fused silica capillary is used to generate a continuous liquid jet. This is orientated in a position relative to the surface to be analysed, similar to the emitter in DESI. A voltage is applied at the emitter to the solution that is pumped through. The high-velocity jet stream continuously erodes the sample surface and generates gas-phase ions, making this technique an interesting option for depth profiling (Takáts *et al.*, 2006).

The method of neutral desorption extractive electrospray ionisation (NDEESI) retains the general desorption and sampling geometry of a DESI experiment, but eliminates both the liquid and high voltage. A room temperature gas jet is used to accomplish the desorption step, and the ionisation of desorbed material occurs by a secondary ESI process (Chen, Huanwen *et al.*, 2007). Predominantly, small, volatile compounds such as amines from spoiled meat have been analysed. Some less volatile explosives have, however, been detected directly from human skin. The limitation of the technique is ultimately what can be aerosolised and transported to the ion source by the gas stream, making volatile and semi-volatile species most amenable.

2.3 Thermal desorption/ionisation

Thermal desorption atmospheric pressure chemical ionisation (TD/APCI) is a sampling technique introduced in the 1970s. Commercially available in the 1980s, it was largely forgotten in the 1990s, but has had reincarnations in recent years. Heat is used to liberate the sample intact from the condensed phase to the vapour phase. The heating is most commonly accomplished through the use of a heated gas passing over the sample. Once in the gas phase, the sample can be ionised; typically, this is by atmospheric pressure chemical ionisation (APCI) (Moini, 2007). It has recently been shown that some organic salts can be thermally desorbed into the gas phase directly as ions, a process termed atmospheric pressure thermal desorption ionisation (APTDI) (Chen, Hao *et al.*, 2006). At the outset, samples were analysed by direct vapour detection and collection of trace particle residue or vapours, followed by TD and corona discharge APCI. Of late, commercial ‘plug-and-play’ corona discharge APCI sources designed for liquid introduction can be used to sample materials from surfaces. McEwen *et al.*, modified such a commercial source to allow the insertion of a glass melting point capillary into a heated gas stream emerging from the heated nebuliser, to allow the rapid analysis of liquids and (semi)volatile solid materials; they named this approach atmospheric pressure solids analysis probe (ASAP) (McEwen *et al.*, 2005). These techniques have proved suitable for the analysis of lipids, capsaicins and carotenoids from fresh biological samples, polymer additives, fatty acids and pharmaceuticals. A new form of TD/APCI has emerged, called laser diode thermal desorption (LDTD), which uses an IR laser to thermally desorb samples that have been deposited onto stainless steel sample wells in a plate (Wu *et al.*, 2007). Thermally desorbed species are carried by heated air through a transfer tube and are ionised by a corona discharge source at the inlet of the mass spectrometer. LDTD has been used as a high-throughput method for the analysis of enzyme inhibition assays.

The aforementioned technique of DART can also be categorised as a thermal desorption/ionisation method. Although the desorption process involved in DART has not been categorically stated, it is likely that TD is a dominant process. DAPCI can also be seen as a simple TD/APCI system, as the sheath gas involved may be heated. A technique similar to DAPCI is desorption atmospheric pressure

photoionisation (DAPPI), where the reagent ion population is initiated by a photoionisation process rather than by corona discharge (Haapala *et al.*, 2007).

Certain plasma-assisted techniques have also emerged, which have, at least in part, a TD component. In plasma-assisted desorption/ionisation (PADI) (Ratcliffe *et al.*, 2007) and dielectric barrier discharge ionisation (DBDI) (Na *et al.*, 2007), a plasma is created in a flowing stream of helium by applying an alternating voltage between two electrodes. The recently described method of atmospheric pressure glow discharge desorption ionisation (APGDDI) utilises an AP glow discharge source for surface sampling, where there is a TD component from plasma heating of the sample (Andrade *et al.*, 2008).

2.4 Laser desorption/ionisation

In laser desorption/ionisation, a pulsed laser beam is focussed at a small spot on the surface to be analysed, material is desorbed/ionised by the laser pulse and the ions generated are analysed by a suitable mass spectrometer. The exact mechanisms of this are often complex and not fully understood. It is common for more neutral species to be generated than ions from a typical laser desorption/ionisation process. The use of a distinct secondary ionisation process therefore allows for the desorption and ionisation conditions to be independently optimised. Such methods can be subdivided into two categories: atmospheric pressure laser desorption (ablation) with secondary ionisation (AP-LD/SI) and atmospheric pressure matrix-assisted laser desorption ionisation (AP-MALDI).

2.4.1 Atmospheric pressure laser desorption (ablation) with secondary ionisation (AP-LD/SI)

Laser ablation inductively coupled plasma (LA-ICP) is the oldest and most established of the AP-LD/SI systems, and is increasingly being employed for the study of biological samples. There is a more focussed discussion of this area in Chapter 5 and therefore will not be elaborated upon here.

Laser desorption atmospheric pressure chemical ionisation (LD/APCI) uses laser desorption to sample a surface, followed by secondary ionisation via an APCI process to produce molecular ionic species. This was initially used in the 1980s for analysis of small biological molecules such as amino acids and vitamins; it has been utilised for screening analytes separated on thin layer chromatography (TLC) plates, and for the analysis of proteins and peptides directly from polyacrylamide gels (Kolaitis and Lubman, 1986). The method of laser desorption electrospray ionisation (LD/ESI) was first reported under the name electrospray-assisted laser desorption/ionisation (ELDI) (Shiea *et al.*, 2005). It works in much the same way as LD/APCI, except that the desorbed material is ionised through reaction with the charged solvent droplets, protonated solvent species, or gas-phase ions produced in the ESI process. This technique offers the advantages of high secondary ionisation efficiency and the ability to form multiply charged species from macromolecules such as proteins. It has, so far, been used for the direct characterisation of chemical compounds on TLC plates, and the detection of intact proteins in biological fluids, bacterial cultures and tissues (Peng *et al.*, 2007). Adding a chemical matrix to the samples converts the process to matrix-assisted laser desorption electrospray ionisation (MALDESI) (Sampson *et al.*, 2006).

While ELDI and MALDESI use a UV laser, a technique has been introduced which uses an IR laser. Termed laser ablation with electrospray ionisation (LAESI) (Nemes and Vertes, 2007) or infrared laser assisted desorption electrospray ionisation (IR-LD/ESI) (Rezenom *et al.*, 2008), this allows for the direct analysis of ‘wet’ biological samples. The method appears to work well on samples with high water content (e.g. fruit), whilst success has been limited with dried or less water-rich analytes, such as bone.

2.4.2 Atmospheric pressure matrix-assisted laser desorption ionisation (AP-MALDI)

First reported in 2000, the use of AP-MALDI (Laiko *et al.*, 2000) has increased significantly as commercial easily interfaced technology has become widely available. The coupling of AP-MALDI sources with ion traps, orthogonal acceleration (oa) Q-TOFs, and hybrid instruments takes advantage of their high duty

cycle and capabilities to perform MS/MS and MSⁿ experiments. Papers published indicate that this technique is therefore being used to provide structural information from analytes. As with other laser desorption/ionisation techniques, AP-MALDI results in less fragmentation than vacuum MALDI, due to the rapid collisional cooling that occurs at atmospheric pressure (Van Berkel *et al.*, 2008).

2.5 Applications of ambient ionisation techniques

Ambient ionisation methods have many advantages that make them attractive to a number of areas of analysis. They are applicable to solids, liquids, frozen solutions and adsorbed gases, usually with little or no sample preparation, and with high sensitivity and a virtually instantaneous response time (Harris *et al.*, 2008).

A point of focus for ambient ionisation techniques has been the area of forensics, such as the detection of explosives, toxic compounds and chemical warfare agents (Takáts *et al.*, 2005b). Early *in situ* MS methods employed trace atmospheric gas analysers, including transportable triple-quadrupole instruments, for the analysis of gas-phase trinitrotoluene (TNT). In these types of applications, the sensitivity, specificity, speed of response and lack of sample preparation are major advantages, all of which can be offered by various ambient ionisation methods. Explosives and chemical warfare agents have been analysed by DESI from surfaces such as metal, brick, paper, cloth and skin (Ifa *et al.*, 2009). The initial publication on DART also demonstrated the direct detection of solid-phase explosives and toxic industrial compounds (Cody *et al.*, 2005).

Other applications have included the identification of natural products in plant material, food adulterants, and chemical imaging by DESI. The latter is a particularly interesting application that has been used to profile analyte bands on TLC plates (Van Berkel *et al.*, 2005), inked lettering and images on paper and fingerprints. DESI is capable of producing an image based on any particular ion within a mass spectrum, recorded at different positions within a latent fingerprint. Endogenous compounds such as fatty acids and lipids, known to be present on the skin, can be detected and identified. Several ambient ionisation methods have been reported to be safe enough for the direct analysis of skin (Ifa *et al.*, 2008). EESI has been used to

detect caffeine from skin following coffee consumption, and cocaine has been detected from skin (Chen, Huanwen et al., 2007). The analysis of phospholipid distributions in rat brain thin tissue sections indicates there is potential for DESI to become a valuable analytical technique for identifying compounds, both endogenous and exogenous, in tissues. DESI imaging has achieved a lateral resolution of 200 μm , although competitive MALDI imaging has shown resolutions of 100 μm to 600 nm. Detection levels and sensitivity will need significant improvement before this can be implemented in a practical, clinical manner.

High throughput analysis of pharmaceuticals, and the identification of drugs and drug metabolite from blood and other biological fluids, is another area of interest. Early DART and DESI publications demonstrated the ability to interrogate compounds directly from tablets and other formulations (Rodriguez-Cruz, 2006), and the techniques have since been used to identify illicit and abused drugs in biological fluids (Kauppila *et al.*, 2007). Counterfeit pharmaceutical drugs are an increasing public health problem, with consumption sometimes causing death (Fernandez *et al.*, 2007). Older methods of tablet analysis are often inconvenient, with LC-MS analyses requiring extensive sample preparation and chromatography time, as well as expensive equipment. The minimal sample preparation and rapid analysis time of ambient ionisation methods offer an attractive alternative, particularly as ionisation sources and instruments become more available and affordable.

Proteins, protein complexes, carbohydrates, oligonucleotides, industrial polymers and small organic molecules have all been analysed by DESI and other ambient ionisation techniques. There is no doubt that ambient ionisation methods have made an important impact in the field of mass spectrometry. There remains, nevertheless, much to be learned about the fundamental mechanisms of many of these methods. If tackled, this could pave the way for further development and optimisation of these analytical tools.

2.6 Aims and objectives

The field of ambient ionisation techniques in mass spectrometry emerged only five years ago and since then has undergone rapid and considerable growth. The work presented here was conducted from 2005 to 2006, in the early stages of method development when both fundamentals and potential applications were being deduced.

The specific aims of my work were:

1. To evaluate the techniques of DESI, DAPCI and DART using commonly available pharmaceuticals.
2. To increase information content available from DESI experiments by using fragmentation and neutral loss/precursor scanning.
3. To investigate the merits of coupling DESI to a then novel polarity switching mass spectrometry approach.

This work has been peer-reviewed and published as the following articles:

Williams *et al.*, (2006) *The use of recently described ionisation techniques for the rapid analysis of some common drugs and samples of biological origin*, Rapid Communications in Mass Spectrometry **20**: 1447-1456

Williams *et al.*, (2006) *Collision-induced fragmentation pathways including odd-electron ion formation from desorption electrospray ionisation generated protonated and deprotonated drugs derived from tandem accurate mass spectrometry*, Journal of Mass Spectrometry **41**: 1277-1286

Williams *et al.*, (2006) *Polarity switching accurate mass measurement of pharmaceutical samples using desorption electrospray ionisation and a dual ion source interfaced to an orthogonal acceleration time-of-flight mass spectrometry*, Analytical Chemistry **78**: 7440-7445

References

- Andrade, F. J., Shelley, J. T., Wetzel, W. C., Webb, M. R., Gamez, G., Ray, S. J., Hieftje, G. M.** (2008). Atmospheric Pressure Chemical Ionization Source. 2. Desorption Ionization for the Direct Analysis of Solid Compounds. *Anal. Chem.* **80**: 2654-2663.
- Chen, H., Ouyang, Z., Cooks, R. G.** (2006). Thermal Production and Reactions of Organic Ions at Atmospheric Pressure. *Angew. Chem.* **45**: 3656-3660.
- Chen, H., Wortmann, A., Zenobi, R.** (2007). Neutral desorption sampling coupled to extractive electrospray ionization mass spectrometry for rapid differentiation of biosamples by metabolomic fingerprinting. *J. Mass Spectrom.* **42**: 1123-1135.
- Cody, R. B., Laramée, J. A., Durst, H. D.** (2005). Versatile New Ion Source for the Analysis of Materials in Open Air under Ambient Conditions. *Anal. Chem.* **77**: 2297-2302.
- Cooks, R. G., Jo, S. C., Green, J.** (2004). Collisions of organic ions at surfaces. *Appl. Surf. Sci.* **231-232**: 13-21.
- Cooks, R. G., Ouyang, Z., Takats, Z., Wiseman, J. M.** (2006). Ambient Mass Spectrometry. *Science* **311**: 1566-1570.
- Costa, A. B., Cooks, R. G.** (2007). Simulation of atmospheric transport and droplet-thin film collisions in desorption electrospray ionization. *Chem. Commun.*: 3915-3917.
- Fernandez, F. M., Green, M. D., Newton, P. N.** (2007). Prevalence and Detection of Counterfeit Pharmaceuticals: A Mini Review. *Ind. Eng. Chem. Res.* **47**: 585-590.
- Haapala, M., Pol, J., Saarela, V., Arvola, V., Kotiaho, T., Ketola, R. A., Franssila, S., Kauppila, T. J., Kostianen, R.** (2007). Desorption Atmospheric Pressure Photoionization. *Anal. Chem.* **79**: 7867-7872.
- Haddad, R., Sparrapan, R., Eberlin, M. N.** (2006). Desorption sonic spray ionization for (high) voltage-free ambient mass spectrometry. *Rapid Commun. Mass Spectrom.* **20**: 2901-2905.
- Haddad, R., Sparrapan, R., Kotiaho, T., Eberlin, M. N.** (2008). Easy Ambient Sonic-Spray Ionization-Membrane Interface Mass Spectrometry for Direct Analysis of Solution Constituents. *Anal. Chem.* **80**: 898-903.
- Harris, G. A., Nyadong, L., Fernandez, F. M.** (2008). Recent developments in ambient ionization techniques for analytical mass spectrometry. *The Analyst* **133**: 1297-1301.
- Hirabayashi, A., Sakairi, M., Koizumi, H.** (2002). Sonic spray mass spectrometry. *Anal. Chem.* **67**: 2878-2882.

- Ifa, D., Jackson, A., Paglia, G., Cooks, R.** (2009). Forensic applications of ambient ionization mass spectrometry. *Anal. Bioanal. Chem.* **394**: 1995-2008.
- Ifa, D., Manicke, N. E., Dill, A. L., Cooks, R. G.** (2008). Latent Fingerprint Chemical Imaging by Mass Spectrometry. *Science* **321**: 805-.
- Kauppila, T. J., Talaty, N., Kuuranne, T., Kotiaho, T., Kostianen, R., Cooks, R. G.** (2007). Rapid analysis of metabolites and drugs of abuse from urine samples by desorption electrospray ionization-mass spectrometry. *The Analyst* **132**: 868-875.
- Kolaitis, L., Lubman, D. M.** (1986). Detection of nonvolatile species by laser desorption atmospheric pressure mass spectrometry. *Anal. Chem.* **58**: 2137-2142.
- Laiko, V. V., Baldwin, M. A., Burlingame, A. L.** (2000). Atmospheric Pressure Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry. *Anal. Chem.* **72**: 652-657.
- McEwen, C. N., McKay, R. G., Larsen, B. S.** (2005). Analysis of Solids, Liquids, and Biological Tissues Using Solids Probe Introduction at Atmospheric Pressure on Commercial LC/MS Instruments. *Anal. Chem.* **77**: 7826-7831.
- Moini, M.** 2007. Atmospheric pressure chemical ionization: principles, instrumentation, and applications. In: Caprioli R, ed. *Encyclopedia of Mass Spectrometry*. Amsterdam: Elsevier. 344.
- Na, N., Zhao, M., Zhang, S., Yang, C., Zhang, X.** (2007). Development of a Dielectric Barrier Discharge Ion Source for Ambient Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **18**: 1859-1862.
- Nemes, P., Vertes, A.** (2007). Laser Ablation Electrospray Ionization for Atmospheric Pressure, in Vivo, and Imaging Mass Spectrometry. *Anal. Chem.* **79**: 8098-8106.
- Peng, I. X., Shiea, J., Loo, R. R. O., Loo, J. A.** (2007). Electrospray-assisted laser desorption/ionization and tandem mass spectrometry of peptides and proteins. *Rapid Commun. Mass Spectrom.* **21**: 2541-2546.
- Ratcliffe, L. V., Rutten, F. J. M., Barrett, D. A., Whitmore, T., Seymour, D., Greenwood, C., Aranda-Gonzalvo, Y., Robinson, S., McCoustra, M.** (2007). Surface Analysis under Ambient Conditions Using Plasma-Assisted Desorption/Ionization Mass Spectrometry. *Anal. Chem.* **79**: 6094-6101.
- Rezenom, Y. H., Dong, J., Murray, K. K.** (2008). Infrared laser-assisted desorption electrospray ionization mass spectrometry. *The Analyst* **133**: 226-232.
- Rodriguez-Cruz, S. E.** (2006). Rapid analysis of controlled substances using desorption electrospray ionization mass spectrometry. *Rapid Commun. Mass Spectrom.* **20**: 53-60.

- Sampson, J. S., Hawkrige, A. M., Muddiman, D. C.** (2006). Generation and Detection of Multiply-Charged Peptides and Proteins by Matrix-Assisted Laser Desorption Electrospray Ionization (MALDESI) Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **17**: 1712-1716.
- Shiea, J., Huang, M.-Z., HSu, H.-J., Lee, C.-Y., Yuan, C.-H., Beech, I., Sunner, J.** (2005). Electrospray-assisted laser desorption/ionization mass spectrometry for direct ambient analysis of solids. *Rapid Commun. Mass Spectrom.* **19**: 3701-3704.
- Takáts, Z., Cotte-Rodriguez, I., Talaty, N., Chen, H., Cooks, R. G.** (2005a). Direct, trace level detection of explosives on ambient surfaces by desorption ionization mass spectrometry. *Chem. Commun.:* 1950-1952.
- Takáts, Z., Czuczy, N., Katona, M., Skoumal, R.** 2006. Jet desorption ionization. *Proceedings of the 54th ASMS Conference on Mass Spectrometry.* Seattle, USA.
- Takáts, Z., Wiseman, J. M., Cooks, R. G.** (2005b). Ambient mass spectrometry using desorption electrospray ionization (DESI): instrumentation, mechanisms and applications in forensics, chemistry, and biology. *J. Mass Spectrom.* **40**: 1261-1275.
- Takáts, Z., Wiseman, J. M., Gologan, B., Cooks, R. G.** (2004). Mass Spectrometry Sampling Under Ambient Conditions with Desorption Electrospray Ionization. *Science* **306**: 471-473.
- Van Berkel, G. J., Ford, M. J., Deibel, M. A.** (2005). Thin-Layer Chromatography and Mass Spectrometry Coupled Using Desorption Electrospray Ionization. *Anal. Chem.* **77**: 1207-1215.
- Van Berkel, G. J., Pasilis, S. P., Ovchinnikova, O.** (2008). Established and emerging atmospheric pressure surface sampling/ionization techniques for mass spectrometry. *J. Mass Spectrom.* **43**: 1161-1180.
- Wu, J., Hughes, C. S., Picard, P., Letarte, S., Gaudreault, M., Levesque, J.-F., Nicoll-Griffith, D. A., Bateman, K. P.** (2007). High-Throughput Cytochrome P450 Inhibition Assays Using Laser Diode Thermal Desorption-Atmospheric Pressure Chemical Ionization-Tandem Mass Spectrometry. *Anal. Chem.* **79**: 4657-4665.

Chapter 3

Materials and Methods

Ambient Ionisation Mass Spectrometry

3.1 Overview of ambient ionisation study

Three different studies were carried out: rapid screening of common drugs using different ambient ionisation techniques, collision-induced dissociation and polarity switching accurate mass measurements.

Non-prescription medications were obtained from commercial suppliers; prescription drugs were obtained under permission from medical professionals.

The complete complement of samples analysed has been summarised in Table 3.1, with corresponding active ingredients detailed in Table 3.2.

3.2 Direct analysis in real time (DART)

DART experiments were carried out on an AccuToF LC TOF mass spectrometer (Jeol, Peabody, MA, USA). A detailed description of the DART source can be found elsewhere (Cody *et al.*, 2005).

The AccuToF instrument was operated with helium flowing into the DART source and a voltage of 2 kV applied to the discharge needle in positive mode of ionisation. Orifice 1 of the interface was set to 27 eV. This voltage can be increased or decreased depending on the amount of fragmentation desired. The gas temperature was maintained at 80 °C and the operating resolution of the instrument was approximately 6000 full width at half maximum (FWHM). Mass spectra were acquired over the mass range m/z 50 – 500 at an acquisition rate of 0.5 spectra per second. For sample analysis, the helium gas was directed towards the sample or allowed to interact with vapour-phase samples. Tablets were broken to expose an uncoated sample surface, before being held with tweezers in the path of the flowing helium at atmospheric pressure. Samples in solution were analysed by placing filter paper (1 cm - 8 cm) in the solution prior to being held in a similar manner. For ointments, approximately 100 mg was applied to the surface of a piece of matt-finished cardboard (1 cm - 2 cm) and held in the same position.

3.3 Desorption electrospray ionisation (DESI)

Experiments were carried out on a Q-ToF I (Waters, Manchester, UK). The instrument was operated in positive and negative mode with a capillary voltage of 3.5 kV and -3.2 kV respectively. The ion source block and nitrogen desolvation gas temperatures were set to 100 °C and 400 °C respectively, and the desolvation gas was set to a flow rate of 300 L/h. The cone voltage was set at 20 V for both mass spectrometry (MS) and tandem mass spectrometry (MS/MS) experiments, and the collision energy used for MS/MS experiments was ramped between 10 and 25 eV during the acquisition. The TOF mass analyser was operated at a resolution of approximately 6000 (FWHM), with spectra acquired over the mass range m/z 50 – 500 at an acquisition rate of 1 spectrum/s. For all MS/MS experiments, argon was used as the collision gas. Each tablet was broken to expose an uncoated sample surface, before being held with tweezers, at an angle of approximately 45° to the solvent spray and a distance of 5 mm from the source sampling cone. Approximately 100 mg of the ointment was applied to the surface of a piece of matt-finished cardboard (1 cm - 2 cm) and held in the same position as the solid tablets. The surface of the tablet or card was then sprayed with a solution of acetonitrile/H₂O (1:1) in negative mode and a solution of acetonitrile/H₂O with 0.2 % formic acid in positive mode at a flow rate of 10 mL/min, using a model 22 syringe pump (Harvard Apparatus, MA, USA). No extensive modification of solvents, buffers and pH was carried out.

3.4 Desorption atmospheric pressure chemical ionisation (DAPCI)

Experiments were carried out on a Q-TOF I (Waters, UK). The instrument was operated in positive and negative mode with a capillary voltage of 3.5 kV and -3.0 kV respectively. The cone voltage was optimised between 10 and 25 V for each sample. The collision energy used for MS/MS experiments was ramped between 10 and 25 eV during the acquisition. The flow rate of the nitrogen desolvation gas was set to 150 L/h. The source and probe temperatures were set to 100 °C and 400 °C, respectively. A solvent mixture of methanol and water (1:1) flowing at 10 mL/min was infused into the heated nebuliser probe where it was converted into an aerosol

which was rapidly heated in a stream of nitrogen gas, forming a vapour at the probe tip. The probe tip directly faced the tablet or ointment (which had been deposited onto the card) positioned between the corona discharge pin and the sampling cone. Reagent ions formed in the corona discharge region reacted with desorbed analyte molecules from the tablet or card forming, depending on the ionisation mode, for the most part, protonated or deprotonated molecules. The same experiments were also performed without any solvent flowing into the heated probe. This solventless DAPCI experiment is similar to the ASAP experiment previously described.

3.5 Accurate mass measurement protocol for the Q-TOF I

Instrumental mass drift was corrected for by using a single internal reference lock mass in MS and MS/MS mode on the Q-TOF. Since the target compounds in this study are known, the precursor ion selected for MS/MS experiments provided the internal reference lock mass in MS/MS mode. Data acquisition and processing were carried out using the digital dead-time correction algorithm embedded in the operating software (MassLynx v3.5, Waters, UK). The high ion counts generated using the DAPCI and DESI techniques caused time-to-digital (TDC) dead-time saturation. The TDC correction software was utilised and displayed a peak centroid with the correct mass and signal intensity.

3.6 Collision-induced dissociation pathways

Experiments were performed in a triple-quadrupole (Q_1qQ_2 , where q is a hexapole) mass spectrometer (Quattro Ultima, Waters, UK) and a hybrid quadrupole time-of-flight mass spectrometer (Q-TOF I, Waters, UK). Experimental details using the Q-TOF I were as previously described. These experiments were performed at Waters Technologies, Manchester, UK with Dr. Brian Green.

The Q_1qQ_2 instrument was equipped with the standard Z-spray electrospray ion source and operated at a source and desolvation temperature of 110 °C and 250 °C, respectively. The desolvation gas was set to a flow rate of 200 L/h. The instrument was operated in the positive and negative modes with capillary voltages of 3.0 kV and -2.6 kV, respectively. MS/MS was carried out using argon as collision gas at a

pressure of 2.5×10^{-3} mbar within the radio frequency-only hexapole collision cell. Each tablet was broken to expose an uncoated sample surface, before being held with tweezers, at an angle of approximately 45° to the solvent spray. The surface of the tablet was sprayed with a solution of acetonitrile/H₂O + 0.2% formic acid at a flow rate of 10 μ l/min using a Model 22 syringe pump (Harvard Apparatus, USA). Mass spectra were acquired in the MCA mode at an acquisition rate of 1 spectrum per 8 seconds. Data acquisition and processing were carried out using MassLynx v3.5.

3.7 Polarity switching accurate mass measurement

This was performed at Waters Technologies, Manchester, UK, with Dr. Richard Lock. Experiments were carried out using a LCT Premier orthogonal acceleration time-of-flight mass spectrometer (oa-TOFMS) fitted with a two-way electrospray ionization source (LockSpray) (Waters, Manchester, UK). The instrument was operated in positive and negative modes with a capillary voltage of 3.0 and -2.5 kV, respectively. The ion source block and nitrogen desolvation gas temperature were set to 120 °C and 350 °C, respectively, and the desolvation gas was set to a flow rate of 300 L/h. The TOF mass analyser was tuned in W-optic mode for an operating resolution of 10,000 (FWHM). Mass spectra were acquired at an acquisition rate of 1 spectrum per 150 ms with an interscan delay of 50 ms in centroid mode.

For solid sample analysis, each tablet was broken to expose an uncoated sample surface, before being held with tweezers, at an angle of 45° to the solvent spray emanating from inlet 1.

The source was open to the laboratory atmosphere to allow manual introduction of the samples. Approximately 100 mg of the ointment or 50 μ L of the liquid medicine was applied to the surface of a piece of matt-finished cardboard (1 - 2 cm) and held in the same position as the solid tablets. The surface of the tablet or card was then sprayed with a solution of acetonitrile/H₂O with 0.1 % formic acid at a flow rate of 10 μ L/min, using an Acquity Binary Solvent Manager pump (Waters, UK). Accurate mass measurement was provided by infusing leucine enkephalin ($[M+H]^+$ 556.2771 and $[M-H]^-$ 554.2615) into inlet two of the dual source and used as a single-point lockmass against which any subsequently acquired mass spectra were mass measured in real time. The lock mass was infused at a flow rate of 5 μ L/min at a concentration

of 200 pg/μL (acetonitrile/H₂O with 0.1 % formic acid) using a single piston Waters Reagent Manager pump. No significant modification of solvents, buffers, and pH was carried out.

References

Cody, R. B., Laramée, J. A., Durst, H. D. (2005). Versatile New Ion Source for the Analysis of Materials in Open Air under Ambient Conditions. *Anal. Chem.* **77**: 2297-2302.

Chapter 4

Results

Ambient Ionisation Techniques

4.1 Rapid screening of common pharmaceuticals and samples of biological origin

An investigation was conducted on a number of prescription and non-prescription pharmaceuticals to evaluate the potential of selected ambient ionisation techniques for the rapid analysis of active ingredients. In addition, human urine was analysed post-ingestion of a non-steroidal anti-inflammatory drug (NSAID). DART, DESI and DAPCI (with and without solvent) were utilised for the study. Any proposed fragmentation schemes are not assumed to be sequential.

4.1.1 Analysis of solid tablets

A solid tablet of Anadin Extra (Wyeth, UK) containing 300 mg aspirin, 45 mg caffeine and 200 mg paracetamol was analysed by DESI, DAPCI and DART, all operated in positive ion mode. Figure 4.1 shows representative spectra obtained for all three techniques. The protonated species for each of the active ingredients can be seen in the DESI and DAPCI spectra, but the protonated aspirin is missing from the DART spectrum. There are a number of other ions observed, which have been labelled in each of the spectra. The peak at m/z 163 corresponds to the loss of water from the protonated aspirin; this is seen in all three spectra, despite the absence of protonated aspirin in the DART analysis. Ammoniated aspirin is seen in the DESI and DAPCI spectra at 198, with the additional observation of sodiated aspirin in the DESI spectrum at 203. The peak at 121, seen in all three spectra, relates to an odd-electron cation from aspirin; the characterisation of this ion will be discussed later. The base peak in the DESI and DAPCI spectra is the protonated caffeine at 195. This was selected for MS/MS, performed by increasing the energy in the collision cell to cause fragmentation by CID. The DAPCI MS/MS spectrum is shown in Figure 4.2. The molecule appears to fragment by the loss of methyl isocyanate (57 Da) to form an even-electron ion at 138. This fragment produces the ion at 110 through the loss of CO.

A Solpadeine Max tablet (GlaxoSmithKline, UK) containing 12.8 mg codeine phosphate and 500 mg paracetamol was also analysed by DESI, DAPCI and DART in positive mode. Representative spectra can be seen in Figure 4.3(a)-(c). The

protonated species for both active ingredients are observed in all three spectra. The DESI and DAPCI analyses also show the dimer of paracetamol at m/z 303. In these spectra, the dimer is more abundant than codeine, whereas using DART the dimer is barely visible. The DAPCI MS/MS spectrum of codeine is shown in Figure 4.3(d); it is very complex, but was shown to be reproducible. The assignments for some of the product ions obtained are given in Table 4.1.

Measured mass	Formula
266.1196	C ₁₇ H ₁₆ NO ₂
225.0922	C ₁₅ H ₁₃ O ₂
215.1094	C ₁₄ H ₁₅ O ₂
199.0765	C ₁₃ H ₁₁ O ₂
183.0800	C ₁₃ H ₁₁ O

Table 4.1 – Fragments from DAPCI MS/MS with proposed elemental structures

4.1.2 Detection of active ingredients in a topical ointment

Proctosedyl is an ointment used for the treatment of haemorrhoids, containing the active ingredients cinchocaine hydrochloride and hydrocortisone. In this investigation, the ointment was analysed by both positive and negative ion DAPCI, spectra from which are shown in Figure 4.4. High ion counts were generated in both modes, indicating the sensitivity of the technique. The base peak in both spectra corresponds to cinchocaine, but hydrocortisone is only observed by positive ion DAPCI. The ions in both acquisitions corresponding to cinchocaine were selected for MS/MS analyses, the spectra from which can be seen in Figure 4.5. In the negative ion mode experiment the prominent peaks are at m/z 200 and 144, corresponding to the loss of C₇H₁₄N₂O from the deprotonated molecule, and further fragmentation loss of C₄H₈. The positive ion mode experiment yielded a product ion at 271, corresponding to the loss of C₄H₁₁N from the protonated molecule. This then undergoes the same 56 Da loss as seen in the negative ion mode to give a fragment of 215. The proposed fragmentation pathways are shown in Figure 4.6.

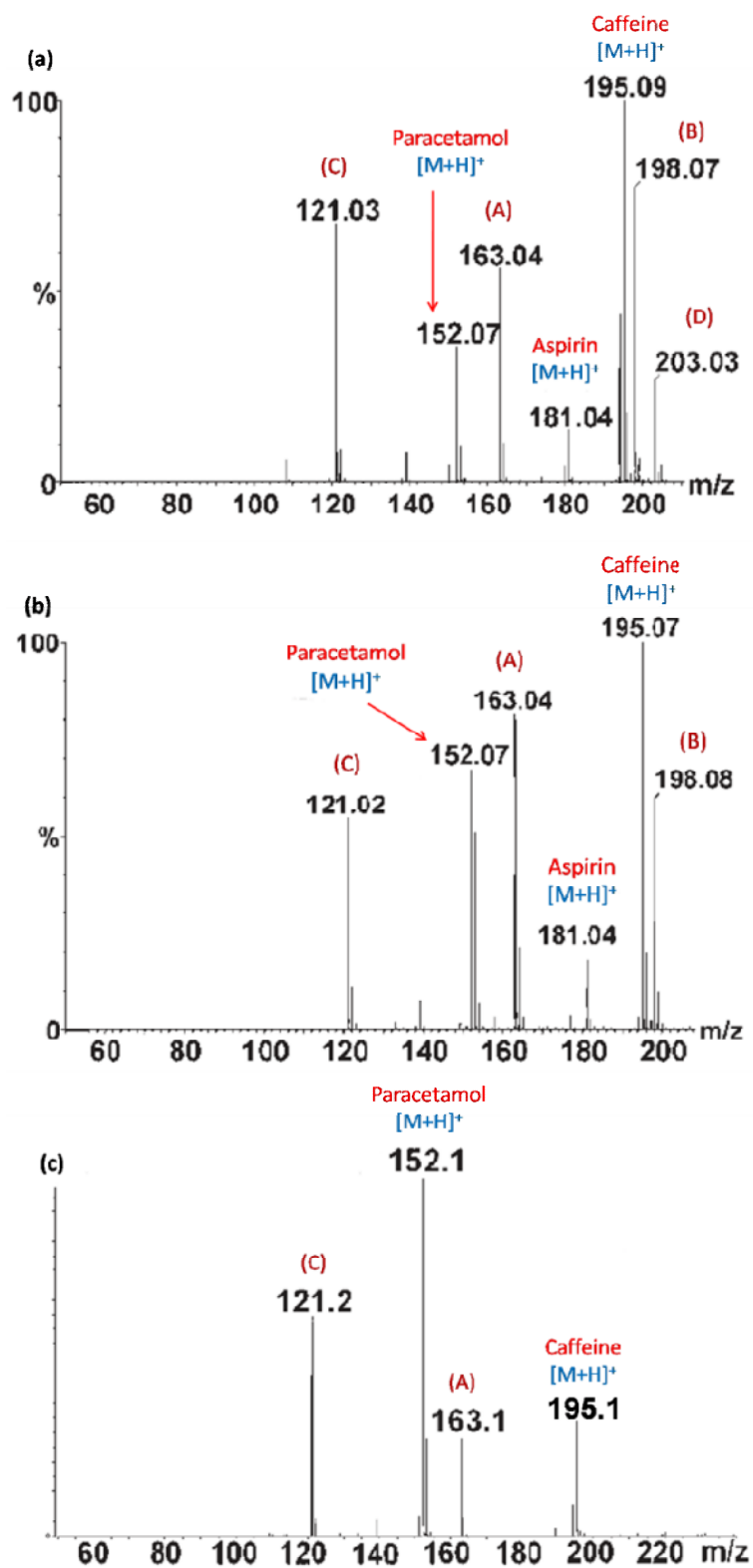


Figure 4.1 – Anadin Extra tablet by (a) DESI, (b) DAPCI and (c) DART, all in positive ion mode.

(A) Loss of H₂O from aspirin, (B) [Aspirin+NH₄]⁺, (C) potential odd-electron ion from caffeine, (D) [Aspirin+Na]⁺

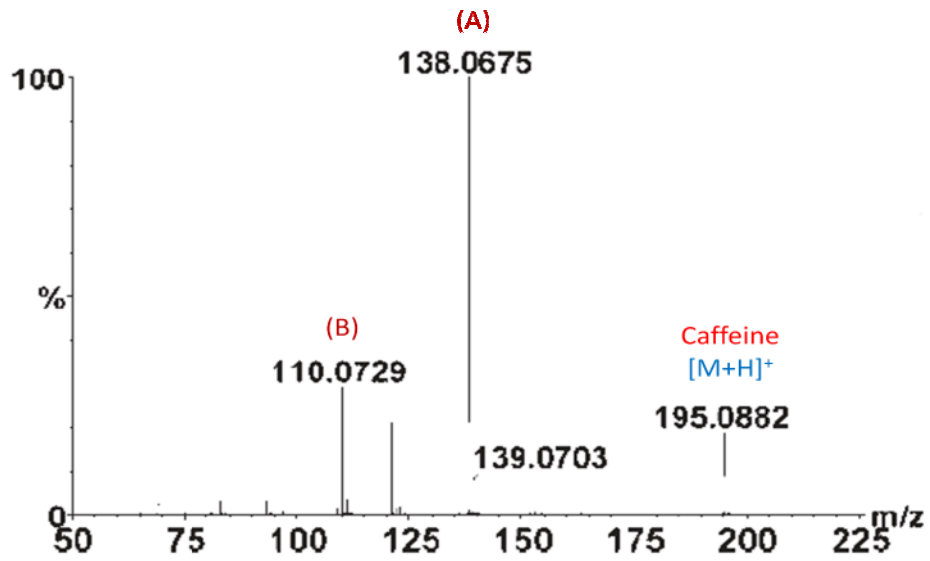


Figure 4.2 – Positive ion DAPCI MS/MS from protonated caffeine,

$[M+H]^+$ of m/z 195

(A) Loss of methyl isocyanate, (B) Loss of CO from 138

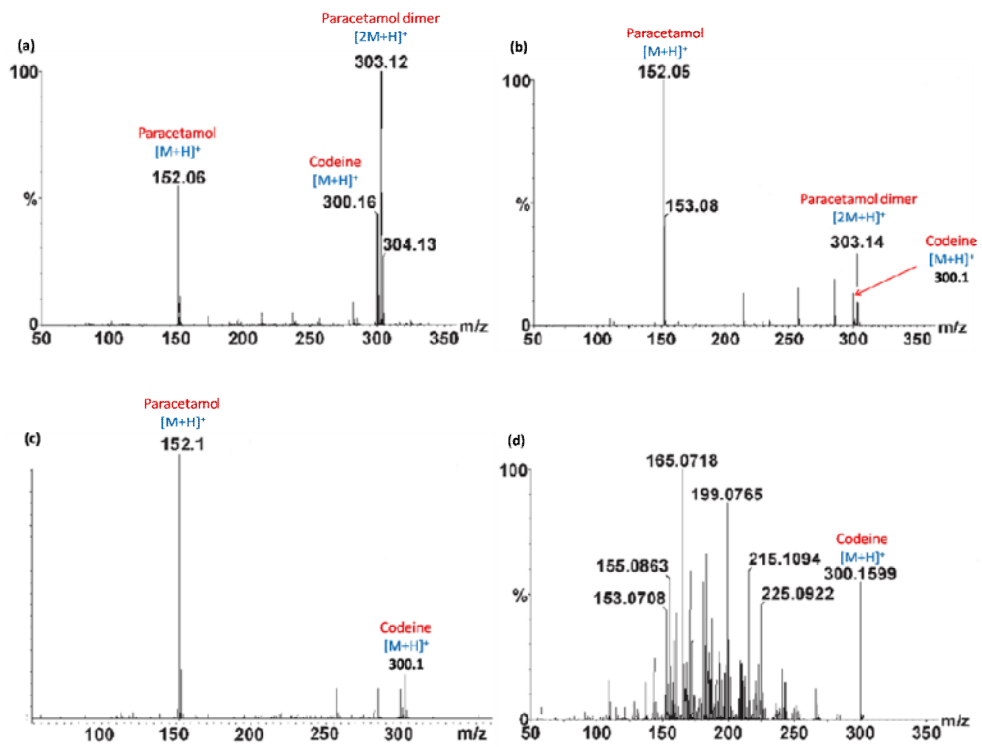


Figure 4.3 – Solpadeine Max tablet by (a) DESI, (b) DAPCI, (c) DART; (d) DAPCI MS/MS of codeine, $[M+H]^+$ m/z 300.

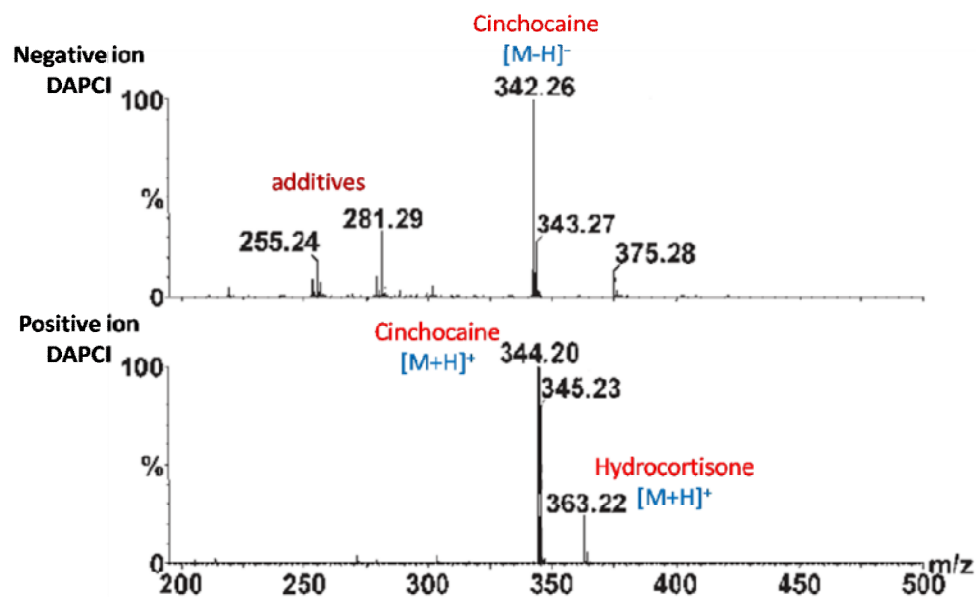


Figure 4.4 – DAPCI analyses of proctosedyl ointment. The active ingredient cinchocaine can be seen in both negative and positive modes, but hydrocortisone is only observed in positive ion mode.

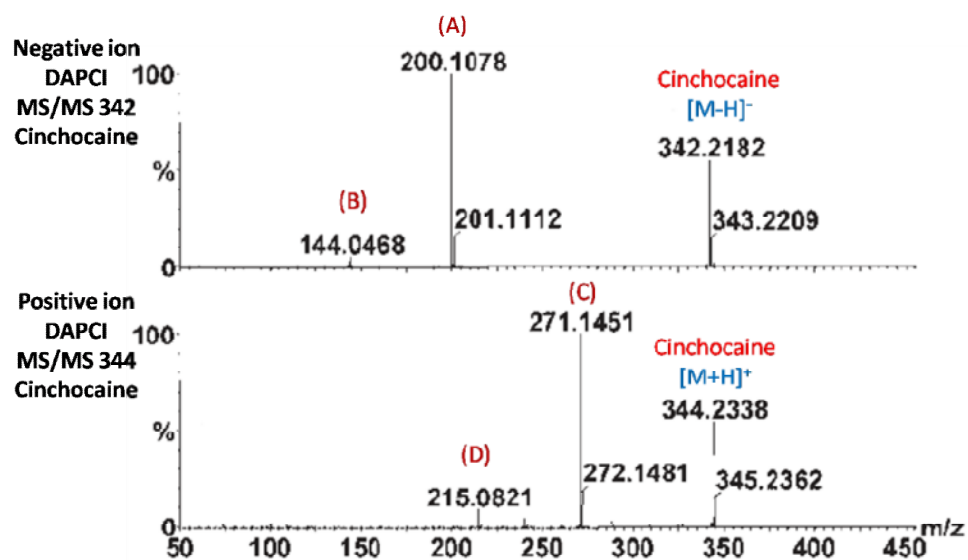


Figure 4.5 – DAPCI MS/MS of cinchocaine in negative and positive ion mode. The masses selected for fragmentation in either mode are indicated on the spectra.

(A) Loss of $C_7H_{14}N_2O$ from the deprotonated ion, (B) Loss of 56 Da from 200, (C) Loss of $C_4H_{11}N$ from the protonated ion, (D) Loss of 56 Da from 271.

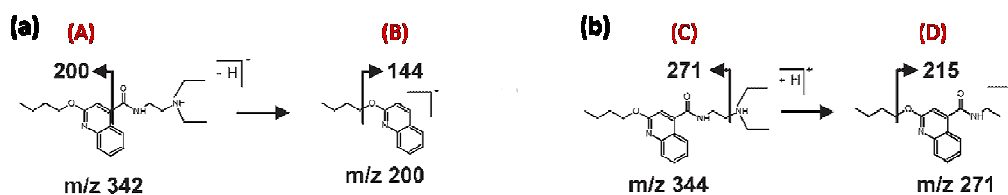


Figure 4.6 – Proposed fragmentation pathways for cinchocaine in (a) negative ion DAPCI MS/MS and (b) positive ion DAPCI MS/MS; (A) – (D) correspond to observed fragments as indicated in Figure 4.5.

The cinchocaine in proctosedyl ointment appears to ionise more efficiently than the hydrocortisone. A comparison was performed between positive ion DESI and DAPCI, the latter with and without solvent; the spectra from this analysis are shown in Figure 4.7. Protonated molecules for both active ingredients were observed by all three techniques; in all experiments the cinchocaine was shown to ionise more efficiently than the hydrocortisone. The hydrocortisone has a relative abundance of approximately 3%, 4% and 35% compared with cinchocaine in DESI, DAPCI without solvent and DAPCI with solvent, respectively. This indicates that DAPCI with solvent provides the most effective detection of both active ingredients in this ointment. When compared with DAPCI (with solvent), DART shows similar results, with a relative hydrocortisone abundance of approximately 10%. This is shown in Figure 4.8.

4.1.3 Desorption of a gel formulation from human skin

A thin layer of ibuprofen gel containing 5% w/w of the active ingredient was applied to the surface of a human finger. The gel was gently massaged until absorbed by the skin. DESI was used in negative ion mode to analyse the skin at the point of application. The deprotonated drug was detected at m/z 205, as the base peak in the spectrum as shown in Figure 4.9(a). The MS/MS spectrum of 205, acquired 20 minutes after application, is shown in Figure 4.9(b). The mass at 161 corresponds to the loss of CO_2 from the deprotonated ibuprofen ion.

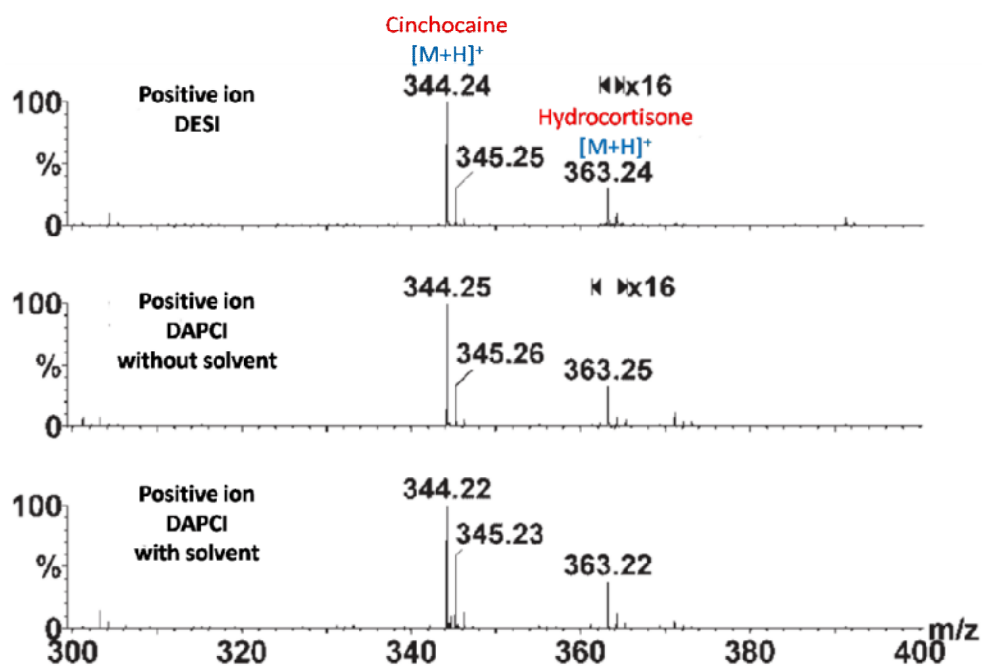


Figure 4.7 – A comparison of DESI, DAPCI without solvent and DAPCI with solvent to analyse proctosedyl ointment. The hydrocortisone shows abundances relative to cinchocaine of 3%, 4% and 35% for the three methods, respectively.

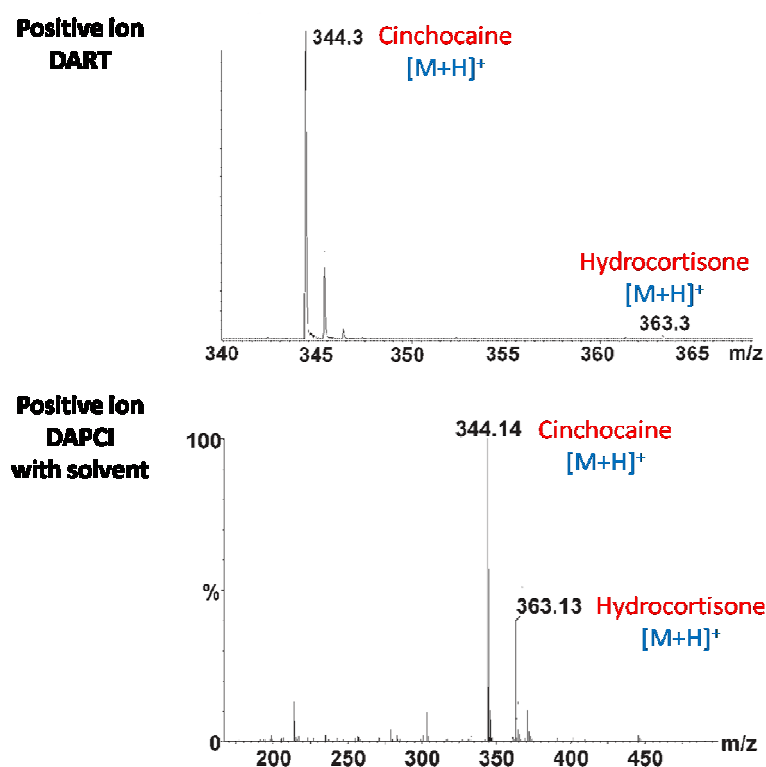


Figure 4.8 – A comparison of DART with DAPCI (with solvent) to analyse proctosedyl ointment.

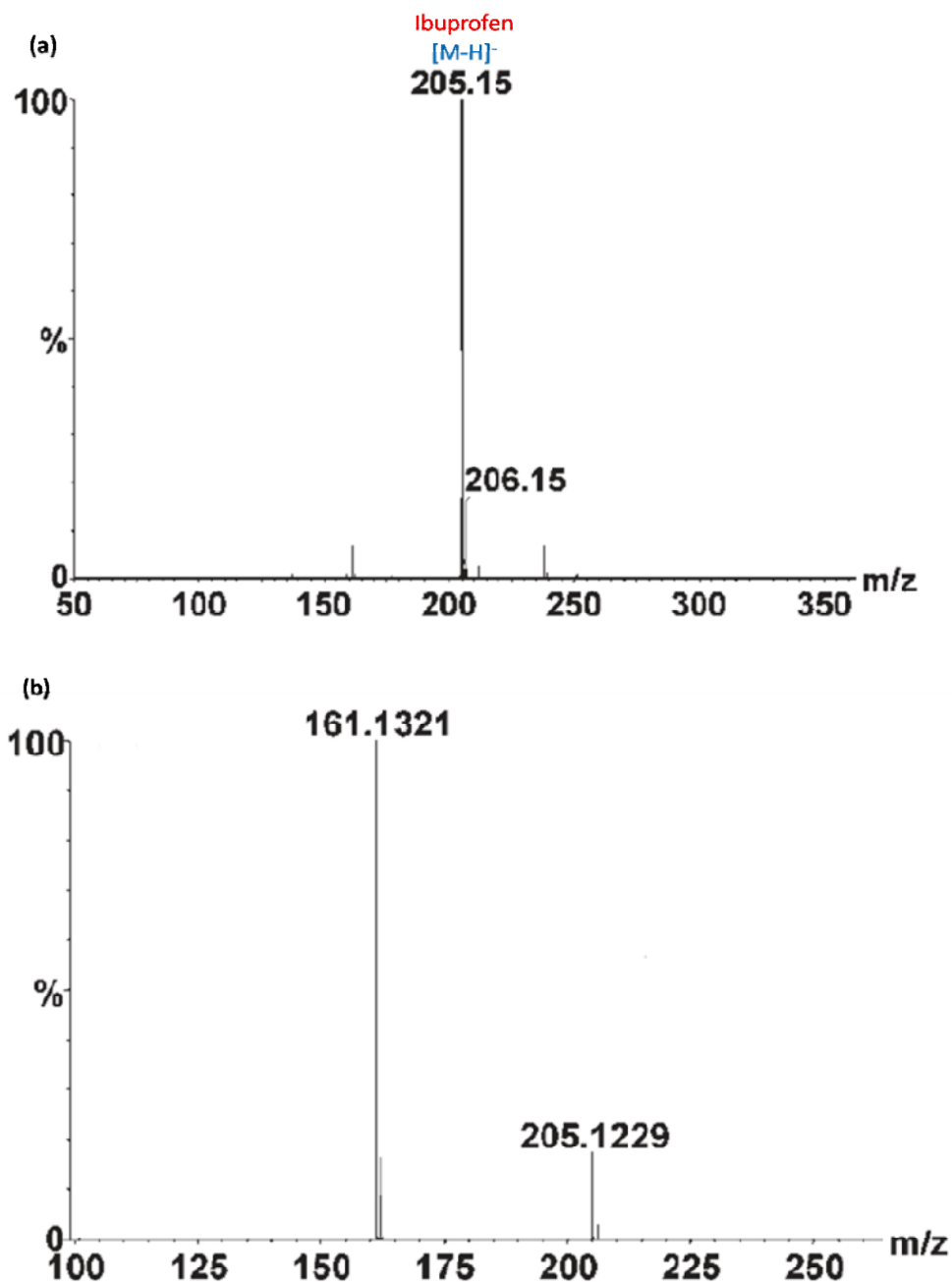


Figure 4.9 – Negative ion DESI of ibuprofen gel desorbed from human skin, (a) detection of the active ingredient by DESI, (b) MS/MS of deprotonated ibuprofen, yielding a fragment at 161 from loss of CO₂.

4.1.4 Comparison of positive and negative DAPCI-MS/MS

Metoclopramide is a pharmaceutical used to treat nausea and vomiting, and to facilitate gastric emptying in patients with gastroparesis. A metoclopramide tablet containing 10 mg of active ingredient was analysed by DAPCI in positive and

negative ion modes; the spectra are shown in Figure 4.10(a) and (b). The deprotonated metoclopramide molecule is seen at 298 and 300 with relative signal intensities of approximately 3:1, consistent with the presence of one chlorine atom in the molecule. Peaks at 283 and 255 are the result of in-source CID, the former due to loss of a methyl radical. The protonated molecule is observed at 300 as the base peak in the spectrum, with a corresponding isotope peak at 302. A very similar spectrum is observed in positive ion DART, shown in Figure 4.10(c).

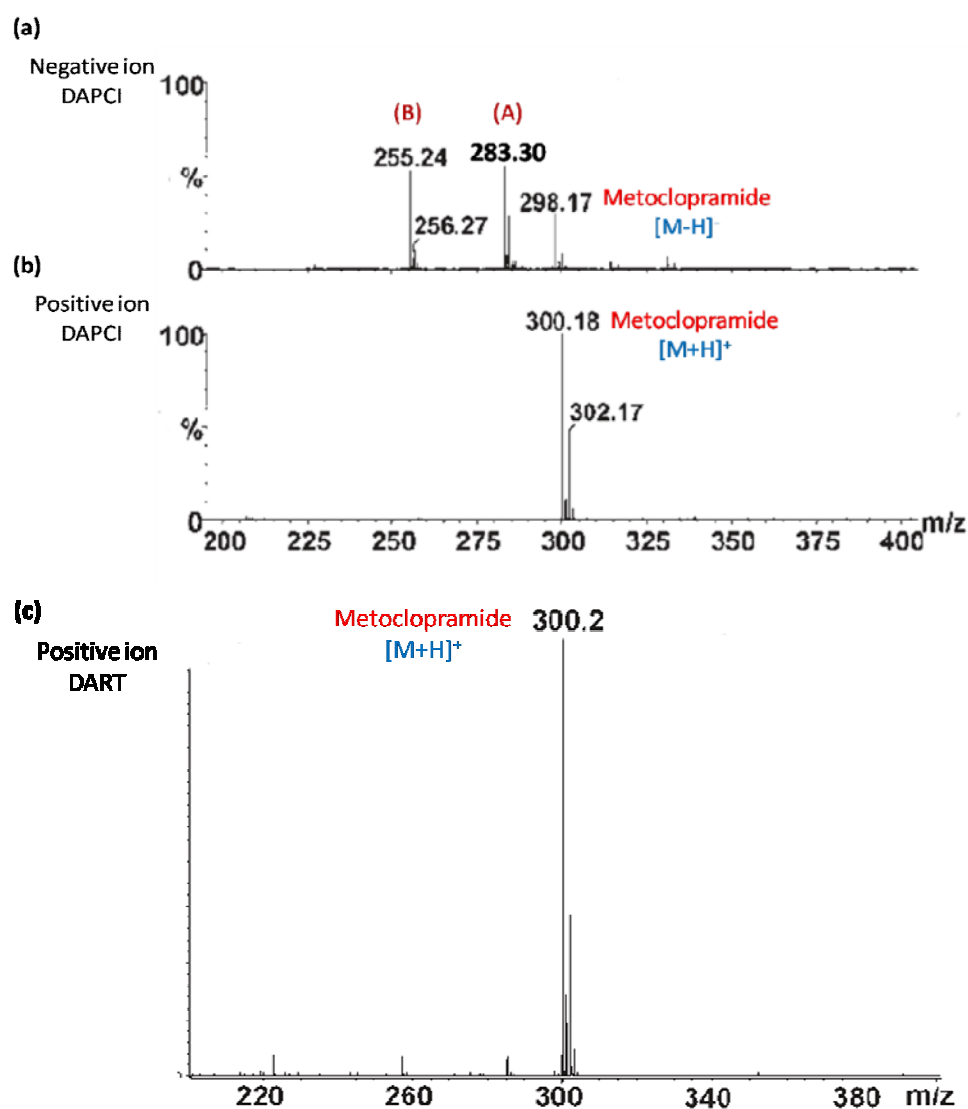


Figure 4.10 – Analysis of a metoclopramide tablet by (a) negative ion DAPCI, (b) positive ion DAPCI and (c) positive ion DART. Fragments labelled (A) and (B) in the negative ion DAPCI spectrum are produced as a result of in-source CID.

The metoclopramide ion was fragmented in both polarities, and the comparative spectra can be seen in Figure 4.11. Four fragment peaks are observed in the negative ion spectrum. As previously stated, the ion at 283 is formed from the loss of a methyl radical. The associated loss of a diethylamine moiety (72 Da) produces the ion at 211. An additional loss of CH_2 generates the fragment at 197. Fragmentation at the substituted carbon, as shown in Figure 4.12, produces the 156 ion. In the positive ion spectrum, there are ions at 227 and 184. The base peak of 227 indicates fragmentation of the protonated molecule by loss of diethylamine. Further fragmentation to 184 occurs by loss of $\text{C}_2\text{H}_5\text{N}$ (43 Da), a proposed scheme for which is also shown in Figure 4.12. All of the fragments observed show an isotope pattern indicative of chlorine, which is useful in confirming potential fragmentation pathways.

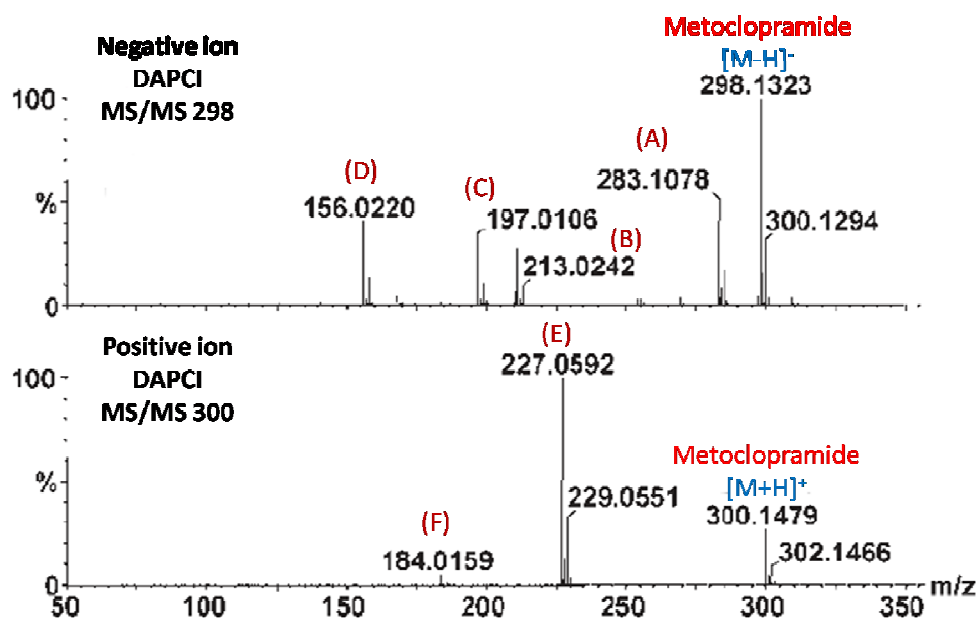


Figure 4.11 – Fragmentation of metoclopramide by negative and positive ion DAPCI.

(A) Loss of methyl radical, (B) Loss of diethylamine from 283, (C) As (B) but with additional CH_2 loss, (D) – (F) See fragmentation schemes below

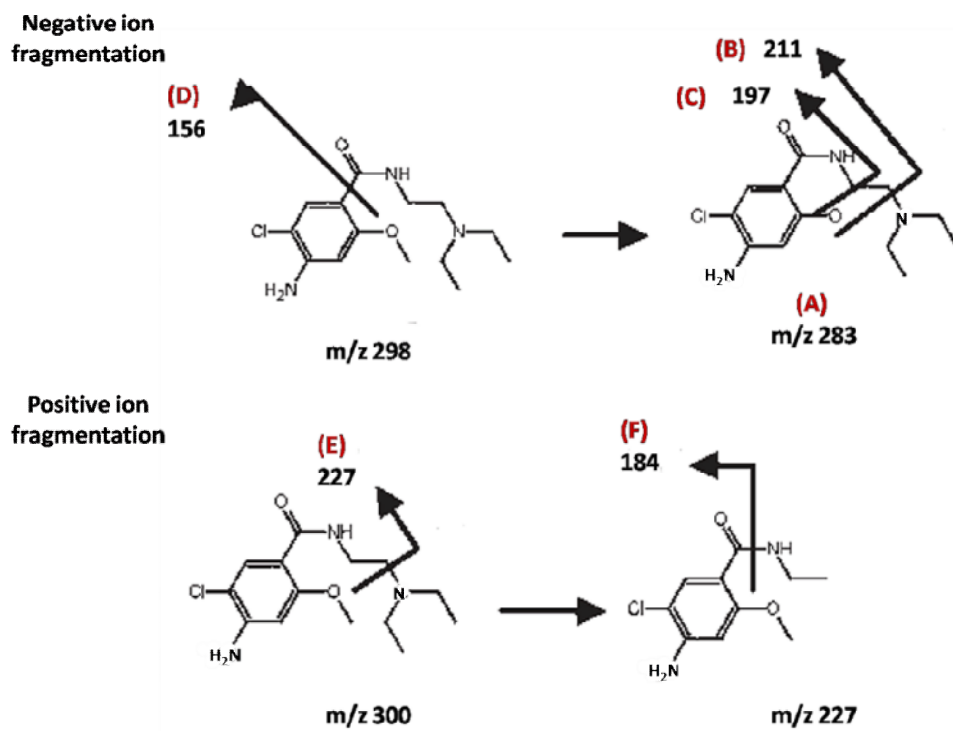


Figure 4.12 – Proposed fragmentation pathways for the deprotonated and protonated metoclopramide molecule. Letters (A) – (F) correspond to peaks indicated in spectra above.

4.1.5 Analysis of a plant alkaloid

Nicotine is a plant alkaloid, known for its natural presence in the tobacco plant. Tobacco was removed from a cigarette and analysed by positive ion DESI, DAPCI and DART. The mass spectra are shown in Figure 4.13(a)-(c). The DESI (a) and DAPCI (b) spectra were performed at low cone voltage and produced a single ion at m/z 163, corresponding to protonated nicotine. The DART data was obtained at a higher cone voltage, resulting in some of the in-source fragmentation observed. Similar fragmentation is seen when nicotine was fragmented by DAPCI MS/MS, the spectrum from which is shown in Figure 4.13(d). The ion at 132 is generated by the loss of CH_3NH_2 from the methyl-substituted pyrrolidine ring. The loss of H_2 from this generates the ion seen at 130, which contains a four-membered ring in its structure. Ions of m/z 117 and 106 could be attributed to the loss of a methyl radical and C_2H_2 , respectively, from 132. The ions represented by peaks at 84 and 80 are formed by the cleavage of the bond between the two rings in the nicotine structure.

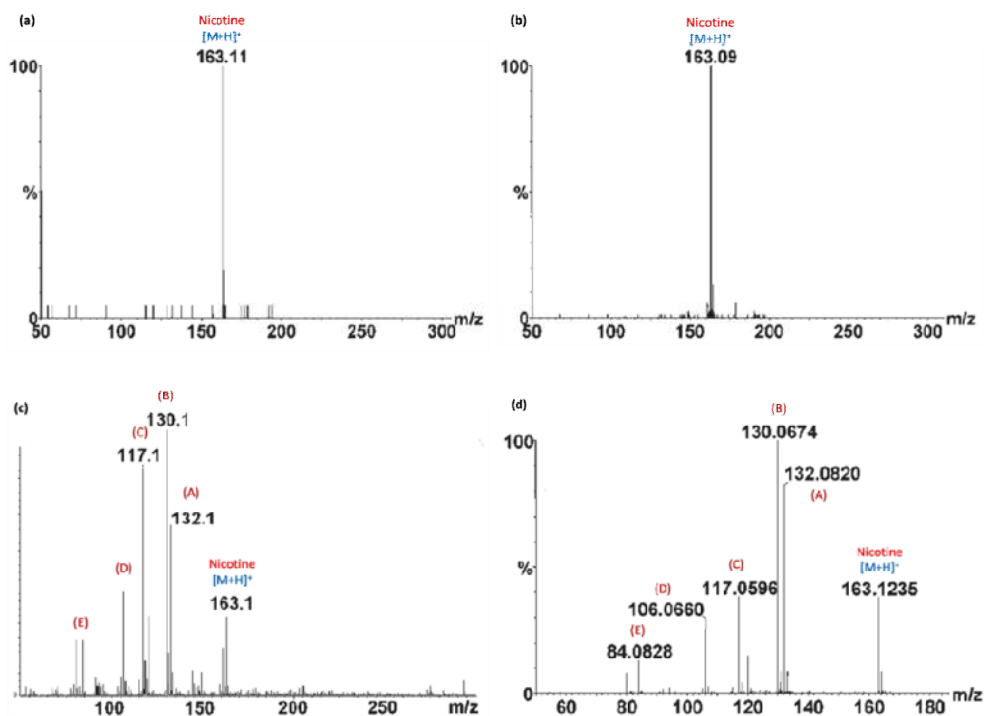


Figure 4.13 – Analysis of tobacco from a cigarette by (a) DESI, (b) DAPCI, (c) DART and (d) DAPCI MS/MS (selected mass 163, nicotine) all in positive ion mode. (A) Loss of 31 Da from the pyrrolidine ring (CH_3NH_2), (B) – (E) correspond to fragments generated in the proposed scheme shown in Figure 4.14 below.

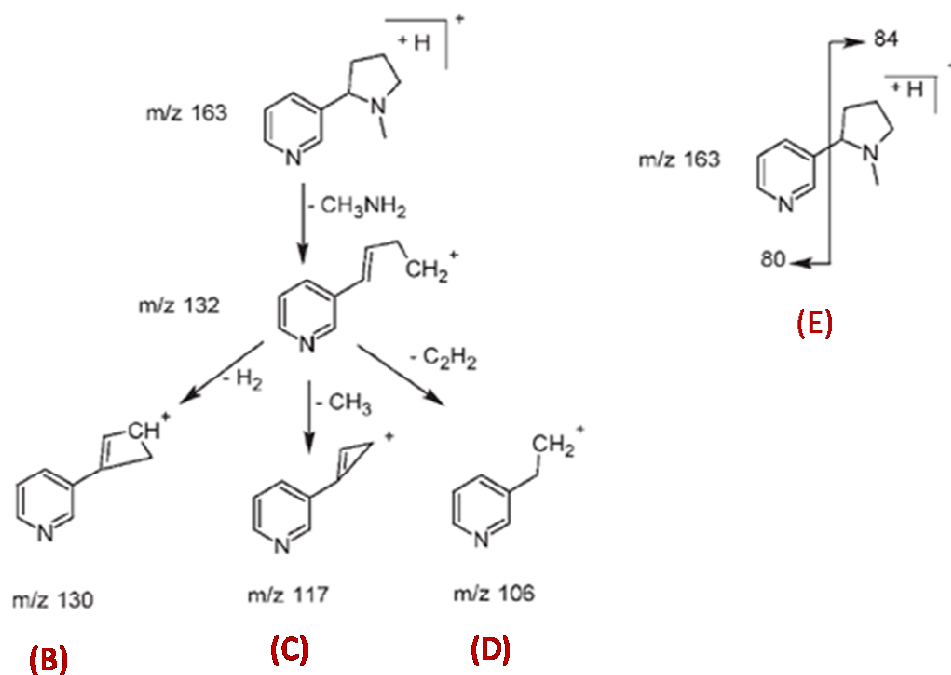


Figure 4.14 – Proposed positive ion fragmentation pathway for protonated nicotine.

4.1.6 Identification of ibuprofen metabolites from human urine

The study of drug metabolites in urine is of interest in medicine, sport regulation and forensics, but can often be challenging due to the low level of metabolites in relation to fairly high levels of endogenous materials such as salts. Samples may, therefore, have to undergo lengthy preparation in order to be analysed. The ability of DAPCI to detect drug metabolites from human urine was evaluated. Ibuprofen was used as the administered pharmaceutical, as it produces metabolites which have previously been characterised (de Oliveira *et al.*, 2005). A urine sample was obtained 75 minutes after ingestion of two ibuprofen tablets, each containing 200 mg active ingredient. Filter paper (1 cm × 8 cm) was dipped into the urine sample and allowed to absorb the liquid. Negative ion mode DAPCI was used with a solvent mixture of methanol and water.

Figure 4.15(a) shows the mass spectrum of the analysis. There are ions which can be attributed to endogenous compounds, some from components of urine (Coleman and Norton, 1986) and some ascribed to background species. Some of these have been assigned as deprotonated pyruvic acid (m/z 87), lactic acid (m/z 89), methylmalonic acid (m/z 117), xanthine (m/z 151) and hippuric acid (m/z 178). An accurate mass acquisition for a zoomed in m/z region can be seen in Figure 4.15(b). The hydroxylated and carboxylated metabolites of ibuprofen can be seen at m/z 221 and 235, respectively. No glucuronide metabolites were observed. The composition of the 221 ion was confirmed by an MS/MS experiment, where the deprotonated hydroxy-ibuprofen fragmented to give a single product ion at m/z 177 through loss of CO₂. This spectrum and the proposed fragmentation are shown in Figure 4.15(c).

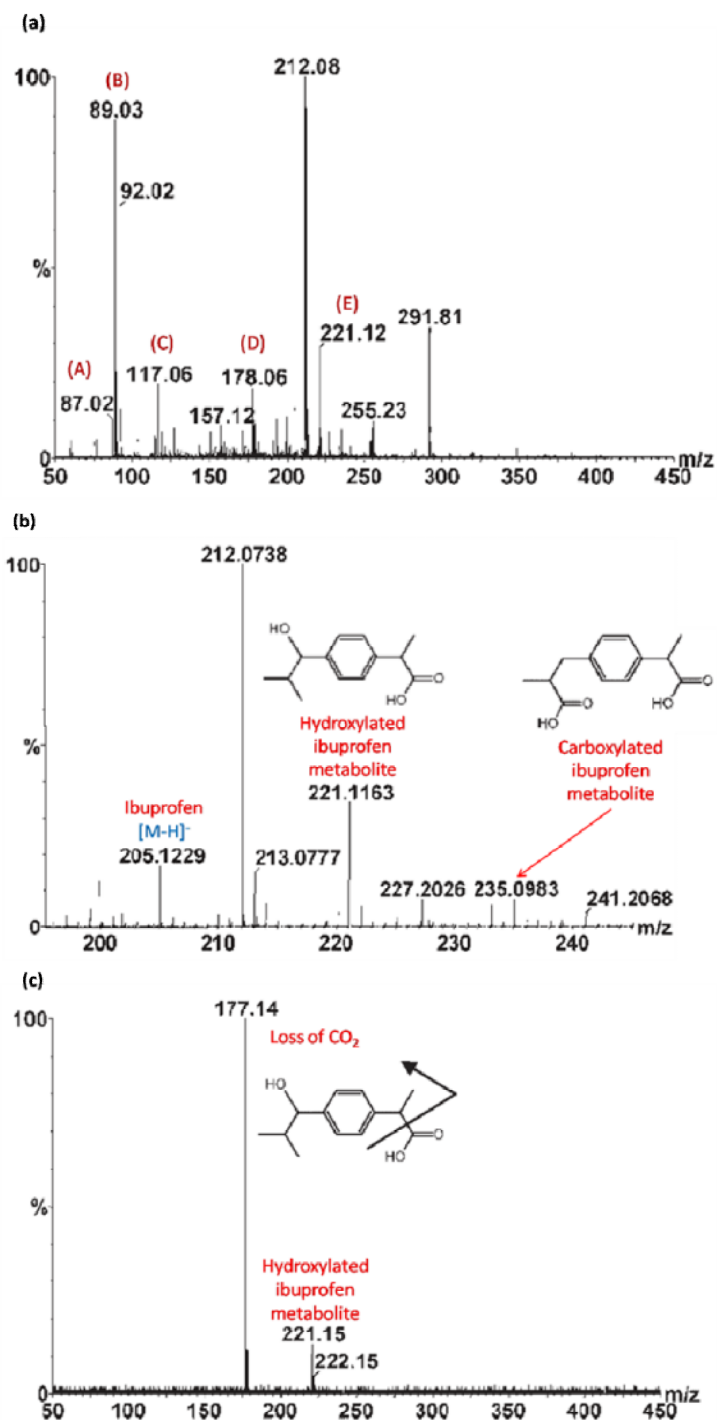


Figure 4.15 – Negative ion DAPCI analysis of a urine sample, 75 minutes after oral administration of 400 mg ibuprofen. (a) DAPCI acquisition, (b) accurate mass DAPCI acquisition, (c) DAPCI-MS/MS of m/z 221 showing loss of CO_2 .

4.2 Determining collision-induced dissociation pathways

A triple quadrupole instrument was used to elucidate fragmentation pathways for some of the product ions observed in the MS/MS spectra obtained. Triple-quadrupole mass spectrometers are extremely useful research tools, as they provide an efficient means of compound identification due to their various scanning capabilities, such as precursor, product and neutral loss scanning. As the tandem quadrupole instrument employed does not provide sufficient accurate mass measurements for unequivocal identification of the product ions studied, accurate mass MS/MS performed in a Q-TOF instrument was also used to provide complementary structural information.

4.2.1 Product ion and neutral loss scanning in negative ion DESI

Three NSAIDs were chosen as model pharmaceuticals to evaluate the use of triple-quadrupole scanning modes coupled with DESI for the detection of targeted, group-specific losses. Tablets of diclofenac, ibuprofen and naproxen were sampled separately during the same acquisition. Negative ion mode was utilised, as acidic functional groups in each of the molecules leads to efficient deprotonation. The DESI mass spectrum is shown in Figure 4.16(a). Deprotonated naproxen gives the base peak at m/z 229, deprotonated ibuprofen is observed at m/z 205, and deprotonated diclofenac at m/z 294. The peak at 250 corresponds to a loss of CO₂ from the diclofenac ion. Product ion spectra were generated for each of the drugs, which are shown in Figure 4.16(b). Under low collision energy, the ibuprofen and diclofenac produced a single product ion at m/z 161 and 250, respectively. Naproxen generated two product ions at m/z 185 and 170. The ions of 161, 250 and 185 are generated by a CO₂ loss; that at 170 is from losses of CO₂ and a methyl group. To confirm the CO₂ loss occurs from all three molecules, a neutral loss scan was performed with a pre-defined mass difference of 44 Da. As expected, the three deprotonated molecules of m/z 205, 229 and 294 were detected. This is the first demonstration of DESI coupled with a neutral loss scan for targeted losses.

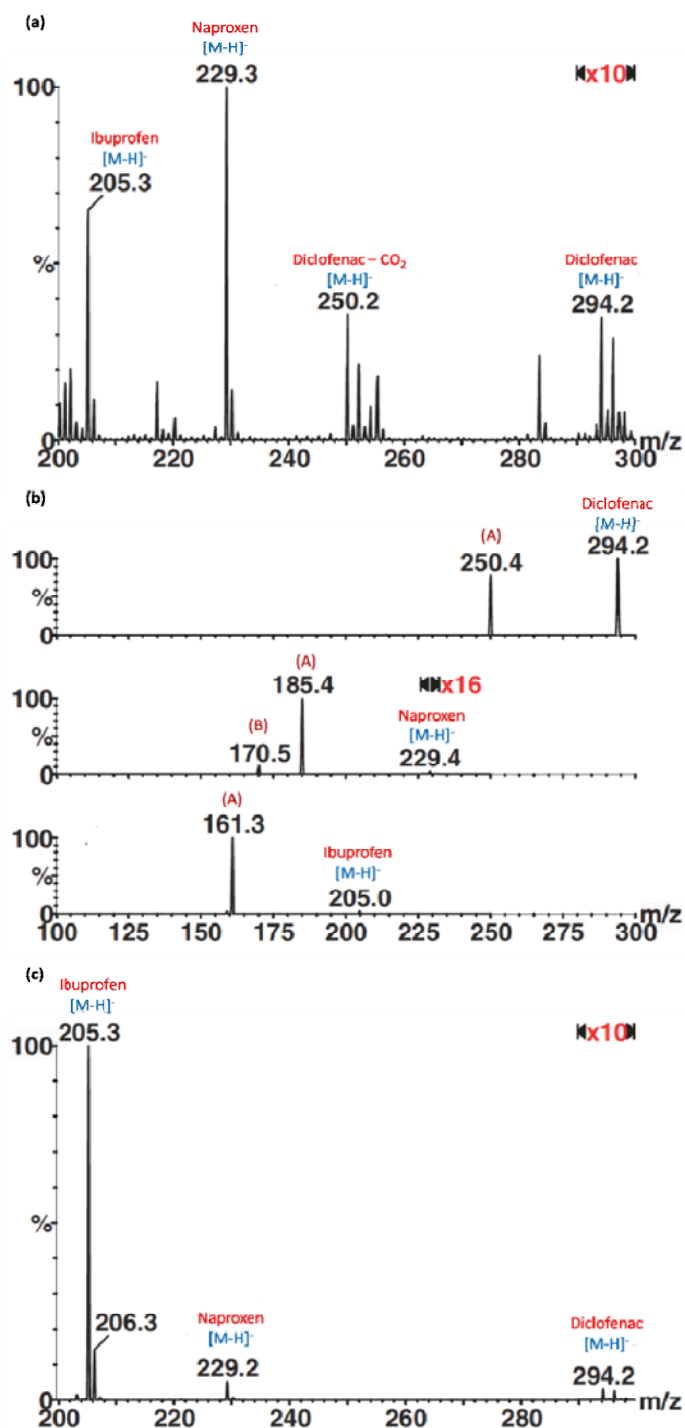


Figure 4.16 – (a) negative ion DESI-MS of ibuprofen, naproxen and diclofenac, (b) DESI MS/MS spectrum of the three deprotonated species, (c) DESI-MS/MS neutral loss spectrum, with pre-determined mass difference of 44 Da.

(A) Loss of CO₂ from deprotonated drug molecule

(B) Simultaneous loss of CO₂ and a methyl group

4.2.2 Product ion and precursor ion scanning in positive ion DESI to analyse pharmaceutical drugs

Triple quadrupole scanning and accurate mass Q-TOF MS/MS were combined to determine the fragmentation pathways of protonated aspirin, caffeine, nicotine and paracetamol. An Anadin Extra tablet (containing 45 mg caffeine, 200 mg paracetamol and 300 mg aspirin) was analysed by DESI on a triple quadrupole instrument. The mass spectrum is shown in Figure 4.17(a), where protonated paracetamol and caffeine and sodiated aspirin can be seen at m/z 152, 195 and 203, respectively. The ions at m/z 110 and 121 are in-source fragments. The 110 Da fragment was thought to originate from caffeine and from paracetamol, as previously stated. A precursor ion scan, shown in Figure 4.17(b), confirms that the ion is formed by the loss of ketene from protonated paracetamol, and by the loss of CO and methyl isocyanate from protonated caffeine. The 121 fragment in the Anadin Extra analysis, resulting from in-source CID, was thought to originate from aspirin, confirmed by the precursor ion scan shown in Figure 4.17(c), which indicates it comes from m/z 139 and 163. These are formed by the loss of ketene and the loss of water from aspirin. The scan also indicates m/z 149 as a precursor of 121. An accurate mass measurement of 149.0243 suggests a formula of $C_8H_5O_3$, which arises from the dissociation of an unknown phthalate, a common contaminant in ESI mass spectra. The proposed fragmentation of protonated aspirin and the phthalate is shown in Figure 4.18.

DESI accurate mass of protonated caffeine generated product ions of m/z 138, 123 and 110, shown in Figure 4.19(a). The 138 Da fragment is formed through loss of methyl isocyanate, which then loses CO to generate the 110 fragment. The accurate mass measurement indicated an elemental formula of $C_5H_5N_3O$ for m/z 123, suggesting the unusual formation of an odd-electron product ion. A precursor ion scan of m/z 123 is shown in Figure 4.19(b), indicating the ion is formed from the 138 fragment. This is proposed to occur through the loss of a methyl radical by cleavage of the N-C bond. The fragmentation pathway resulting in the formation of the 110, 123 and 138 product ions is shown in Figure 4.20.

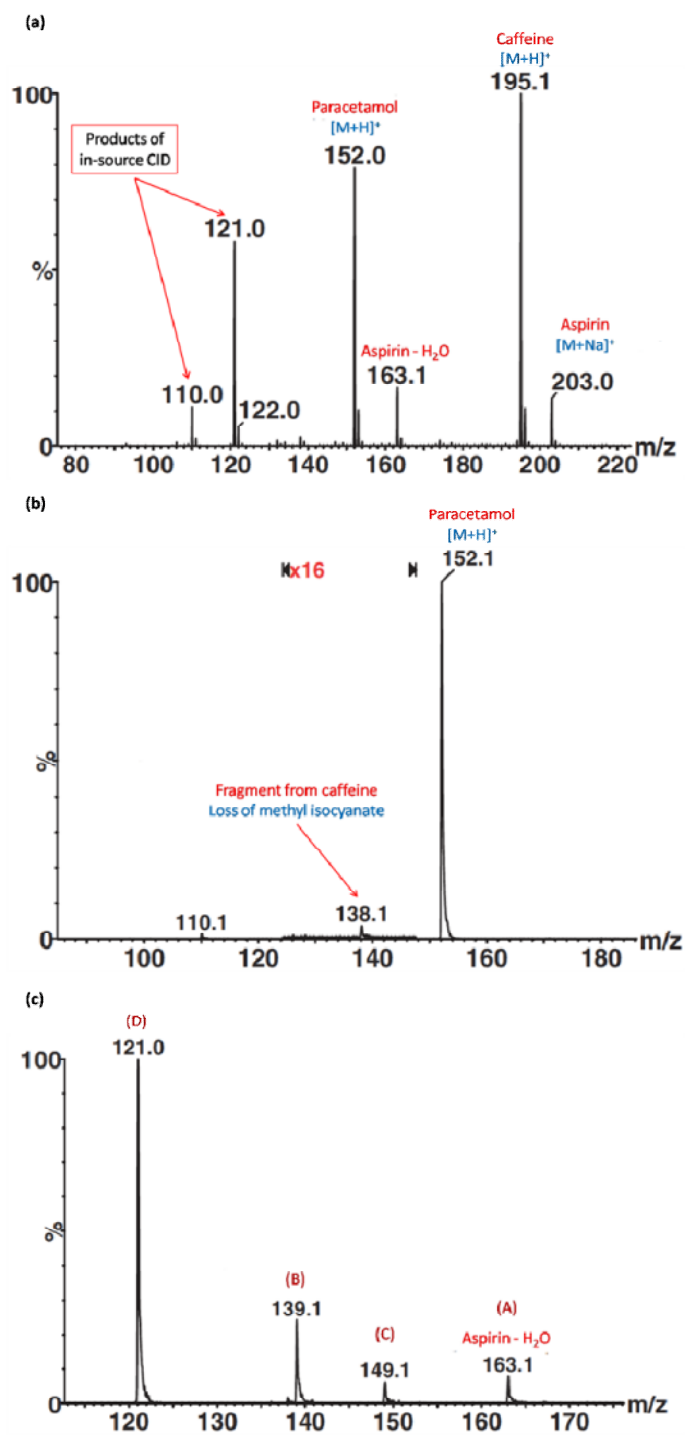


Figure 4.17 – (a) positive ion DESI-MS of Anadin Extra, (b) precursor ion spectrum of fragment ion 110, (c) precursor ion scan of fragment ion 121.

(A) Loss of water from aspirin, (B) Loss of ketene from aspirin, (C) Fragment from dissociation of phthalate contaminant, (D) selected ion; these assignments correspond to assignments in the fragmentation pathway shown in Figure 4.18.

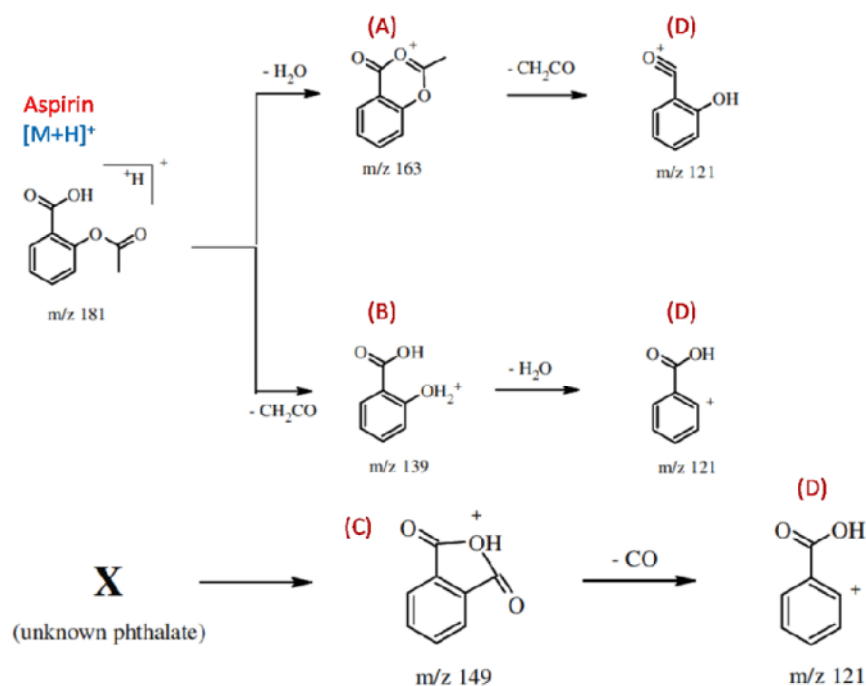


Figure 4.18 – Proposed fragmentation pathways of protonated aspirin by means of DESI-MS/MS

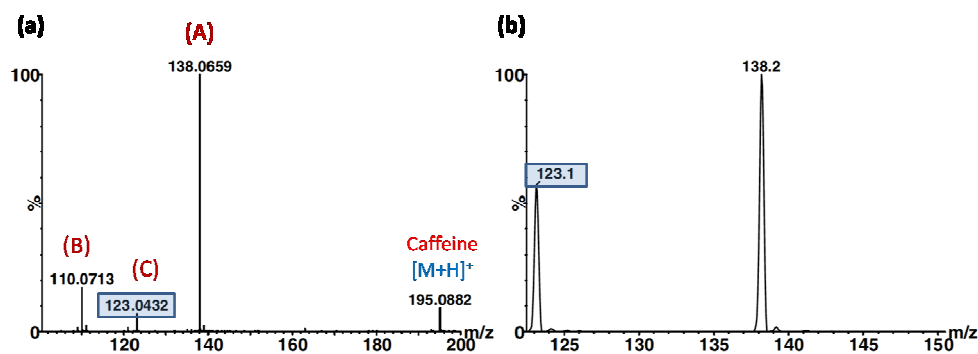


Figure 4.19 – (a) positive ion DESI-MS/MS of protonated caffeine, m/z 195, (b) precursor ion scan of the 123 fragment, highlighted.

(A) Loss of methyl isocyanate, (B) Loss of CO from 138, (C) precursor ion scan shows that this fragment is generated from the 138 ion, which is proposed to occur via loss of a methyl radical. The assignments correspond to assignments in the fragmentation pathway shown in Figure 4.20.

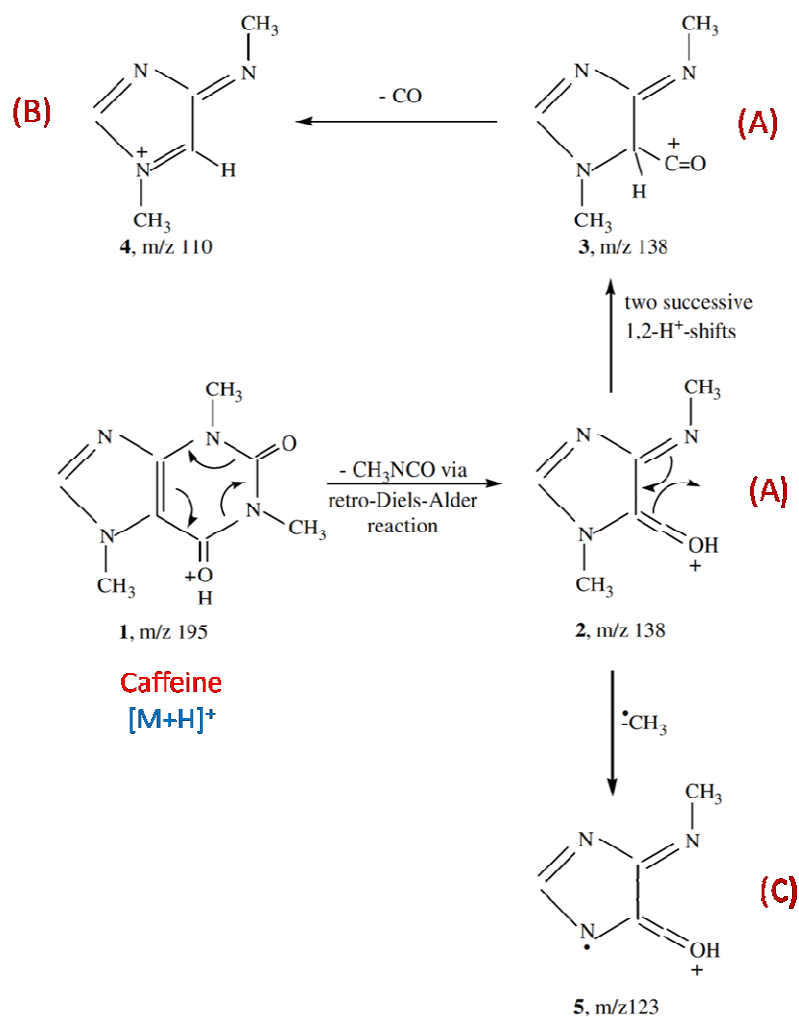


Figure 4.20 - Proposed fragmentation pathways of protonated caffeine by means of DESI-MS/MS, (A)-(C) correspond to annotated peaks in previously shown spectra.

4.2.3 Product ion and precursor ion scanning in positive ion DESI to analyse nicotine from cigarette tobacco

Accurate mass measurements and triple quadrupole scanning were utilised to determine more detailed fragmentation pathways for nicotine as sampled from cigarette tobacco. The DESI-MS/MS spectrum of protonated nicotine is shown in Figure 4.21, where a number of fragment peaks can be observed. Fragmentation pathways which may lead to these ions were proposed, as shown in Figure 4.14. The product ion of m/z 117 is an unusual odd-electron ion, and a precursor scan was performed to more categorically determine its origins. The spectrum from this scan,

shown in Figure 4.22(a), confirms that the 117 fragment is formed from the loss of a methyl radical from the 132 nicotine fragment. A precursor scan of the 106 fragment was also performed, the spectrum from which can be seen in Figure 4.22(b). This indicates that the ion of m/z 106 is formed solely by the loss of C_3H_7N from the protonated nicotine molecule, contrary to the previous assignment. The revised fragmentation pathway for nicotine is summarised in Figure 4.23.

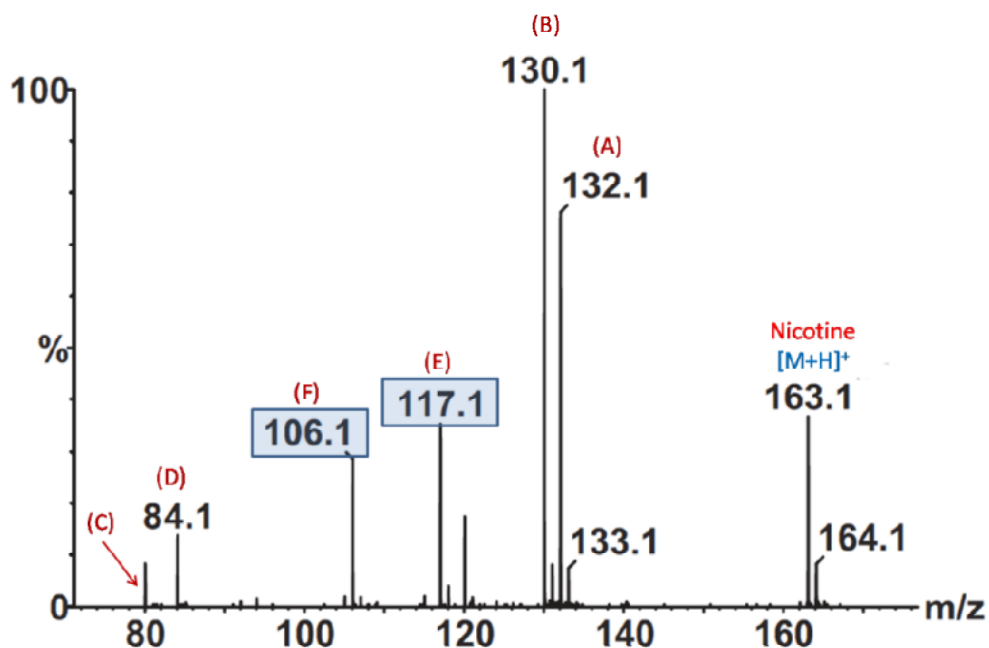


Figure 4.21 – Positive ion DESI-MS/MS spectrum of protonated nicotine from cigarette tobacco. (A)-(E) correspond to assignments as shown in Figure 4.23. Ions of m/z 106 and 117 (highlighted) were selected for precursor ion scans.

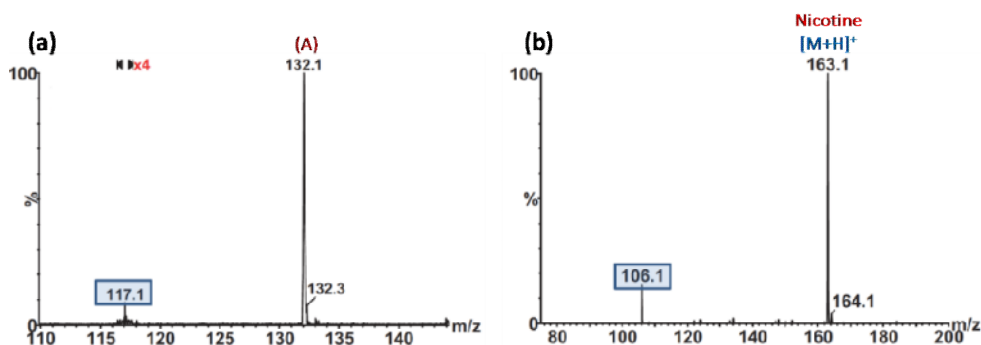


Figure 4.22 – (a) precursor ion spectrum of m/z 117 from MS/MS spectrum of nicotine, (A) corresponds to the fragmentation assignment shown in Figure 4.23, (b) precursor ion spectrum of m/z 106, which shows that this ion is generated directly from the protonated nicotine molecule.

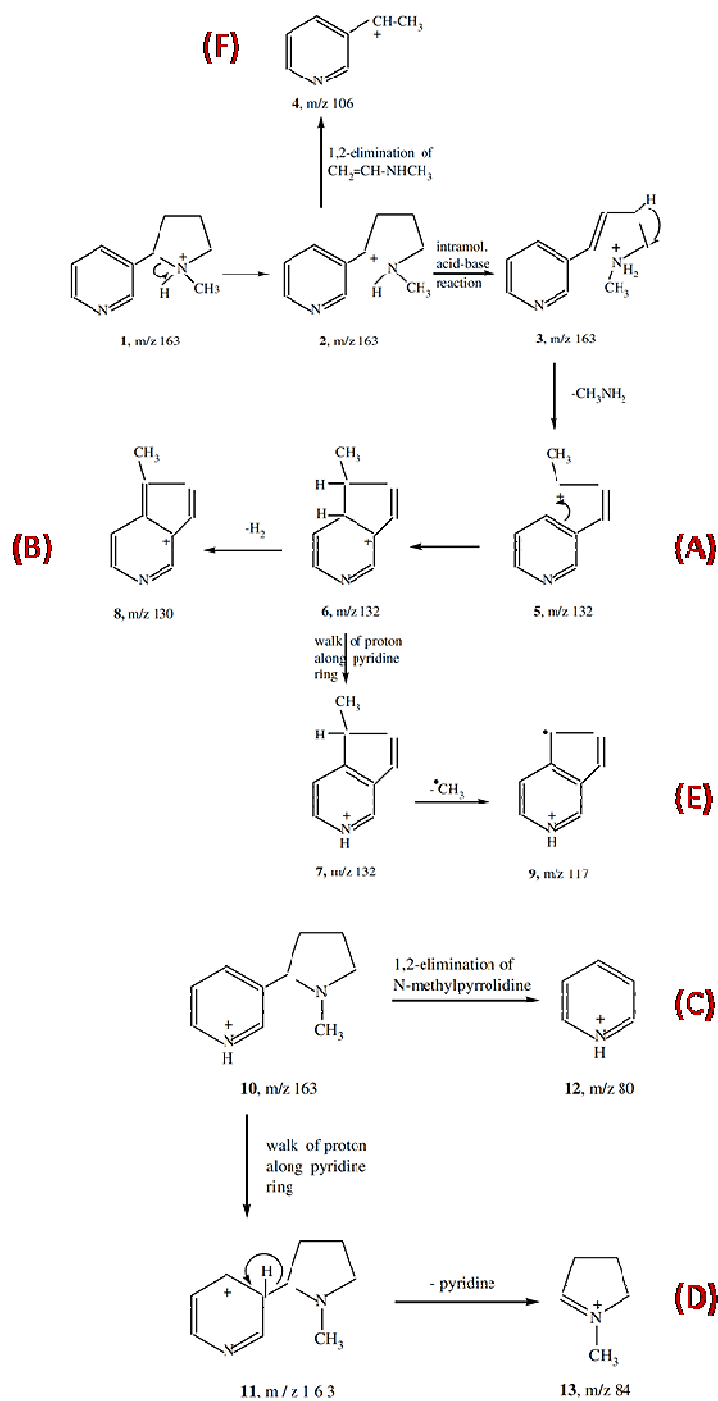


Figure 4.23 - Proposed fragmentation pathways of protonated nicotine by means of DESI-MS/MS, (A)-(F) correspond to annotated peaks in previously shown spectra.

4.3 Polarity switching accurate mass measurement of pharmaceutical samples using DESI

For the rapid analysis of pharmaceutical formulations, as demonstrated so far, it would be useful to extend the sampling approach to simultaneously detect both basic and acidic active ingredients that easily protonate and deprotonate in the same acquisition. Many medications contain more than one active ingredient, with different preferential ionisation. This would negate the need for repeat analysis, which can be both time-consuming and costly, and would also be of benefit for compound screening where the ionisation mode for the active ingredient is unknown or uncertain. DESI was coupled to a TOF instrument fitted with a two-way electrospray ionisation source (Waters, UK). Accurate mass measurements are more simply acquired, as a reference compound is infused into a separate inlet rather than sampling before or after the drug under investigation. Sampling pharmaceutical formulations can generate very high ion counts, as the amount of active ingredient can range from 1 to 1000 mg. When using an instrument with a time-to-digital (TDC) detector, as is the case here, mass accuracy can be compromised at high ion counts due to dead-time saturation. During this study, we have successfully used a single TOF instrument incorporating dynamic range extension (DRE) technology together with polarity switching, for routine accurate mass measurement of DESI generated ions. The DRE technique employs modified transfer optics capable of reducing the transmission of the ion beam into the orthogonal sampling (pushout) region by means of an applied voltage. These optics are arranged to defocus the ion beam normal to the plane of the TOF-MS, resulting in minimal effect on resolution and mass measurement.

4.3.1 Preferential ionisation

During any single acquisition, data from each of the two sprays are sampled independently and stored as separate functions within the same data file. This means that four functions are acquired: positive ion reference, negative ion reference, positive ion analyte and negative ion analyte. Figure 4.24 shows the reference compound leucine enkephalin (a) and a solid tablet containing 500 mg of the macrolide antibiotic, erythromycin (b).

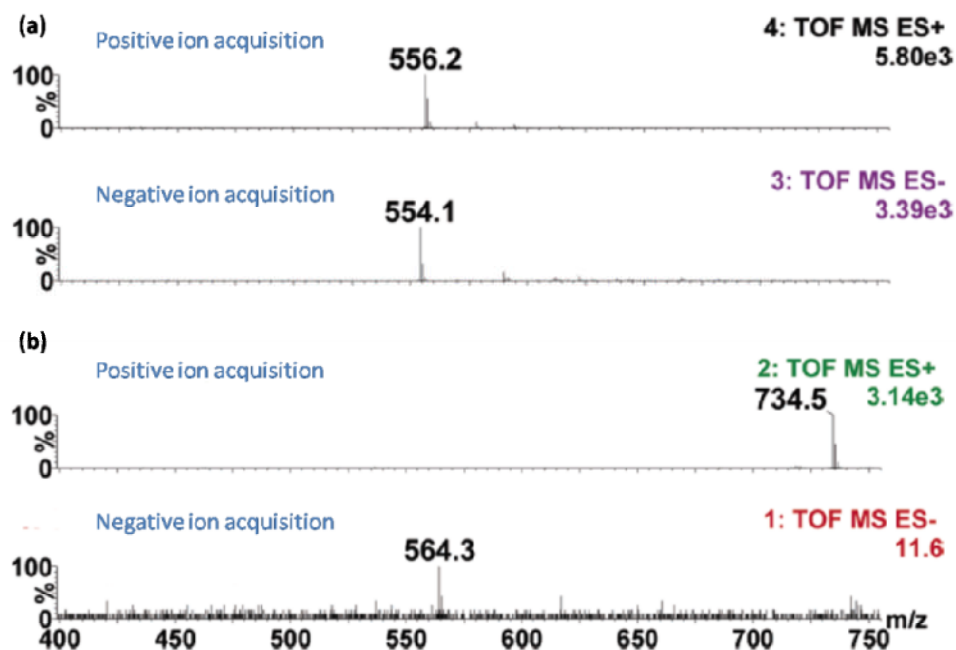


Figure 4.24 – Mass spectra for DESI-MS of (a) leucine enkephalin, where ions are seen in both polarities, and (b) erythromycin, where significant ions are only seen in positive ion mode. The peak observed at 564.3 can only be interpreted as baseline noise.

Leucine enkephaline is observed in both ion modes, whilst erythromycin is only seen in positive ion mode; this is expected as erythromycin is known to preferentially protonate rather than deprotonate. Although a high ion count was seen, no carry-over of sample is seen. Accurate mass measurements were performed on fragments of erythromycin generated by in-source CID, which was possible despite the high ion counts due to the DRE technology employed. Five prominent product ions are observed in the spectrum, shown in Figure 4.25, and can be assigned using the accurate mass data. The fragment at m/z 716 results from the loss of water from the deprotonated drug molecule, m/z 576 is the loss of the cladinose sugar moiety, m/z 558 is either loss of water and cladinose, or the loss of water from m/z 576, m/z 540 is either the loss of $(\text{H}_2\text{O})_2$ and the loss of cladinose or the loss of H_2O from m/z 558, and m/z 158 corresponds to the desosamine sugar moiety.

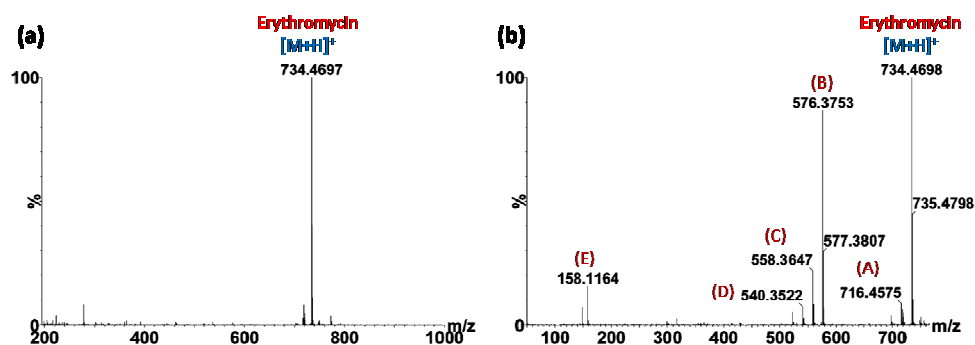


Figure 4.25 – (a) accurate mass spectrum obtained for protonated erythromycin, (b) accurate mass in-source CID mass spectrum of protonated erythromycin.

4.3.2 Analysis of multiple active ingredients from a single tablet

An Anadin Extra tablet, containing aspirin, caffeine and paracetamol, was analysed to attempt the simultaneous detection of all three active ingredients using polarity switching in the same acquisition. Figure 4.26(a) shows results from the positive ion mode analysis. Two peaks dominate the spectrum, protonated paracetamol at m/z 152 and protonated caffeine at m/z 195. No protonated aspirin was observed. In the negative ion mode analysis, the spectrum from which is shown in Figure 4.26(b), three ions were detected. Deprotonated aspirin and deprotonated paracetamol are seen at m/z 179 and 150, respectively. The peak at 137 corresponds to the loss of ketene from aspirin. This indicates the power of a polarity switching experiment, as all three drugs were detected in the same acquisition, thereby negating the need for analysis to be repeated in each ion mode.

4.3.3 Analysis of a pharmaceutical ointment

The eye ointment chloramphenicol, containing 1 % w/v of active ingredient, was desorbed from a piece of cardboard. The drug was only detected in negative ion mode. The obtained mass spectrum shows excellent agreement to the theoretical masses which would be expected, as can be seen in Figure 4.27(a). Accurate mass measurements generated the spectrum shown in Figure 4.27(b). Three peaks are seen, separated by two m/z units, with relative intensities of 9:6:1, which is consistent with the presence of two chlorine atoms in the molecule.

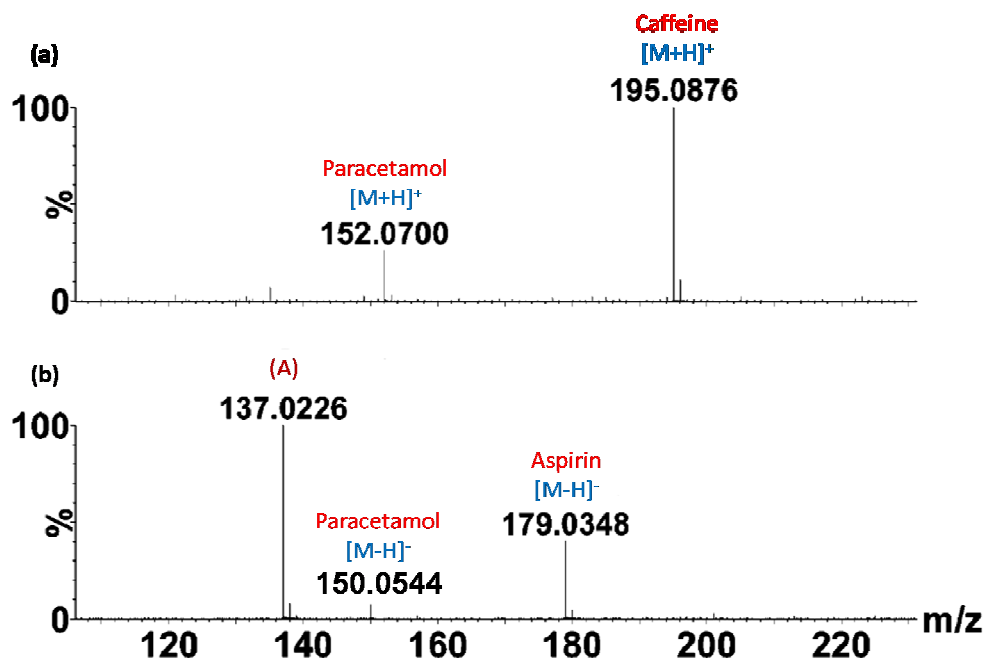


Figure 4.26 – Anadin Extra tablet data in (a) positive ion mode, and (b) negative ion mode. (A) fragment generated by the loss of ketene from deprotonated aspirin.

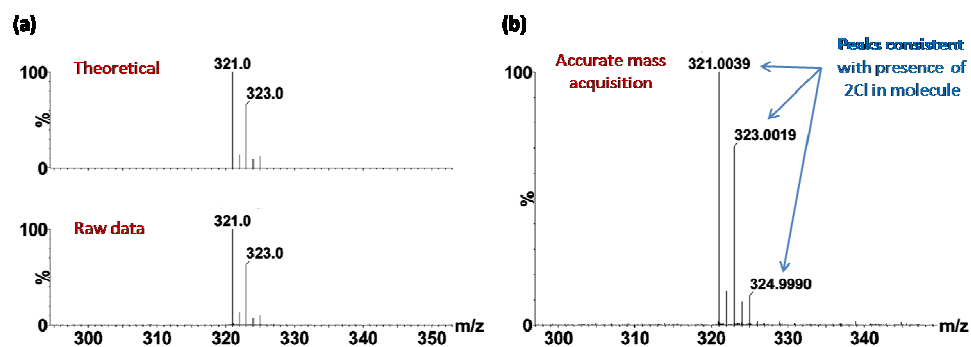


Figure 4.27 – (a) theoretical isotope distribution for deprotonated compared with data acquired, showing excellent agreement, (b) accurate mass spectrum for deprotonated chloramphenicol showing isotope distribution consistent with the presence of two chlorine atoms.

4.4 Conclusions

DESI, DAPCI and DART techniques have provided a highly robust means of interrogating the active ingredients of a variety of pharmaceutical formulations. Sampling the formulation is rapid and ionisation occurs almost instantly. The techniques DART, DESI and DAPCI (solvent and solventless) have complex, potentially inter-related mechanisms but have been shown to provide complementary information on a range of compounds of interest. Whilst all the methods gave spectra, some are better than others depending on the type of sample under study; DESI showed more efficient detection for those molecules which were polar or of higher molecular weight, whereas DAPCI appeared to be more efficient for analysis of non-polar compounds. The direct analysis of urine and molecules adsorbed on skin is of particular interest, due to their medical and forensic application potential.

The use of collisionally induced dissociation, coupled with capabilities such as neutral loss scanning, enhanced the information content available from a DESI experiment, providing added insight into fragmentation pathways of drugs of interest. This shows potential as a rapid and powerful analytical tool for unequivocal determination of the identity of observed species, and their subsequent fragmentation. Data acquired when employing DESI with a novel polarity switching instrument show that this approach has many advantages in the rapid screening of pharmaceuticals. This could be of benefit in industrial quality control, forensics and the identification of counterfeit medication, with many advantages in terms of its high-throughput capabilities.

Ambient ionisation has moved forward vastly since its inception in 2004, now being coupled to emerging ion mobility technology (Kaur-Atwal *et al.*, 2007; Weston *et al.*, 2005; Williams and Scrivens, 2008) and used to image drugs and metabolites in tissues (Wiseman *et al.*, 2008). This area of research appears to have firmly established roots in the field of MS and will likely continue its rapid evolution for the foreseeable future.

References

Northern Illinois University www.niu.edu/AnalyticalLab/ms/troubleshooting

- Coleman, M. D., Norton, R. S.** (1986). Observation of drug metabolites and endogenous compounds in human urine by ^1H nuclear magnetic resonance spectroscopy. *Xenobiotica: The fate and safety evaluation of foreign compounds in biological systems* **16**: 69 - 77.
- de Oliveira, A. R. M., de Santana, F. J. M., Bonato, P. S.** (2005). Stereoselective determination of the major ibuprofen metabolites in human urine by off-line coupling solid-phase microextraction and high-performance liquid chromatography. *Anal. Chim. Acta* **538**: 25-34.
- Kaur-Atwal, G., Weston, D. J., Green, P. S., Crosland, S., Bonner, P. L. R., Creaser, C. S.** (2007). Analysis of tryptic peptides using desorption electrospray ionisation combined with ion mobility spectrometry/mass spectrometry. *Rapid Commun. Mass Spectrom.* **21**: 1131-1138.
- Weston, D. J., Bateman, R., Wilson, I. D., Wood, T. R., Creaser, C. S.** (2005). Direct Analysis of Pharmaceutical Drug Formulations Using Ion Mobility Spectrometry/Quadrupole-Time-of-Flight Mass Spectrometry Combined with Desorption Electrospray Ionization. *Anal. Chem.* **77**: 7572-7580.
- Williams, J. P., Scrivens, J. H.** (2008). Coupling desorption electrospray ionisation and neutral desorption/extractive electrospray ionisation with a travelling-wave based ion mobility mass spectrometer for the analysis of drugs. *Rapid Commun. Mass Spectrom.* **22**: 187-196.
- Wiseman, J. M., Ifa, D. R., Zhu, Y., Kissinger, C. B., Manicke, N. E., Kissinger, P. T., Cooks, R. G.** (2008). Desorption electrospray ionization mass spectrometry: Imaging drugs and metabolites in tissues. *Proc. Natl. Acad. Sci. USA* **105**: 18120-18125.

Chaper 5

Introduction to metallomics

5.1 Metals and proteins

It has been estimated that over 30% of proteins either incorporate metal atoms, or require them in order to function properly. Metal ions are required by biological systems for a variety of fundamental processes: signalling, gene expression and catalysis to name but a few. In order to fully characterise the biochemistry of a cell, therefore, one needs to look not only at the genome and proteome, but also at the **metallome**: the entirety of individual metal species in the cell (Szpunar, 2004). This encompasses free metal ions, and those in complex with biomolecules; of particular interest here are the metal species associated with proteins, i.e. the metalloproteome. A representation of the involvement of metals in various aspects of biology can be seen in Figure 5.1. The term ‘metallomics’ was proposed by Haraguchi in 2003 as a new field integrating research related to biometals, which should be considered at the same level of significance as genomics and proteomics due to the hugely important role played by metals in biological processes (Haraguchi, 2004). Major groups of proteins of interest in the metallomics field include metalloenzymes, metal-transport proteins and metal stress proteins (Szpunar, 2005).

A complete fingerprint of the metal-binding components in a cell is likely to provide new insight into the role of metal ions in biochemistry. The current analysis methods available are not without their limitations, one in particular being the ability to scan samples for multiple trace elements. Whilst much work is being carried out in this burgeoning field, it is very much in the stages of technique evaluation and there is much potential to increase information gained on the metalloprotein content of a cell. If the metalloproteome refers to the metal-site structures in all proteins of the proteome of a given organism (or given tissue) under a given set of conditions at a given stage of development, then metalloproteins and their bound metals could be considered as biological markers for physiological differences (Lobinski *et al.*, 2006). The ability to perform high-throughput screens of metal content and protein-metal binding could provide significant insight into cellular processes.

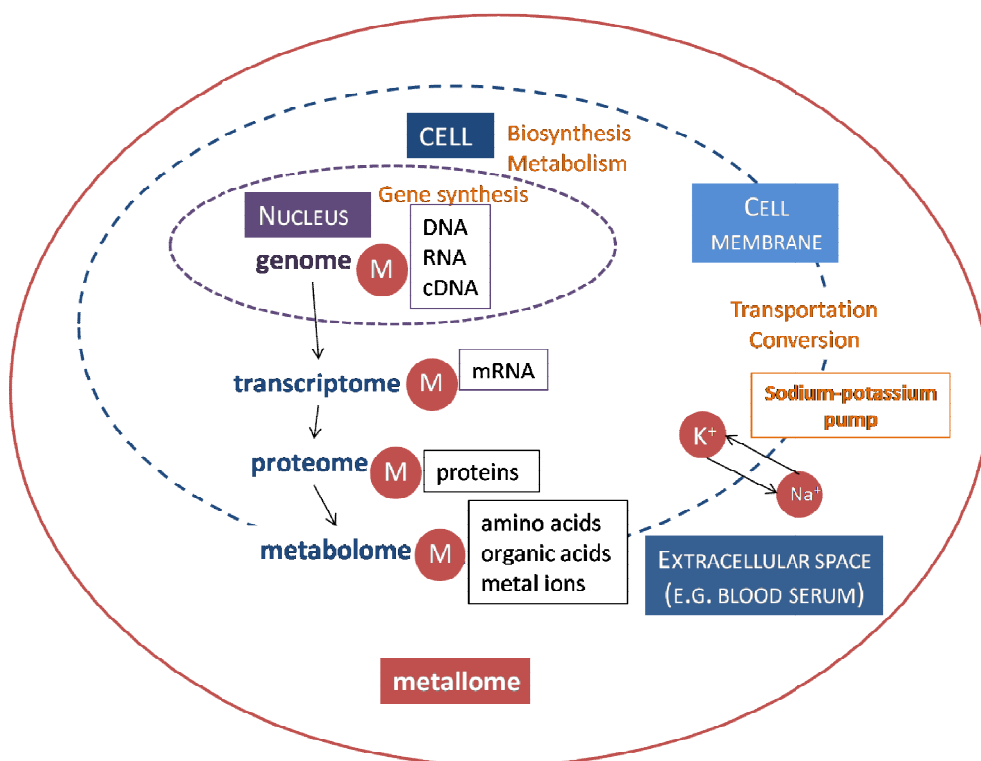


Figure 5.1 – A schematic model of a biological system, showing the relationships between genomics, proteomics, metabolomics and metallomics. The involvement of metal ions is depicted by ‘M’. Adapted from (Haraguchi, 2004).

5.2 Multi-technique approaches to studying metalloproteins

The analysis of metalloproteins includes three major components. Firstly, a separation technique, such as liquid chromatography or 2D-PAGE is required to separate out the proteins from a complex mixture isolated from the biological system of interest. Then, an atomic detector is needed for elemental quantification of the metal, and a detector capable of tandem mass spectrometry (MS/MS) for characterisation of the biomolecule (protein or peptide). Using these techniques together provides complementary data for the sensitive detection, quantification and identification of metalloproteins (Lobinski et al., 2006). Traditional mass spectrometry techniques are widely used for the characterisation of biological molecules, but their application in the analysis of metalloproteins has its limitations. Probing metal-protein interactions by ESI-MS is hampered by its poor tolerance of

salt buffers and solubilising agents (e.g. detergents) which are often used to maintain stability of the sample. Little success has been achieved with MALDI, as protein complexes are often not preserved during the ionisation process. There has been some success for analysing metal-containing non-covalent complexes using these types of mass spectrometry (which will be discussed later) but they are unable to provide information on metal content. This is where the need arises for an atomic technique capable of identifying and quantifying inorganic constituents of a protein or protein complex.

Analytical atomic spectrometry has not always been particularly popular for the speciation of metallobiomolecules, with the majority of applications driven by analytical chemistry and not biochemistry (Jakubowski *et al.*, 2004). Limitations have prevented the use of atomic techniques for biological work. Detection limits are often unsuitable for biologically relevant concentrations, and problems are often encountered when dealing with the complex sample matrix that a sample often requires. The exception, and therefore the approach almost exclusively used for inorganic biochemistry, is inductively coupled plasma mass spectrometry (ICP-MS), favoured over alternative techniques such as inductively coupled plasma optical emission spectrometry (ICP-OES) due to its extremely low detection limits. The latter is of particular importance when considering samples at biological concentrations. An overview of the current techniques being employed in metalloprotein analysis can be seen in Figure 5.2.

5.3 ICP-MS for metal identification and quantification

5.3.1 Principles of ICP-MS

Inductively coupled plasma mass spectrometry (ICP-MS) has become an important part of metallomics, its main advantage being the ability to discriminate between metal-containing and metal-free species. ICP-MS combines the use of plasma formed from a gas (usually argon) at atmospheric pressure with a mass spectrometer operating under vacuum conditions (Houk *et al.*, 1980). A stable, high-temperature plasma is generated from the gas by seeding it with a spark from a Tesla unit (similar

to that used on a car spark plug). The plasma torch is designed in such a way as to allow a sample to be injected directly into the heart of the plasma.

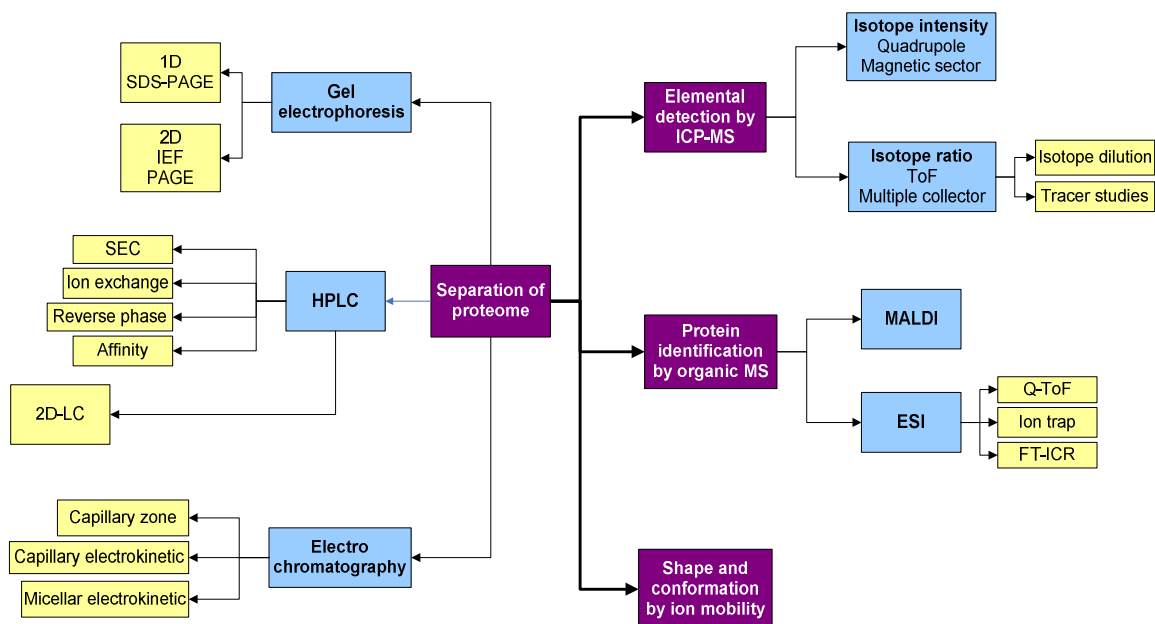


Figure 5.2 – Current hyphenated techniques for the study of metalloproteomes. Adapted from (Lobinski et al., 2006).

The sample consists of a fine aerosol, which can come from any number of sources including, but not limited to, nebulised liquids and ablated solids. As the sample passes through the plasma, it collides with free electrons, argon cations and neutral argon atoms. The result is that any molecules present in the sample are quickly and completely broken down to charged atoms. These ions are then passed through a series of apertures (cones) into a high vacuum mass analyser. The isotopes of the elements present are identified by their mass-to-charge ratio (m/z), and the intensity of a specific peak in the mass spectrum is proportional to the amount of that isotope in the original sample. Ions produced by the ICP are principally atomic and singly-charged, making this technique ideal for atomic analysis. A schematic of the process can be seen in Figure 5.3.

The three main separation principles used in ICP-MS systems are quadrupole, magnetic sector, and time-of-flight (TOF). Quadrupole MS is used in the majority of ICP-MS instruments, although some systems utilise a magnetic sector analyser, typically employed when higher resolution is required.

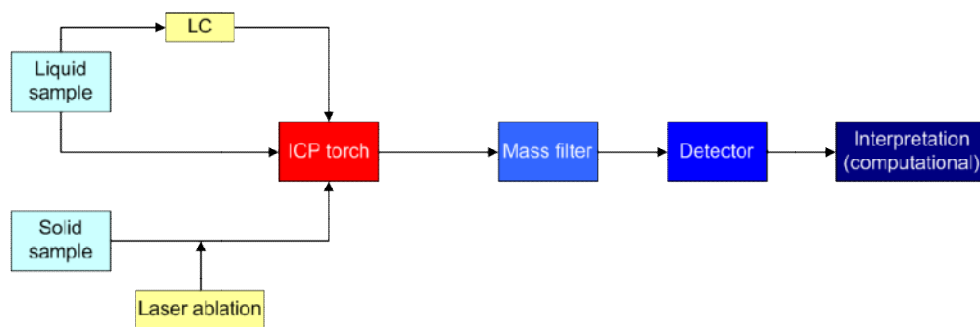


Figure 5.3 – An overview of the processes involved in ICP-MS. Sample introduction can be from liquid or solid state, and the type of mass analyser will vary according to specific instrumentation.

A diagram of a quadrupole ICP-MS instrument can be seen in Figure 5.4. Widely appreciated for its isotope specificity, versatility and sensitivity, this technique was originally applied to geological samples, but is being increasingly used for the analysis of biological samples (Muller *et al.*, 2005). This is largely due to developments in ICP-MS which tackle previous limitations such as interferences and sample handling.

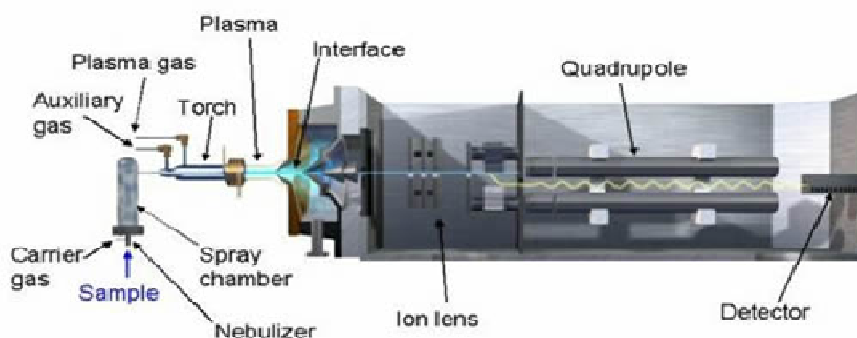


Figure 5.4 – Sample is introduced into the instrument from either a liquid or solid source. The sample, carried in argon gas, is nebulised and then enters the ICP torch where it is ionised; the ions pass to the mass filter, then to the detector. The detector is typically an electron multiplier which converts the incoming electrons into an electrical signal which is interpreted by software.

5.3.2 Overcoming interferences

Interferences in ICP-MS can be classified into two major groups: spectral and non-spectral. Spectral interferences arise from other elements (isobaric interferences), polyatomic ions or (to a lesser extent) doubly charged ions (Rowan and Houk, 1989). The constant nature of isotope ratios for most naturally occurring elements means that elemental isobaric interferences can be corrected mathematically by monitoring the intensity of an isotope of the interfering element which is free from spectral interferences (Van Veen *et al.*, 1994). Polyatomic ions can prove more challenging. These molecular interferences can be produced by the combination of two or more atoms and/or ions, and are usually associated with the argon plasma, atmospheric gases, or matrix components of the solvent or sample (Tan and Horlick, 1987). For example, a common interference when looking at the biologically relevant element, iron, is argon oxide; $^{40}\text{Ar}^{16}\text{O}$ gives the sample signal response as the ^{56}Fe isotope. Some methods employed to overcome such interferences are to choose an interference-free isotope to monitor (Vanhaecke *et al.*, 2002), removing the matrix (Evans and Giglio, 1993), the use of mathematical equations (Cao *et al.*, 2001), using cool plasma conditions (Hasan *et al.*, 2001), using a high resolution mass analyser, or utilising a collision or reactor cell (Tanner and Baranov, 1999). This latter approach has been employed in the work described here. A quadrupole instrument was used with an incorporated reaction cell, known as a dynamic reaction cell (DRC). A schematic can be seen in Figure 5.5.

5.3.3 Sample introduction

The main goal of an ICP-MS sample introduction system is to introduce a normally wet sample aerosol into the plasma, without causing destabilisation and resultant extinction of the plasma. It is also desirable to transfer the maximum amount of analyte into the plasma in the most suitable form. Most commonly, the sample is supplied as a liquid, preferred due to its homogeneity, ease of handling and the ability to simply prepare calibration standards as required (Mora *et al.*, 2003). The calibration process for liquid samples often achieves very high accuracy (Becker, J. Sabine and Jakubowski, 2009). The liquid sample must be converted to an aerosol by a nebuliser, which can take several forms. An overview of these is given in

Figure 5.6. A filter chamber then filters the aerosol and transports it to the plasma; a desolvation system can also be used to present a drier aerosol which can improve sensitivity for aqueous solutions. To analyse organic solutions, a desolvator may be necessary, as the resultant solvent present in the aerosol after passing through a normal spray chamber may be too much for the plasma to tolerate. In this case the spray chamber would be replaced by a desolvating system. The use of autosamplers has also made liquid sample introduction a quick and efficient process. Less commonly, laser ablation (LA) has been used as a means of sample introduction.

In this method, a laser is focused on the sample and creates a plume of ablated material which can be swept into the plasma. The use of laser ablation in the analysis of biological samples is attractive as it is potentially fast and robust, requiring no reaction or derivatisation. Ablation of a metalloprotein can be carried out directly from a gel or Western blot, eliminating the problems related to recovery of protein from a gel (Feldmann, Ingo *et al.*, 2006). LA-ICP-MS has also been developed in recent times as the method of choice for imaging elements in thin cross sections of biological tissues. At present, LA-ICP-MS is accepted as one of the most sensitive techniques for imaging of biological tissues (Becker, J. Sabine *et al.*, 2008). A representation of a laser ablation cell is shown in Figure 5.7. Elements such as copper, zinc, iron and cadmium have been identified from biological samples using LA-ICP-MS (Becker, J. Sabine *et al.*, 2004; Becker, J. Susanne *et al.*, 2007; Binet *et al.*, 2003) and it has also shown potential in the study of protein phosphorylation (Bandura *et al.*, 2004). One problem associated with the use of laser ablation is that the laser heats the sample in the vicinity of the laser shot. This causes water to evaporate from the sample, which can cause interference issues, and leads to drying of the sample. In order to overcome this, cryogenically cooled ablation cells have been developed (Feldmann, Jorg *et al.*, 2002).

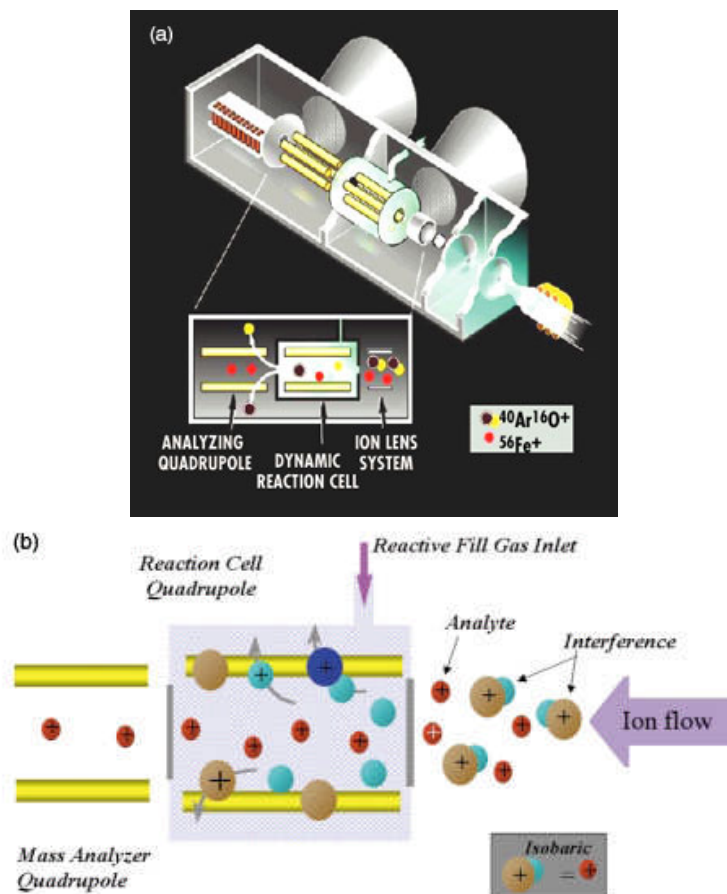


Figure 5.5 – (a) The DRC is located between the ion optics and the mass analyzer quadrupole. It consists of a quadrupole placed inside an enclosed reaction chamber. This quadrupole eliminates polyatomic interferences caused by the combination of plasma gases and sample-matrix constituents before they can enter the analyzing quadrupole. Gas inlets pressurize the reaction chamber with a low flow of reaction gas, in this case, helium. The reaction gas is selected based on its predictable ability to undergo a gas phase chemical reaction with the interfering species and remove the interference. Interference removal can occur through various processes, including collisional dissociation, electron transfer, proton transfer and oxidation.

(b) Analyte and interfering ions from the ICP enter the DRC. The reaction gas combines with the interfering ions, creating a non-interfering reaction product at a different mass. The DRC eliminates reaction by-products using a function known as Dynamic Bandpass Tuning (DBT). The DBT mechanism ejects the precursor ions before they can react to form new interferences – a real concern with complex sample matrices, as are often required for biological samples. Figure from Perkin Elmer technical note.

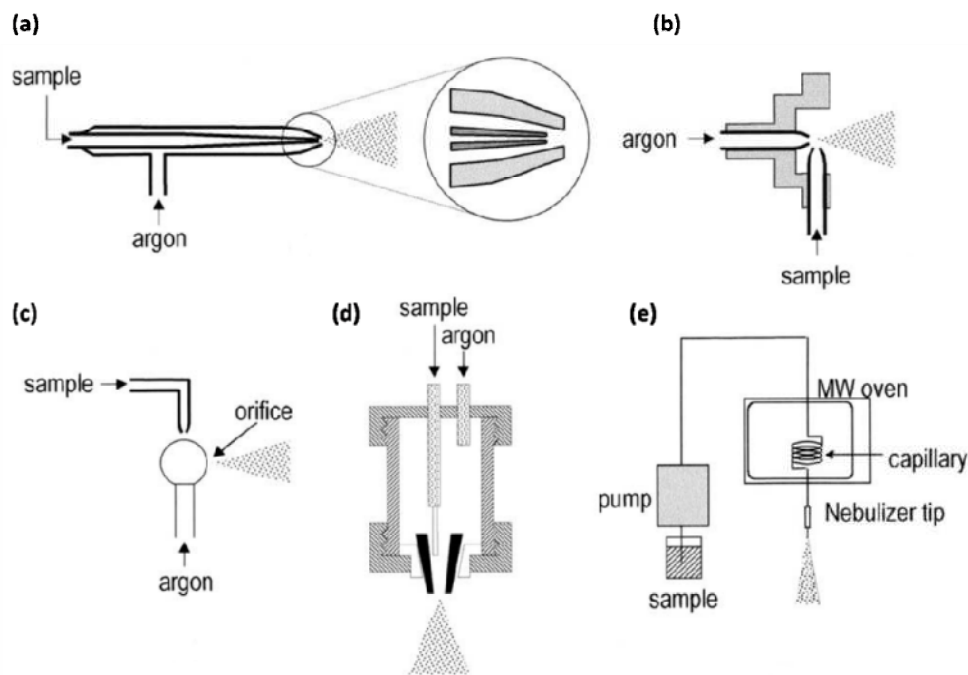


Figure 5.6 – Pneumatic nebulisers for liquid sample introduction in ICP-MS analysis, (a) concentric, (b) cross-flow, (c) Babington, (d) single-bore high-pressure, (e) microwave thermal. Adapted from (Mora et al., 2003).

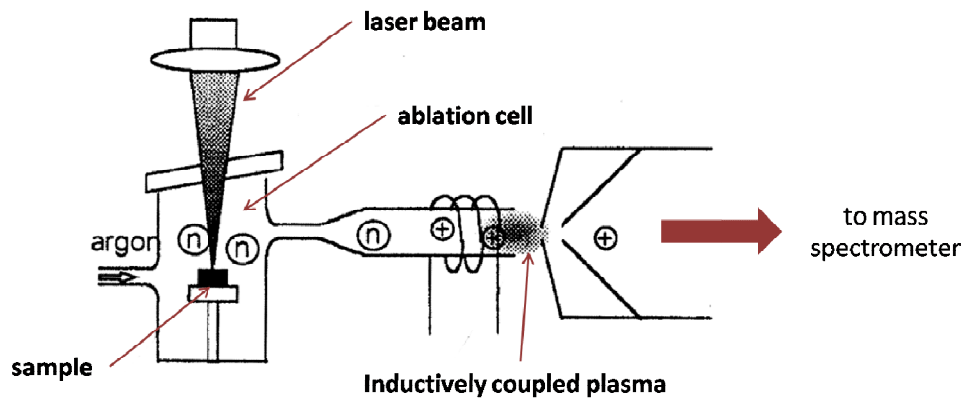


Figure 5.7 – A laser ablation inductively coupled plasma source.

5.4 Organic mass spectrometry methods for the study of metalloproteins

Mass spectrometry has become increasingly important in the study of protein structure, in areas such as protein complex assembly, and the interactions of subunits and ligands. In the field of metalloprotein research, as inorganic MS only provides information on the presence of the metal atoms, the complementary ability to probe the overall structure of the protein of interest is crucial. The inception of ESI and MALDI gave huge impetus to the structural study of proteins and non-covalent protein complexes by mass spectrometry, as these soft ionisation techniques enable proteins and complexes to remain intact. ESI is used as the method of choice for non-covalent analysis, although success has also been achieved with MALDI (Wenzel *et al.*, 2005; Yanes *et al.*, 2006). Often, however, the matrix required is not an ideal environment to keep proteins under physiologically relevant conditions (Heck and van den Heuvel, 2004). Over the years, ESI has emerged as a powerful tool for producing intact ions in vacuo from large and complex species in solution (Fenn *et al.*, 1989; Loo, 2000). The typical ESI mass spectrum of a protein consists of an envelope of peaks attributed to a series of multiply charged gas-phase ions that can indicate the stability and compactness of its structure in the gas phase (Chowdhury *et al.*, 1990); multiply charged ions are produced by proton attachment, predominantly to exposed basic sites on the protein. Recent work suggests that those ions of lowest charge are most representative of native structure (Scarff *et al.*, 2008).

Non-covalent complexes usually appear with lower net charge compared to ions produced from denatured proteins of similar size, possibly due to the fact that basic amino acids are buried in the structure. This translates to high m/z values, often outside the range of quadrupole analysers, meaning that study of non-covalent complexes is predominantly performed on TOF instruments. Although TOFs theoretically have an unlimited m/z range, it was observed that detecting large complexes was problematic unless the pressures in the first vacuum chambers were increased (Tahallah *et al.*, 2001). Larger ions generated by ESI may acquire excessive energy resulting in their missing of the detector, either partially or completely (Chernushevich and Thomson, 2004). The ions can be thermalised by collisions at elevated pressure with gas molecules. This collisional cooling dissipates the excess energy, allowing for more efficient transmission (Douglas and French,

1992; Krutchinsky *et al.*, 1998). Questions have been raised as to whether the gas-phase measurements of MS can be correlated to solution-phase (i.e. physiological) events. Whilst there are certain examples where ESI-MS data are not completely faithful to solution-phase characteristics (Robinson *et al.*, 1996), there is an increasing body of work that indicates MS can provide an indication of native structure and behaviour (Hunter *et al.*, 1997; Kaddis *et al.*, 2007; Loo, 1997). The ability to maintain protein structures and protein complexes in ESI-MS enables the study of proteins binding ligands, such as metal ions, in a near-native state. This has been successfully demonstrated for a number of metal cofactors, such as calcium binding to calbindin (Veenstra *et al.*, 1997).

5.5 Ion mobility

The mobility of an ion is a measure of how rapidly it moves through a buffer gas under the influence of a weak electric field. This time is related to the rotationally averaged collision cross section, mass and charge of the ion (Mesleh *et al.*, 1996). The development of high-resolution ion mobility spectrometry (IMS) techniques coupled with MS has provided a powerful tool for the determination of molecular structure (Kanu *et al.*, 2008). Traditional IMS is measured as drift time, i.e. the time for the ion to move through the cell containing the buffer gas, and in low-field conditions can be thought of as directed diffusion. Under these conditions, the velocity of the ion is directly proportional to the electric field. This proportionality is called the ion mobility constant (K) and is related to the ion's collision cross section by the following equation:

$$K = \left(\frac{3q}{16N} \right) \left(\frac{2\pi}{kT} \right)^{1/2} \left(\frac{m + M}{mM} \right)^{1/2} \left(\frac{1}{\Omega} \right)$$

q is the charge on the ion, N is the number density of the buffer gas, k is the Boltzmann constant, T is the absolute temperature, m is the mass of the buffer gas, M is the mass of the ion and Ω is the collision cross section of the ion. There are two modes of drift-time IMS: reduced-pressure (RPIMS) and ambient-pressure (APIMS). RPIMS has the main advantage of efficient transfer of ions from the mobility cell to

the mass spectrometer (McAfee *et al.*, 1967); APIMS is predominantly used as a field-deployable stand-alone instrument for the separation and identification of explosives, drugs and chemical warfare agents (Creaser *et al.*, 2004).

Aspiration ion spectrometry is practically employed in the design of the IMCell. The flow of buffer gas is perpendicular to the direction of the electric field. Both positive and negative ions can be measured, as they travel in opposite directions. Another method passes ions between two flat parallel electrodes or two concentric cylinder electrodes with a flowing buffer gas similar to the aspiration design. An alternating electric field is placed between the two electrodes such that the ions move perpendicular to the gas flow in alternating directions (Guevremont, 2004). Alternating-field (AFIMS) has also been called field-ion spectrometry (FIS), field-asymmetric waveform (FAIMS) and differential-mobility spectrometry (DMS) (Kanu *et al.*, 2008).

A novel method has recently been developed, described as travelling-wave ion mobility spectrometry (TWIMS). The travelling-wave mobility cell comprises a series of electrodes arranged orthogonally to the ion transmission axis. Opposite phases of an RF voltage are applied to adjacent electrodes, creating pulses of travelling waves (T-waves) along which the ions are propelled through a background gas. The drag due to the presence of gas causes ions to periodically slip over the waves, with ions of higher mobility slipping over less often than those of lower mobility and so exiting the device first (Riba-Garcia *et al.*, 2008). A T-wave separator has been incorporated into a commercial Q-TOF instrument, the Synapt HDMS (Waters, Manchester, UK), a schematic of which is shown in figure 5.8 (Pringle *et al.*, 2007). The T-wave device does not allow for absolute cross-sectional measurements to be obtained, although these may be estimated by using reference samples of known cross sections (Ruotolo *et al.*, 2005) (Thalassinos *et al.*, 2008).

The combination of IMS with MS (often abbreviated to IMMS) adds another level of information in structural studies, providing data on shape as well as mass. IMMS has emerged as a technique complementary to the well-established methods of X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy for three-dimensional analysis (van den Heuvel and Heck, 2004). Multiple studies have shown good agreement between rotationally averaged cross-sectional measurements

obtained from X-ray and NMR experiments and those obtained by ion mobility (Scarff *et al.*, 2008; Shelimov *et al.*, 1997; Shelimov and Jarrold, 1997).

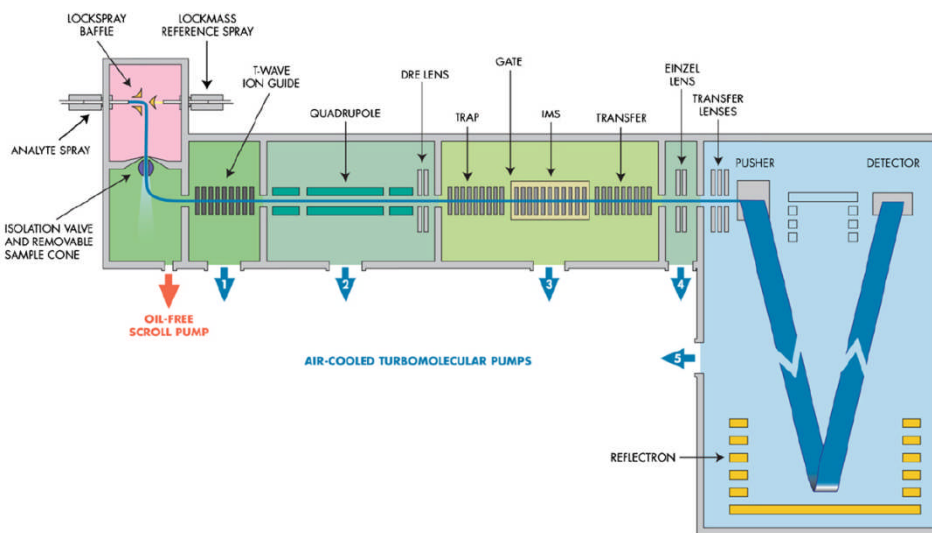


Figure 5.8 – A schematic of the Synapt HDMS instrument, which combines a quadrupole/IM separator with an orthogonal acceleration-TOF. The ion mobility section comprises three T-wave ion guides as labelled. From (Pringle *et al.*, 2007).

5.6 Particulate methane monooxygenase

Methanotrophic bacteria use methane as their sole source of carbon and energy, and are, therefore, of significant environmental interest. Not only do they play a key role in carbon cycling, they are crucially involved in the removal of the potent greenhouse gas, methane, from their surroundings (Smith *et al.*, 2000). The metabolism of methane is performed by the organisms using a methane monooxygenase enzyme, either particulate (pMMO), soluble (sMMO) or ammonia (AMO). The commonly accepted methane oxidation pathway is highlighted in Figure 5.9. With one exception (Theisen *et al.*, 2005) all methanotrophs produce pMMO, when copper concentrations are high (Murrell *et al.*, 2000). Under conditions where copper is limited in the environment, several strains can also produce sMMO. The soluble version has been well-characterised, and is known to have a carboxylate-bridged dinuclear iron centre. Despite vast study, there remains

disagreement over the metal centre of the pMMO enzyme. A summary of proposed compositions is given in Table 5.1.

The crystal structure of pMMO, purified from the organism *Methylococcus capsulatus* (Bath), was recently determined by researchers at Northwestern University (Lieberman and Rosenzweig, 2005a). The complex is composed of three subunits, pmoB (α , ~47 kDa), pmoA (β , ~24 kDa) and pmoC (γ , ~22 kDa). Three copies each of the three subunits form a cylindrical trimer, $\alpha_3\beta_3\gamma_3$, with the soluble regions derived predominantly from pmoB, and pmoA and pmoC residing primarily in the membrane. Each protomer in the trimer comprises single copies of the pmoB, pmoA and pmoC subunits. Three metal centres were identified per protomer. The first and second are located in pmoB, both assigned as copper sites, one mononuclear and one dinuclear. The third has been modelled as a mononuclear zinc site within the lipid bilayer, but inductively coupled plasma atomic emission spectroscopy (ICP-AES) analysis indicates that this is derived from zinc acetate in the crystallisation buffer. The *in vivo* occupation of this site, therefore, remains open for discussion. Additional structural studies performed by electron microscopy (EM) at Warwick show that pMMO, as described above, forms a supramolecular complex with methane dehydrogenase, the second enzyme employed in the methane oxidation pathway (Myronova *et al.*, 2006).

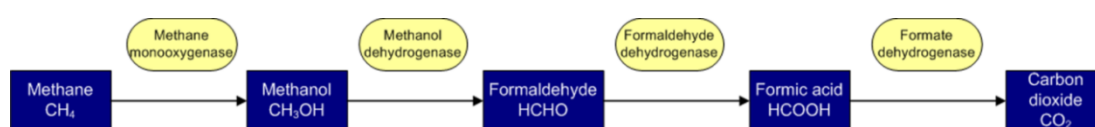


Figure 5.9 – The methane oxidation pathway of methanotrophic bacteria.

The elucidated crystal structure is shown in Figure 5.10, with the metal sites highlighted. The EM-determined complex can be seen in Figure 5.11. The oxidation of methane to methanol is challenging, as methane is the most inert hydrocarbon (104 kcal/mol C–H bond) (Lieberman and Rosenzweig, 2005a). Some studies have determined possible roles for the copper ions, but a full characterisation has not been achieved (Lieberman *et al.*, 2006; Yoshizawa and Shiota, 2006). Further information

regarding the third, unassigned metal centre in pMMO could provide crucial insight into the enzyme's catalytic mechanism.

Organism	Group	Mol Cu per 200 kDa	Mol Fe per 200 kDa	Reference
<i>Methylococcus capsulatus</i> (Bath)	Chan	24.8	~0	(Nguyen <i>et al.</i> , 1998)
		30	~0	(Nguyen <i>et al.</i> , 1998)
		27.2	~0	(Yu <i>et al.</i> , 2003)
	Dalton	4	2	(Basu <i>et al.</i> , 2003)
	DiSpirito	29	5	(Zahn and DiSpirito, 1996)
		16-20	4	(Choi <i>et al.</i> , 2003)
	Rosenzweig	4-6	1-2	(Lieberman <i>et al.</i> , 2003)
<i>Methylosinus trichosporium</i> OB3b	Okura	25.6	1.8	(Takeguchi <i>et al.</i> , 1998)
		4	0	(Miyaji <i>et al.</i> , 2002)

Table 5.1 – Proposed compositions of the metal sites in pMMO, provided as moles per 200 kDa, as gel purification had previously indicated this as the molecular mass of the complex. Adapted from (Lieberman and Rosenzweig, 2004).

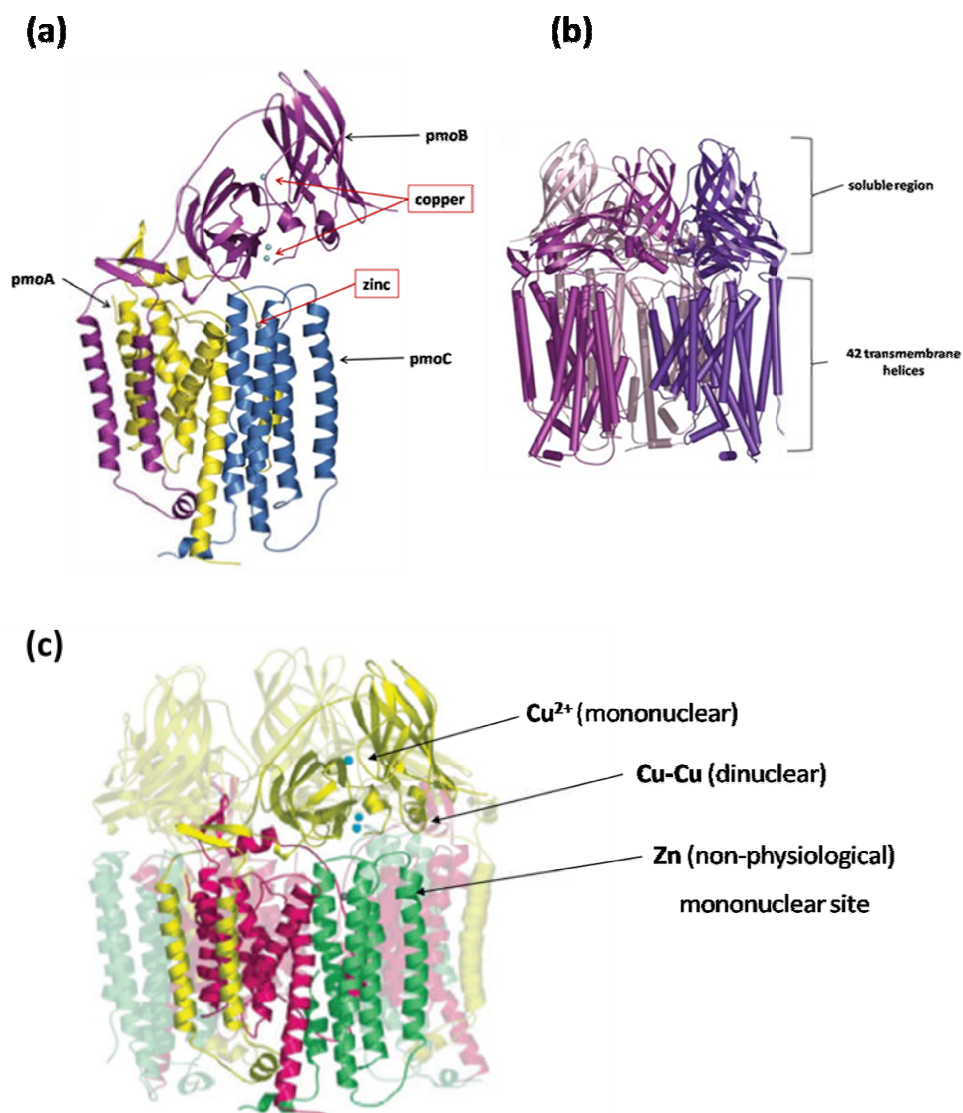


Figure 5.10 – (a) a single pMMO protomer with the constituent polypeptides labelled, the copper centres are indicated, as is the third site which crystallised with a zinc ion, (b) the pMMO trimer viewed parallel to the membrane normal (both adapted from (Lieberman and Rosenzweig, 2005a)), (c) the pMMO trimer viewed parallel to the membrane normal with one protomer and its metal centres indicated (adapted from (Lieberman and Rosenzweig, 2005b)).

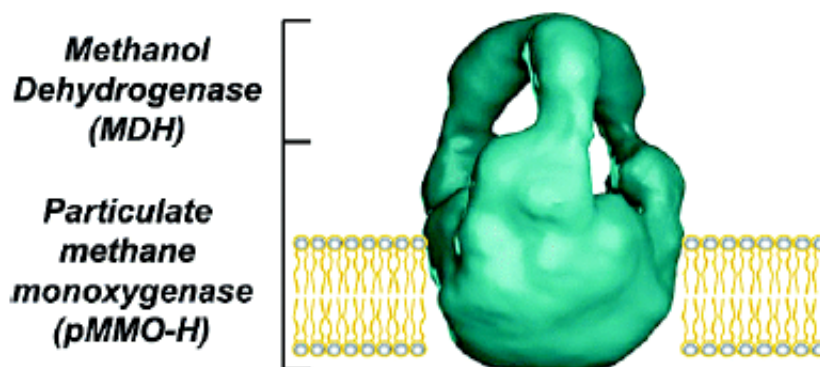


Figure 5.11 – Three-dimensional structure of pMMO as determined by EM. The membrane graphic has been added to enable visualisation of the positioning of the complex in vivo. Figure reproduced with kind permission from Dr. Nataliia Myronova.

5.7 Hemoglobin and hemoglobin disorders

Hemoglobin (Hb) is the major oxygen-transport protein found in the red blood cells of all vertebrates. It has a tetrameric structure comprising four chains, two α - and two β -, each associated with a heme group, as shown in Figure 5.12. Inherited hemoglobin disorders are the commonest diseases attributable to single defective genes. Approximately 7 % of the world's population are carriers, and 300,000 to 500,000 babies with severe forms of such disorders are born each year (WHO 1989). Although these disorders are most frequent in tropical regions, they are now encountered in most countries because of migrations of populations (Weatherall *et al.*, 2006). As such, hemoglobin has been extensively studied.

Inherited hemoglobin disorders fall into two main groups: the structural hemoglobin variants, which are predominantly caused by single amino acid mutations, and the thalassemias, which are caused by defective globin production. More than 700 structural variants have been identified, the most debilitating of these being sickle cell anaemia. The sickle cell mutation results in the production of a β -chain where the sixth residue has changed from glutamate, which is negatively charged, to valine, which is hydrophobic (Ingram, 1957). This changes the conformation of the assembled tetramer, lowering solubility and allowing molecular stacking.

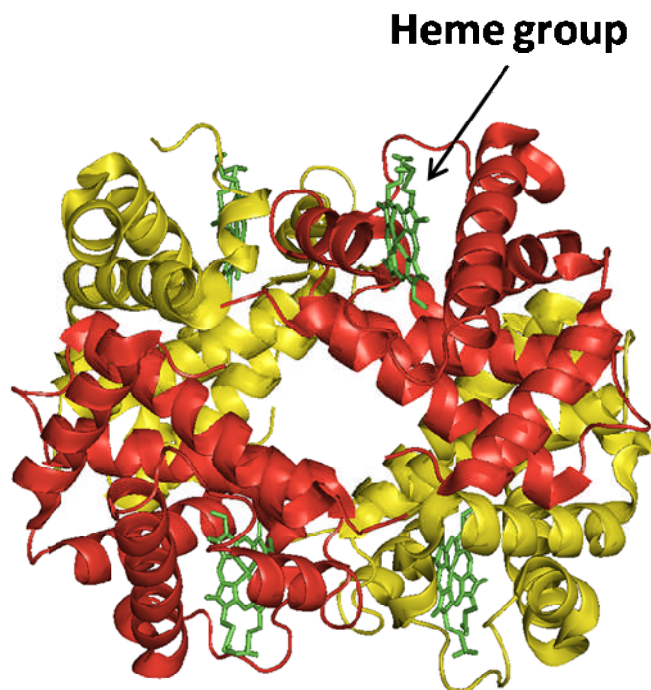


Figure 5.12 – The tetrameric structure of human hemoglobin, with the α - and β -chains shown in different colours, and the heme groups in green.

Polymerisation of the sickle hemoglobin molecule (HbS) in deoxygenated blood causes a characteristic alteration in shape of red blood cells from biconcave to crescentic (Murayama, 1967). The disease of sickle cell anaemia is recessive, and occurs when a person inherits two copies of the mutated β -globin gene. The misshapen HbS leads to shortened red cell survival and a tendency to block small blood vessels (Bunn, 1997). Clinical symptoms often present as anaemia, such as fatigue, shortness of breath and dizziness, but more acute presentations can involve bone pain, sequestration of blood into lungs, liver or spleen, or thrombosis of cerebral vessels causing strokes (Weatherall et al., 2006).

ESI-MS has been widely used to detect Hb variants in hemoglobin (Daniel *et al.*, 2005; Shackleton *et al.*, 1991; Wild *et al.*, 2004) and to investigate its structural assembly into a non-covalent complex (Boys and Konermann, 2007; Griffith and Kaltashov, 2003, 2007) and its corresponding disassembly (Versluis and Heck, 2001). The exact assembly pathway for hemoglobin is still under debate. One α - and one β -monomer come together to form a heterodimer, and two of these associate to form

the tetramer. The α - and β -monomers can exist in heme-free (apo, α^a and β^a) and heme-bound (holo, α^h and β^h) forms (Boys and Konermann, 2007). It is unclear whether the heme groups are attached to the monomers prior to dimer formation, or whether their association with each other leads to recruitment of the heme groups. One suggestion is the formation of a heme-deficient dimer intermediate, consisting of a natively folded holo- α -globin (α^h) and a partially folded apo- β -globin (β^a), before complete dimer formation leading to the correct tetrameric arrangement (Griffith and Kaltashov, 2007). Other work reports, however, that the heme-deficient dimer is seen only when using commercially available lyophilised protein, and not when using freshly prepared samples (Boys *et al.*, 2007). This study of the acid-induced denaturation of bovine Hb concluded that a highly symmetric mechanism took place:



Further insight into the assembly of the hemoglobin tetramer could be of particular importance in understanding the effects of structural Hb disorders.

5.8 Aims and objectives

The multi-technique approach to metallomic studies, highlighted in section 5.2, is summarised in Figure 5.13. Each of the three types of methodology, inorganic MS, organic MS and ion mobility, were utilised for the study of particular biological problems, with the following aims:

1. To confirm or challenge previous characterisation of the metal sites in the pMMO complex using inorganic mass spectrometry, by both laser ablation and liquid sample introduction.
2. To further elucidate the structural properties of the hemoglobin tetramer and its components, and to determine whether conformational differences between HbA and HbS can be observed by travelling wave-based ion mobility MS.

The work towards (1) was part of my CASE Award project, conducted at the industrial site of Intertek MSG, formerly ICI Measurement Science Group, under the supervision of Dr. Jeff Franks.

The research involved in (2) has been peer reviewed and published:

Scarff et al., (2009) *Probing hemoglobin structure by means of travelling-wave ion mobility mass spectrometry*, Journal of the American Society for Mass Spectrometry 20(4): 625-631

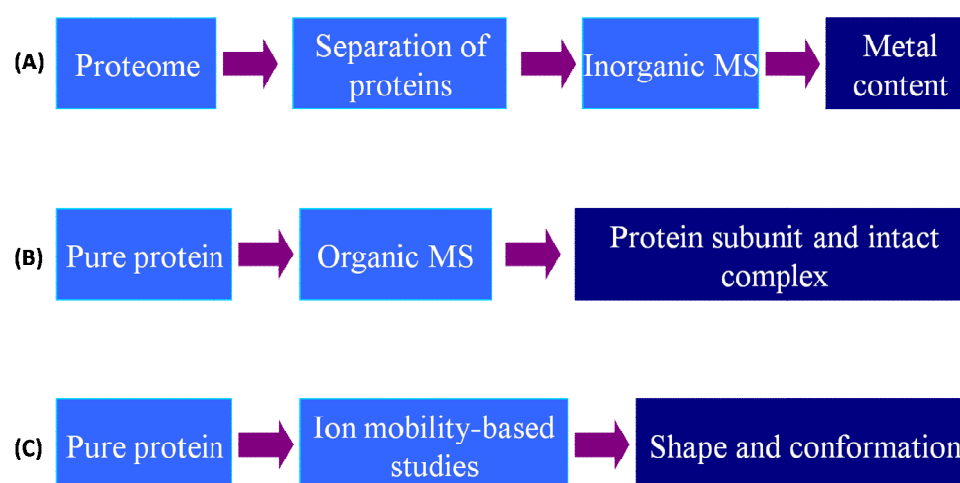


Figure 5.13 – Approaches to the study of metalloproteins, (A) was used for work on pMMO, (B) and (C) for investigating hemoglobin structure.

References

- Bandura, D. R., Ornatsky, O. I., Liao, L.** (2004). Characterization of phosphorus content of biological samples by ICP-DRC-MS: potential tool for cancer research. *J. Anal. At. Spectrom.* **19**: 96-100.
- Basu, P., Katterle, B., Andersson, K. K., Dalton, H.** (2003). The membrane-associated form of methane mono-oxygenase from *Methylococcus capsulatus* (Bath) is a copper/iron protein. *Biochem. J.* **369**: 417-427.
- Becker, J. S., Jakubowski, N.** (2009). The synergy of elemental and biomolecular mass spectrometry: new analytical strategies in life sciences. *Chem. Soc. Rev.* **38**: 1969-1983.
- Becker, J. S., Zoriy, M., Krause-Buchholz, U., Becker, J. S., Pickhardt, C., Przybylski, M., Pompe, W., Rodel, G.** (2004). In-gel screening of phosphorus and copper, zinc and iron in proteins of yeast mitochondria by LA-ICP-MS and identification of phosphorylated protein structures by MALDI-FT-ICR-MS after separation with two-dimensional gel electrophoresis. *J. Anal. At. Spectrom.* **19**: 1236-1243.
- Becker, J. S., Zoriy, M., Przybylski, M., Becker, J. S.** (2007). High resolution mass spectrometric brain proteomics by MALDI-FTICR-MS combined with determination of P, S, Cu, Zn and Fe by LA-ICP-MS. *Int. J. Mass spectrom.* **261**: 68-73.
- Becker, J. S., Zoriy, M., Wu, B., Matusch, A., Becker, J. S.** (2008). Imaging of essential and toxic elements in biological tissues by LA-ICP-MS. *J. Anal. At. Spectrom.* **23**: 1275-1280.
- Binet, M. R. B., Ma, R., McLeod, C. W., Poole, R. K.** (2003). Detection and characterization of zinc- and cadmium-binding proteins in *Escherichia coli* by gel electrophoresis and laser ablation-inductively coupled plasma-mass spectrometry. *Anal. Biochem.* **318**: 30-38.
- Boys, B. L., Konermann, L.** (2007). Folding and Assembly of Hemoglobin Monitored by Electrospray Mass Spectrometry Using an On-line Dialysis System. *J. Am. Soc. Mass Spectrom.* **18**: 8-16.
- Boys, B. L., Kuprowski, M. C., Konermann, L.** (2007). Symmetric Behavior of Hemoglobin α - and β - Subunits during Acid-Induced Denaturation Observed by Electrospray Mass Spectrometry. *Biochemistry* **46**: 10675-10684.
- Bunn, H. F.** (1997). Pathogenesis and Treatment of Sickle Cell Disease. *N Engl J Med* **337**: 762-769.
- Cao, X., Yin, M., Wang, X.** (2001). Elimination of the spectral interference from polyatomic ions with rare earth elements in inductively coupled plasma mass spectrometry by combining algebraic correction with chromatographic separation. *Spectrochim. Acta, Part B* **56**: 431-441.

- Chernushevich, I. V., Thomson, B. A.** (2004). Collisional Cooling of Large Ions in Electrospray Mass Spectrometry. *Anal. Chem.* **76**: 1754-1760.
- Choi, D.-W., Kunz, R. C., Boyd, E. S., Semrau, J. D., Antholine, W. E., Han, J. I., Zahn, J. A., Boyd, J. M., de la Mora, A. M., DiSpirito, A. A.** (2003). The Membrane-Associated Methane Monooxygenase (pMMO) and pMMO-NADH:Quinone Oxidoreductase Complex from *Methylococcus capsulatus* Bath. *J. Bacteriol.* **185**: 5755-5764.
- Chowdhury, S. K., Katta, V., Chait, B. T.** (1990). Probing conformational changes in proteins by mass spectrometry. *J. Am. Chem. Soc.* **112**: 9012-9013.
- Creaser, C. S., Griffiths, J. R., Bramwell, C. J., Noreen, S., Hill, C. A., Thomas, C. L. P.** (2004). Ion mobility spectrometry: a review. Part 1. Structural analysis by mobility measurement. *The Analyst* **129**: 984-994.
- Daniel, Y. A., Turner, C., Haynes, R. M., Hunt, B. J., Dalton, R. N.** (2005). Rapid and specific detection of clinically significant haemoglobinopathies using electrospray mass spectrometry-mass spectrometry. *Br. J. Haematol.* **130**: 635-643.
- Douglas, D. J., French, J. B.** (1992). Collisional focusing effects in radio frequency quadrupoles. *J. Am. Soc. Mass Spectrom.* **3**: 398-408.
- Evans, E. H., Giglio, J. J.** (1993). Interferences in inductively coupled plasma mass spectrometry. A review. *J. Anal. At. Spectrom.* **8**: 1 - 18.
- Feldmann, I., Koehler, C. U., Roos, P. H., Jakubowski, N.** (2006). Optimisation of a laser ablation cell for detection of hetero-elements in proteins blotted onto membranes by use of inductively coupled plasma mass spectrometry. *J. Anal. At. Spectrom.* **21**: 1006-1015.
- Feldmann, J., Kindness, A., Ek, P.** (2002). Laser ablation of soft tissue using a cryogenically cooled ablation cell. *J. Anal. At. Spectrom.* **17**: 813-818.
- Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., Whitehouse, C. M.** (1989). Electrospray ionization for mass spectrometry of large biomolecules. *Science* **246**: 64-71.
- Griffith, W. P., Kaltashov, I. A.** (2003). Highly asymmetric interactions between globin chains during hemoglobin assembly revealed by electrospray ionization mass spectrometry. *Biochemistry* **42**: 10024-10033.
- Griffith, W. P., Kaltashov, I. A.** (2007). Protein Conformational Heterogeneity as a Binding Catalyst: ESI-MS Study of Hemoglobin H Formation. *Biochemistry* **46**: 2020-2026.
- Guevremont, R.** (2004). High-field asymmetric waveform ion mobility spectrometry: A new tool for mass spectrometry. *J. Chromatogr.* **1058**: 3-19.

- Haraguchi, H.** (2004). Metallomics as integrated biometal science. *J. Anal. At. Spectrom.* **19**: 5-14.
- Hasan, T., Praphairaksit, N., Houk, R. S.** (2001). Low flow, externally air cooled torch for inductively coupled plasma atomic emission spectrometry with axial viewing. *Spectrochim. Acta, Part B* **56**: 409-418.
- Heck, A. J. R., van den Heuvel, R. H. H.** (2004). Investigation of intact protein complexes by mass spectrometry. *Mass Spectrom. Rev.* **23**: 368-389.
- Houk, R. S., Fassel, V. A., Flesch, G. D., Svec, H. J., Gray, A. L., Taylor, C. E.** (1980). Inductively coupled argon plasma as an ion source for mass spectrometric determination of trace elements. *Anal. Chem.* **52**: 2283-2289.
- Hunter, C. L., Mauk, A. G., Douglas, D. J.** (1997). Dissociation of heme from myoglobin and cytochrome b(5): Comparisons of behaviour in solution and the gas phase. *Biochemistry* **36**: 1018 - 1025.
- Ingram, V.** (1957). Gene Mutations in Human Haemoglobin: the Chemical Difference Between Normal and Sickle Cell Haemoglobin. *Nature* **180**: 326-328.
- Jakubowski, N., Lobinski, R., Moens, L.** (2004). Metallobiomolecules. The basis of life, the challenge of atomic spectroscopy. *J. Anal. At. Spectrom.* **19**: 1-4.
- Kaddis, C. S., Lomeli, S. H., Yin, S., Berhane, B., Apostol, M. I., Kickhoefer, V. A., Rome, L. H., Loo, J. A.** (2007). Sizing Large Proteins and Protein Complexes by Electrospray Ionization Mass Spectrometry and Ion Mobility. *J. Am. Soc. Mass Spectrom.* **18**: 1206-1216.
- Kanu, A. B., Dwivedi, P., Tam, M., Matz, L., Hill, H. H. J.** (2008). Ion mobility-mass spectrometry. *J. Mass Spectrom.* **43**: 1-22.
- Krutchinsky, A. N., Chernushevich, I. V., Spicer, V. L., Ens, W., Standing, K. G.** (1998). Collisional Damping Interface for an Electrospray Ionization Time-of-Flight Mass Spectrometer. *J. Am. Soc. Mass Spectrom.* **9**: 569-579.
- Lieberman, R. L., Kondapalli, K. C., Shrestha, D. B., Hakemian, A. S., Smith, S. M., Telser, J., Kuzelka, J., Gupta, R., Borovik, A. S., Lippard, S. J., Hoffman, B. M., Rosenzweig, A. C., Stemmler, T. L.** (2006). Characterization of the Particulate Methane Monooxygenase Metal Centers in Multiple Redox States by X-ray Absorption Spectroscopy. *Inorg. Chem.* **45**: 8372-8381.
- Lieberman, R. L., Rosenzweig, A. C.** (2004). Biological methane oxidation: Regulation, biochemistry, and active site structure of particulate methane monooxygenase. *Crit. Rev. Biochem. Mol. Biol.* **39**: 147-164.
- Lieberman, R. L., Rosenzweig, A. C.** (2005a). Crystal structure of a membrane-bound metalloenzyme that catalyses the biological oxidation of methane. *Nature* **434**: 177-182.

- Lieberman, R. L., Rosenzweig, A. C.** (2005b). The quest for the particulate methane monooxygenase active site. *Dalton Transactions*: 3390-3396.
- Lieberman, R. L., Shrestha, D. B., Doan, P. E., Hoffman, B. M., Stemmler, T. L., Rosenzweig, A. C.** (2003). Purified particulate methane monooxygenase from *Methylococcus capsulatus* (Bath) is a dimer with both mononuclear copper and a copper-containing cluster. *Proc. Natl. Acad. Sci. USA* **100**: 3820-3825.
- Lobinski, R., Schaumlöffel, D., Szpunar, J.** (2006). Mass spectrometry in bioinorganic analytical chemistry. *Mass Spectrom. Rev.* **25**: 255-289.
- Loo, J. A.** (1997). Studying noncovalent protein complexes by electrospray ionization mass spectrometry. *Mass Spectrom. Rev.* **16**: 1-23.
- Loo, J. A.** (2000). Electrospray ionization mass spectrometry: a technology for studying noncovalent macromolecular complexes. *Int. J. Mass spectrom.* **200**: 175-186.
- McAfee, K. B., Sipler, D., Edelson, D.** (1967). Mobilities and Reactions of Ions in Argon. *Phys. Rev.* **160**: 130.
- Mesleh, M. F., Hunter, J. M., Shvartsburg, A. A., Schatz, G. C., Jarrold, M. F.** (1996). Structural information from ion mobility measurements: Effects of the long-range potential. *J. Phys. Chem.* **100**: 16082-16086.
- Miyaji, A., Kamachi, T., Okura, I.** (2002). Improvement of the purification method for retaining the activity of the particulate methane monooxygenase from *Methylosinus trichosporium* OB3b. *Biotechnol. Lett.* **24**: 1883-1887.
- Mora, J., Maestre, S., Hernandis, V., Todolí, J. L.** (2003). Liquid-sample introduction in plasma spectrometry. *TrAC, Trends Anal. Chem.* **22**: 123-132.
- Muller, S. D., Diaz-Bone, R. A., Felix, J., Goedecke, W.** (2005). Detection of specific proteins by laser ablation inductively coupled plasma mass spectrometry (LA-ICP-MS) using gold cluster labelled antibodies. *J. Anal. At. Spectrom.* **20**: 907-911.
- Murayama, M.** (1967). Structure of Sick Cell Hemoglobin and Molecular Mechanism of the Sickling Phenomenon. *Clin. Chem.* **13**: 578-588.
- Murrell, J. C., McDonald, I. R., Gilbert, B.** (2000). Regulation of expression of methane monooxygenases by copper ions. *Trends Microbiol.* **8**: 221-225.
- Myronova, N., Kitmitto, A., Collins, R. F., Miyaji, A., Dalton, H.** (2006). Three-Dimensional Structure Determination of a Protein Supercomplex That Oxidizes Methane to Formaldehyde in *Methylococcus capsulatus* (Bath). *Biochemistry* **45**: 11905-11914.
- Nguyen, H.-H. T., Elliott, S. J., Yip, J. H.-K., Chan, S. I.** (1998). The Particulate Methane Monooxygenase from *Methylococcus capsulatus* (Bath) Is a Novel

- Copper-containing Three-subunit Enzyme. Isolation and characterization. *J. Biol. Chem.* **273**: 7957-7966.
- Pringle, S. D., Giles, K., Wildgoose, J. L., Williams, J. P., Slade, S. E., Thalassinos, K., Bateman, R. H., Bowers, M. T., Scrivens, J. H.** (2007). An investigation of the mobility separation of some peptide and protein ions using a new hybrid quadrupole/travelling wave IMS/oa-ToF instrument. *Int. J. Mass spectrom.* **261**: 1-12.
- Riba-Garcia, I., Giles, K., Bateman, R. H., Gaskell, S. J.** (2008). Evidence for Structural Variants of a- and b-Type Peptide Fragment Ions Using Combined Ion Mobility/Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **19**: 609-613.
- Robinson, C. V., Chung, E. W., Kragelund, B. B., Knudsen, J., Aplin, R. T., Poulsen, F. M., Dobson, C. M.** (1996). Probing the Nature of Noncovalent Interactions by Mass Spectrometry. A Study of Protein CoA Ligand Binding and Assembly. *J. Am. Chem. Soc.* **118**: 8646-8653.
- Rowan, J. T., Houk, R. S.** (1989). Attenuation of Polyatomic Ion Interferences in Inductively Coupled Plasma Mass Spectrometry by Gas-Phase Collisions. *Appl. Spectrosc.* **43**: 976-980.
- Ruotolo, B. T., Giles, K., Campuzano, I., Sandercock, A. M., Bateman, R. H., Robinson, C. V.** (2005). Evidence for Macromolecular Protein Rings in the Absence of Bulk Water. *Science*: 1658-1660.
- Scarff, C. A., Thalassinos, K., Hilton, G. R., Scrivens, J. H.** (2008). Travelling wave ion mobility mass spectrometry studies of protein structure: biological significance and comparison with X-ray crystallography and nuclear magnetic resonance spectroscopy measurements. *Rapid Commun. Mass Spectrom.* **22**: 3297-3304.
- Shackleton, C. H. L., Falick, A. M., Green, B. N., Witkowska, H. E.** (1991). Electrospray mass spectrometry in the clinical diagnosis of variant hemoglobins. *J. Chromatogr.* **562**: 175-190.
- Shelimov, K. B., Clemmer, D. E., Hudgins, R. R., Jarrold, M. F.** (1997). Protein Structure in Vacuo: Gas-Phase Conformations of BPTI and Cytochrome c. *J. Am. Chem. Soc.* **119**: 2240-2248.
- Shelimov, K. B., Jarrold, M. F.** (1997). Conformations, Unfolding, and Refolding of Apomyoglobin in Vacuum: An Activation Barrier for Gas-Phase Protein Folding. *J. Am. Chem. Soc.* **119**: 2987-2994.
- Smith, K. A., Dobbie, K. E., Ball, B. C., Bakken, L. R., Sitaula, B. K., Hansen, S., Brumme, R., Borken, W., Christensen, S., Priemé, A., Fowler, D., Macdonald, J. A., Skiba, U., Klemedtsson, L., Kasimir-Klemedtsson, A., Degórska, A., Orlanski, P.** (2000). Oxidation of atmospheric methane in Northern European soils, comparison with other ecosystems, and uncertainties in the global terrestrial sink. *Global Change Biol.* **6**: 791-803.

- Szpunar, J.** (2004). Metallomics: a new frontier in analytical chemistry. *Anal. Bioanal. Chem.* **378**: 54-56.
- Szpunar, J.** (2005). Advances in analytical methodology for bioinorganic speciation analysis: metallomics, metalloproteomics and heteroatom-tagged proteomics and metabolomics. *The Analyst* **130**: 442-465.
- Tahallah, N., Pinkse, M. M., C. S. , Heck, A. J. R.** (2001). The effect of the source pressure on the abundance of ions of noncovalent protein assemblies in an electrospray ionization orthogonal time-of-flight instrument. *Rapid Commun. Mass Spectrom.* **15**: 596-601.
- Takeguchi, M., Miyakawa, K., Okura, I.** (1998). Properties of the membranes containing the particulate methane monooxygenase from *Methylosinus trichosporium* OB3b. *BioMetals* **11**: 229-234.
- Tan, S. H., Horlick, G.** (1987). Matrix-effect observations in inductively coupled plasma mass spectrometry. *J. Anal. At. Spectrom.* **2**: 745 - 763.
- Tanner, S. D., Baranov, V. I.** (1999). A dynamic reaction cell for inductively coupled plasma mass spectrometry (ICP-DRC-MS). II. Reduction of interferences produced within the cell. *J. Am. Soc. Mass Spectrom.* **10**: 1083-1094.
- Thalassinos, K., Grabenauer, M., Slade, S. E., Hilton, G. R., Bowers, M. T., Scrivens, J. H.** (2008). Characterization of Phosphorylated Peptides Using Traveling Wave-Based and Drift Cell Ion Mobility Mass Spectrometry. *Anal. Chem.* **81**: 248-254.
- Theisen, A. R., Ali, M. H., Radajewski, S., Dumont, M. G., Dunfield, P. F., McDonald, I. R., Dedysh, S. N., Miguez, C. B., Murrell, J. C.** (2005). Regulation of methane oxidation in the facultative methanotroph *Methylocella silvestris* BL2. *Mol. Microbiol.* **58**: 682-692.
- van den Heuvel, R. H. H., Heck, A. J. R.** (2004). Native protein mass spectrometry: from intact oligomers to functional machineries. *Curr. Opin. Chem. Biol.* **8**: 519-526.
- Van Veen, E. H., Bosch, S., De Loos-Vollebregt, M. T. C.** (1994). Spectral interpretation and interference correction in inductively coupled plasma mass spectrometry. *Spectrochim. Acta, Part B* **49**: 1347-1361.
- Vanhaecke, F., Balcaen, L., Wannemacker, G. D., Moens, L.** (2002). Capabilities of inductively coupled plasma mass spectrometry for the measurement of Fe isotope ratios. *J. Anal. At. Spectrom.* **17**: 933-943.
- Veenstra, T. D., Johnson, K. L., Tomlinson, A. J., Naylor, S., Kumar, R.** (1997). Determination of Calcium-Binding Sites in Rat Brain Calbindin D28K by Electrospray Ionization Mass Spectrometry. *Biochemistry* **36**: 3535-3542.

- Versluis, C., Heck, A. J. R.** (2001). Gas-phase dissociation of hemoglobin. *Int. J. Mass spectrom.* **210-211**: 637-649.
- Weatherall, D., Akinyanju, O., Fucharoen, S., Olivieri, N., Musgrove, P.** (2006). *Inherited Disorders of Hemoglobin*. Oxford University Press.
- Wenzel, R. J., Matter, U., Schultheis, L., Zenobi, R.** (2005). Analysis of Megadalton Ions Using Cryodetection MALDI Time-of-Flight Mass Spectrometry. *Anal. Chem.* **77**: 4329-4337.
- Wild, B. J., Green, B. N., Stephens, A. D.** (2004). The potential of electrospray ionization mass spectrometry for the diagnosis of hemoglobin variants found in newborn screening. *Blood Cells, Mol. Dis.* **33**: 308-317.
- Yanes, O., Nazabal, A., Wenzel, R., Zenobi, R., Aviles, F. X.** (2006). Detection of Noncovalent Complexes in Biological Samples by Intensity Fading and High-Mass Detection MALDI-TOF Mass Spectrometry. *Journal of Proteome Research* **5**: 2711-2719.
- Yoshizawa, K., Shiota, Y.** (2006). Conversion of Methane to Methanol at the Mononuclear and Dinuclear Copper Sites of Particulate Methane Monooxygenase (pMMO): A DFT and QM/MM Study. *J. Am. Chem. Soc.* **128**: 9873-9881.
- Yu, S. S. F., Chen, K. H. C., Tseng, M. Y. H., Wang, Y.-S., Tseng, C.-F., Chen, Y.-J., Huang, D.-S., Chan, S. I.** (2003). Production of High-Quality Particulate Methane Monooxygenase in High Yields from *Methylococcus capsulatus* (Bath) with a Hollow-Fiber Membrane Bioreactor. *J. Bacteriol.* **185**: 5915-5924.
- Zahn, J. A., DiSpirito, A. A.** (1996). Membrane-associated methane monooxygenase from *Methylococcus capsulatus* (Bath). *J. Bacteriol.* **178**: 1018-1029.

Chapter 6

Materials and Methods

Metallomic Studies

6.1 Purification of pMMO

This purification was performed with Dr. Nataliia Myronova. Experiments used pMMO from *Methylocella capsulatus* (Bath) from the University of Warwick culture collection. Cultivation, isolation and purification were performed as previously described (Myronova *et al.*, 2006). In brief, *M. capsulatus* (Bath) was cultured in nitrate minimal salt medium with a final copper sulphate concentration of 40 μ M. Cells were harvested, washed with 25 mM piperazine-1,4-bis(2-ethanesulfonic acid) (PIPES), pH 7.2, and resuspended in the same buffer. Concentrated cells were frozen and stored at -80 °C. Thawed cells, resuspended in PIPES, were broken and debris and soluble protein were removed. Membrane-bound protein was solubilised and fractionated by gel filtration chromatography. Solubilisation was performed by adding the dissolved anionic detergent dodecyl- β -D-maltoside (Ultrol grade, Calbiochem, USA) to give a detergent/protein (w/w) ratio of 1.5 to the particulate extract (Basu *et al.*, 2003). Fractions containing pMMO were purified using a Mono Q-10 anion-exchange column (GE Healthcare, Bucks, UK). The activity of pMMO was assayed at each stage of membrane isolation and purification by following the oxidation of propylene to propylene oxide by gas chromatography. Fractions with protein concentrations of 0.08 – 0.22 mg/ml were generated, and some retained for in-solution ICP-MS analysis.

6.2 EDTA treatment

To prepare samples for control analysis, ethylenediamine tetraacetic acid (EDTA) was used to remove any adventitious metals that may have been associated with the protein. Samples were incubated with 10 mM EDTA for one hour on ice, then de-salted using a pre-cooled HiTrap column (GE Healthcare, UK).

6.3 Gel electrophoresis

Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) was performed on pMMO and EDTA-treated pMMO according to Laemmli (Laemmli, 1970) at 200V at room temperature. Gels of 17 % acrylamide were used, with protein loadings of 12 – 165 μ g per well. Proteins bands were visualised using

Coomassie brilliant blue R-250 (0.1 % w/v in 10 % (v/v) methanol and 10 % (v/v) glacial acetic acid). The gel was destained by incubation in a solution of methanol, acetic acid and water (4:1:5 v/v).

Protein standards of myoglobin (equine, Sigma), alcohol dehydrogenase (ADH, yeast and equine, Sigma) were loaded on separate acidic native gels to be used as standards. Gels were prepared according to McLellan (Garfin, 2003). Replicate standard gels were dried onto filter paper (Whatman) using a vacuum system. Blank SDS gels were spiked with copper sulphate, zinc nitrate and ferric sulphate to be used as standards for laser ablation work (Binet *et al.*, 2003).

Blue native PAGE was also performed, only on non-treated pMMO, with an 8-13 % gradient gel and a protein loading of 48 – 144 µg per well. Samples were mixed with 0.5 M aminocaproic acid and loaded on the top of a 4 % stacking gel. The gel was run at 8 °C at 16 mA constant current and 70 – 120 V. When the samples had completely passed through the stacking gel, the voltage was gradually increased to 300 V. The 0.02 % Coomassie-containing cathode buffer was used until the dye front was one-third of the way from the top of the gel, when it was replaced with 0.002 % Coomassie-containing buffer. Following electrophoresis, gels were further destained in 25 % methanol and 10 % acetic acid.

6.4 Inorganic analysis

These experiments were performed under the supervision of Dr. Jeff Franks in the laboratory of the Measurement Science Group, ICI Wilton (now Intertek MSG).

6.4.1 Laser ablation

A neodymium:yttrium aluminium garnet (Nd:YAG) laser system with an output at the fundamental wavelength of 213 nm (UP213, New Wave, UK) was used to ablate gel samples. Ablated material was carried into the plasma of the ICP-MS instrument by helium gas flow. The laser was tuned on NIST 612 glass, which contains trace elements at 50 parts per million (ppm). Conditions were optimised, using blank gel regions, at nebuliser gas flow of 1 L/min, laser power at 50-75 % power and 10 Hz, with 5 µm depth pass and an ablation area of 105 µm. Calibration was performed using polypropylene discs containing known amount of the metals of interest,

namely copper, iron and zinc. Gels under study were dissected into slices of a suitable size for the ablation chamber and placed on a glass or plastic slide prior to analysis. Dried gels were affixed using double-sided tape; wet gels were placed without any adhesive.

6.4.2 Digestion of gel bands

Gel bands were cut out using a plastic knife to avoid any metal contamination. Each band was placed in a separate plastic tube and 0.2 ml of concentrated nitric acid added. These were left in a beaker of water at ~90 °C on a hotplate for 30 minutes. 0.05 ml of rhodium standard was added to give a concentration of 1 ppm, and each tube made up to 5 ml final volume with 4.75 ml water.

6.4.3 ICP-MS

An Elan DRC II (Perkin Elmer, USA) was used for all ICP-MS analysis with a concentric nebuliser. For in-solution analysis an auto-sampler system was used for sample introduction, with nebuliser gas optimised at 0.7 L/minute. Ablated material from LA and nebulised material from in-solution analysis was transported by argon as a carrier gas into the plasma. The instrument was operated in DRC mode, using 5 % hydrogen in argon pumped at 0.2 ml/min into the cell. Calibration in solution mode was performed using standard solutions of iron, copper and zinc at 5, 10, 20, 50 and 100, 200, 300 and 500 parts per billion (ppb) in 4% nitric acid with 1 ppm rhodium as internal standard. For digested gel samples, these were acid-matched calibration standards. For purified pMMO fractions, standards were prepared in 25 mM PIPES buffer. Multiple isotopes were monitored for each of the three metals of interest, to allow for monitoring of interferences and signal consistency. These were: ^{63}Cu , ^{65}Cu , ^{54}Fe , ^{56}Fe , ^{57}Fe , ^{64}Zn , ^{66}Zn and ^{67}Zn .

6.5 Hemoglobin sample preparation

Samples of fresh whole blood were supplied by University Hospitals Coventry and Warwickshire NHS Trust. Sample preparation for mass spectrometric analysis was

adapted from that detailed by Ofori-Acquah (Ofori-Acquah *et al.*, 2001). Samples (20 μ l) were diluted 10-fold in 10 mM ammonium acetate pH 6.8 and spun at 3000 g for 15 minutes in centrifugal filter units with a 10 kDa cut-off (Microcon[®] YM-10, Millipore Corporation, MA, USA). Sample retained on the filter was diluted a further 20-fold with 10 mM ammonium acetate and desalted by agitating for two ten-minute periods with approximately 5 mg of ion-exchange mixed bed resin (AG 501-X8, Bio-Rad Laboratories, Hercules, CA, USA) that had been prepared for use by rinsing twice in liquid chromatography MS grade water. The resulting solutions were introduced into the ESI source of a Synapt HDMS System (Waters, UK) by means of fused silica nanospray needles. All solvents and calibration and protein standards were obtained from Sigma (St. Louis, USA).

6.6 IMMS analysis

6.6.1 Data acquisition

Data were acquired by means of a Synapt HDMS System in ESI positive mode with a capillary voltage of 1.2 kV from 1000-4500 m/z . The time-of-flight (TOF) mass analyser was calibrated using 2 mg/ml cesium iodide in 50% aqueous propan-2-ol. Instrument acquisition parameters were adjusted to provide the optimal ion mobility separation. The cone voltage was 60 V and the collision energy in the trap region was 10 eV. Source temperature and gas flow were 110°C and 35 ml/min respectively. Nitrogen was used as the gas in the ion mobility cell and the indicated pressure within the cell was 0.68 mbar, equivalent to a flow rate of 38 ml/min. The backing pressure was increased in increments from 2 to 8 mbar to identify the ideal pressure conditions for transmission of the relevant ionic species. The travelling wave velocity and wave height were altered in increments from 100-600 m/s and 8-20V respectively, and the conditions that provided the optimal mobility separation were used for all following experiments.

The synchronisation of gated release of ions into the ion mobility separator with TOF acquisition allows arrival time distributions of ions to be obtained. Ions are released from the trapping cell to the mobility cell in a pulse every 100 μ s, and 200 orthogonal acceleration pushes of the TOF analyser are recorded to form one ion

mobility experiment. The overall mobility recording time is $200 \times t_p$, where t_p is the pusher period (Pringle *et al.*, 2007). The pusher period depends on the mass acquisition range; for these experiments, a pusher period of 120 μ s was used giving a mobility recording time of 24 ms. Equine myoglobin at a concentration of 10 μ M in 50% aqueous acetonitrile containing 0.2% formic acid was used to provide data which were used to create a calibration curve for cross-sectional measurements.

Data obtained for each hemoglobin tetramer over the m/z range 3000-4500 were deconvoluted on to a true mass scale using the MaxEnt software to provide an estimate of molecular mass. Experiments were carried out in triplicate.

6.6.2 Calibration, modelling and cross-section estimation

The equine myoglobin data was used to create a calibration curve for each set of experiments. Absolute cross-sections for equine myoglobin were obtained from drift-time ion mobility mass spectrometry (DTIMS) studies (Prof. Michael T. Bowers, personal communication). The calibration was performed using a procedure developed in-house based on previously published work (Ruotolo *et al.*, 2005; Scrivens *et al.*, 2006; Wildgoose *et al.*, 2006). In brief, normalised cross-sections (corrected for charge and reduced mass) were plotted against corrected arrival times (corrected to exclude time spent outside the ion mobility cell) to create a calibration with a power series fit. The calibration allows for estimation of the cross-section of a molecule of interest provided that the mobilities (corrected arrival times) for that molecule lie within the mobilities observed for the calibrant, irrespective of the size range of cross-sections for the calibrant (Shvartsburg and Smith, 2008; Thalassinos *et al.*, 2008). The calibration was used to estimate rotationally averaged collision cross-sections of hemoglobin monomer, dimer and tetramer for the different charge states observed based on their arrival time distributions, provided that their corrected arrival times fell along the calibration curve.

To compare the experimental cross-sections for the normal and sickle hemoglobin tetramers with accepted values, cross-sections were calculated using MOBCAL, a program to calculate mobilities (Mesleh *et al.*, 1996; Shvartsburg and Jarrold, 1996) from published X-ray structures held at the RCSB Protein Data Bank (Berman *et al.*, 2000). MOBCAL facilitates the use of three approximations to calculate cross

sections. The projection approximation (PA) typically results in an underestimation of the cross-section of a large ion. It calculates the cross-section by averaging the projections produced by every orientation of a molecule and so does not take into account interactions with the buffer gas. The trajectory method (TM) takes into account all interactions but is computationally intense. The exact hard sphere scattering model (EHSS) carries out trajectory calculations whilst ignoring long-range interactions but nevertheless gives values within a few percent of the TM approximations (Jarrold, 1999; Scarff *et al.*, 2008). For this work, cross-sections were calculated using the PA and EHSS methods to reduce computational time.

References

- Basu, P., Katterle, B., Andersson, K. K., Dalton, H.** (2003). The membrane-associated form of methane mono-oxygenase from *Methylococcus capsulatus* (Bath) is a copper/iron protein. *Biochem. J.* **369**: 417-427.
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., Bourne, P. E.** (2000). The Protein Data Bank. *Nucleic Acids Res.* **28**: 235-242.
- Binet, M. R. B., Ma, R., McLeod, C. W., Poole, R. K.** (2003). Detection and characterization of zinc- and cadmium-binding proteins in *Escherichia coli* by gel electrophoresis and laser ablation-inductively coupled plasma-mass spectrometry. *Anal. Biochem.* **318**: 30-38.
- Garfin, D. E.** 2003. Gel electrophoresis of proteins. In: Davey J and Lord M, eds. *Essential Cell Biology*. Oxford: Oxford University Press. 197-268.
- Jarrold, M. F.** (1999). Unfolding, Refolding, and Hydration of Proteins in the Gas Phase. *Acc. Chem. Res.* **32**: 360-367.
- Laemmli, U. K.** (1970). Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature* **227**: 680-685.
- Mesleh, M. F., Hunter, J. M., Shvartsburg, A. A., Schatz, G. C., Jarrold, M. F.** (1996). Structural information from ion mobility measurements: Effects of the long-range potential. *J. Phys. Chem.* **100**: 16082-16086.
- Myronova, N., Kitmitto, A., Collins, R. F., Miyaji, A., Dalton, H.** (2006). Three-Dimensional Structure Determination of a Protein Supercomplex That Oxidizes Methane to Formaldehyde in *Methylococcus capsulatus* (Bath). *Biochemistry* **45**: 11905-11914.
- Ofori-Acquah, S. F., Green, B. N., Davies, S. C., Nicolaides, K. H., Serjeant, G. R., Layton, D. M.** (2001). Mass spectral analysis of asymmetric hemoglobin hybrids: Demonstration of Hb FS ($\alpha_2\gamma\beta^S$) in sickle cell disease. *Anal. Biochem.* **298**: 76-82.
- Pringle, S. D., Giles, K., Wildgoose, J. L., Williams, J. P., Slade, S. E., Thalassinos, K., Bateman, R. H., Bowers, M. T., Scrivens, J. H.** (2007). An investigation of the mobility separation of some peptide and protein ions using a new hybrid quadrupole/travelling wave IMS/oa-ToF instrument. *Int. J. Mass spectrom.* **261**: 1-12.
- Ruotolo, B. T., Giles, K., Campuzano, I., Sandercock, A. M., Bateman, R. H., Robinson, C. V.** (2005). Evidence for Macromolecular Protein Rings in the Absence of Bulk Water. *Science*: 1658-1660.
- Scarff, C. A., Thalassinos, K., Hilton, G. R., Scrivens, J. H.** (2008). Travelling wave ion mobility mass spectrometry studies of protein structure: biological

significance and comparison with X-ray crystallography and nuclear magnetic resonance spectroscopy measurements. *Rapid Commun. Mass Spectrom.* **22**: 3297-3304.

Scrivens, J. H., Thalassinos, K., Hilton, G., Slade, S. E., Pinheiro, T. J. T., Bateman, R. H., Bowers, M. T. 2006. Use of a Travelling Wave-Based Ion Mobility Approach to Resolve Proteins of Varying Conformation. *Proc. 55th ASMS Conf. Mass Spectrometry and Allied Topics*. Indianapolis.

Shvartsburg, A. A., Jarrold, M. F. (1996). An exact hard-spheres scattering model for the mobilities of polyatomic ions. *Chem. Phys. Lett.* **261**: 86-91.

Shvartsburg, A. A., Smith, R. D. (2008). Fundamentals of Traveling Wave Ion Mobility Spectrometry. *Anal. Chem.* **80**: 9689-9699.

Thalassinos, K., Grabenauer, M., Slade, S. E., Hilton, G. R., Bowers, M. T., Scrivens, J. H. (2008). Characterization of Phosphorylated Peptides Using Traveling Wave-Based and Drift Cell Ion Mobility Mass Spectrometry. *Anal. Chem.* **81**: 248-254.

Wildgoose, J. L., Giles, K., Pringle, S. D., Koeniger, S. J., Valentine, R. H., Bateman, R. H., Clemmer, D. E. 2006. A comparison of travelling wave and drift tube ion mobility separations. *Proc. 54th ASMS Conf. Mass Spectrometry and Allied Topics*. Seattle.

Chapter 7

Results

Metallomic Studies

7.1 Analysis of gel-resolved proteins by LA-ICP-MS

7.1.1 Technical aspects of the laser ablation system

The laser system used for the LA-ICP-MS work was a newly-released instrument provided by New Wave (Cambs, UK) for assessment. One particular advantage was the use of helium, rather than the traditional argon, as a carrier gas in the ablation cell, as ablation in helium results in smaller particles produced at the ablation site. Helium is less dense than argon, has a higher thermal conductivity and can remove energy away from the ablation site faster. It has a higher ionisation energy than argon resulting in a smaller plasma above the surface of the sample (Hergenroder, 2006). These particles are transported and processed by the plasma more efficiently, and are not deposited back on the sample. Although it was superior to the previously available infrared laser, a number of technical issues were encountered. Visualising the gels using the built-in camera system was difficult, with a lack of sufficient contrast between the background and the protein bands, and limited control of the light settings which could assist with this issue. An example of the imaging is shown in Figure 7.1. Due to the size of the visualisation area, only a small amount of the gel could be seen at any one time. In order to obtain the view required for analysis of protein bands, a visualisation map had to be created which could take up to ten minutes. Whilst not a problem when looking at dried gels, it was a significant drawback when studying wet gels. The heat in the ablation chamber would often cause wet gels to dry and distort before ablation could commence. An additional problem was encountered when opening the chamber to change samples. This process would cause the inductively coupled plasma (ICP) to extinguish; re-igniting this must be a gradual procedure to prevent the plasma extinguishing again, therefore slowing the throughput of samples. Although this was partially overcome by careful control of the nebuliser gas flow, it was not completely solved.

7.1.2 Analysis of dried gels

Laser ablation (LA) of dried gels was performed on gels containing the standard proteins of myoglobin and alcohol dehydrogenase (ADH). Images of the ablated gels can be seen in Figure 7.2. Although the gels appeared to hold their structural

integrity well during the drying process, they soon suffered the effects of transportation and handling, as is evident in the cracks highlighted. Blank areas of gels were initially analysed to provide background readings to which data from the protein bands could be compared, but the amount of ablated material proved insufficient to provide necessary signal when transferred to the mass spectrometer. The laser energy and ablation area were both increased to try and overcome this; the result, however, was the scorching of the gel right through to the backing filter paper. It was decided to attempt analysis of wet gels in consideration of the laboratory and instrument time available.

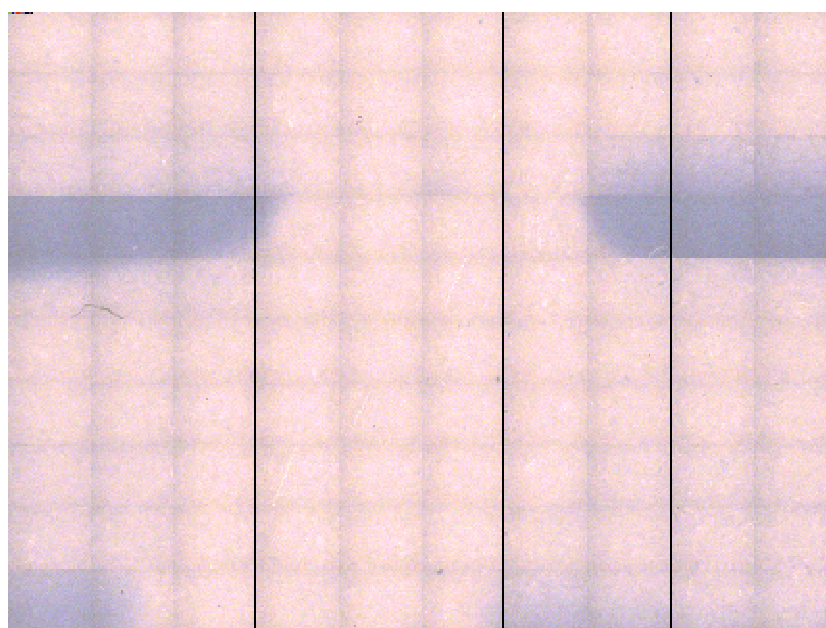


Figure 7.1 – An example of gel visualisation in the laser ablation chamber

7.1.3 Analysis of wet gels

Blank gels spiked with metal salts were prepared to be used for calibrating background readings (Binet *et al.*, 2003). Gels containing copper sulphate did not polymerise, but gels containing iron and zinc were used. Despite the technical difficulties described above, some data was obtained. The monitoring of multiple isotopes allowed for the monitoring of interferences, as specific isobaric interferences only affect particular isotopes of certain elements, e.g. $^{40}\text{Ar}^{16}\text{O}$ to ^{56}Fe .

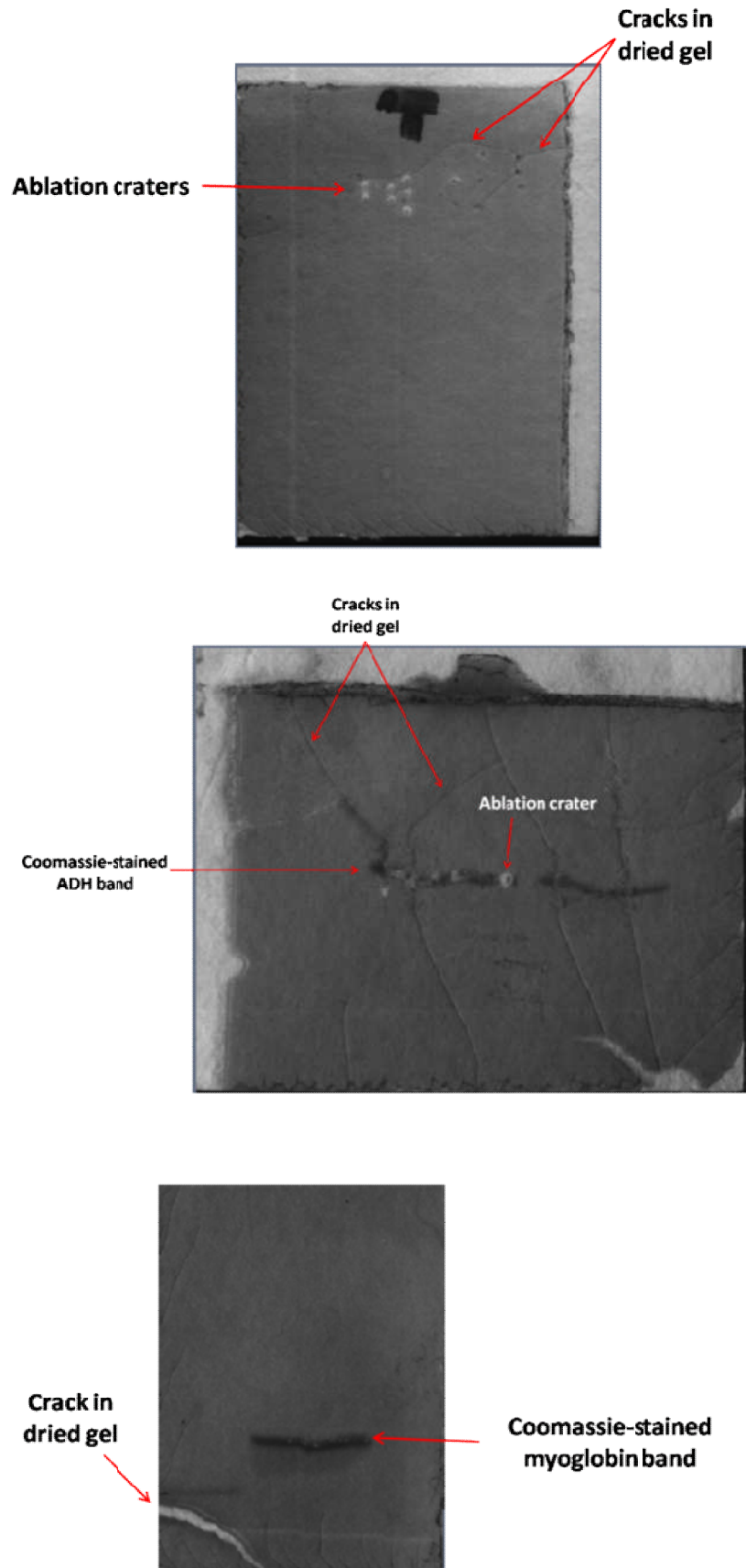


Figure 7.2 – Laser ablation of dried polyacrylamide gels

For the iron-containing gels, ^{53}Fe and ^{56}Fe were monitored, along with ^{31}S , ^{32}S , and ^{33}S and ^{13}C , as sulphur and carbon are components of the gel. For the zinc-containing gels, ^{63}Zn , ^{65}Zn and ^{66}Zn were monitored. A region of the gel was ablated over a period of 60 seconds; the aim was to generate a stable reading for the metal ions, so that significant readings in protein bands could be confirmed. As shown in Figure 7.3, stable readings could not be obtained from the spiked gels, and they were therefore not suitable to be used to calibrate the LA-ICP-MS system. Following calibration using polypropylene discs spiked with known concentrations of metal, gel bands from ADH and pMMO were analysed. Examples of data obtained are given in Figure 7.4. As with the spiked gels, stable readings could not be obtained. Significant readings for zinc would be expected for ADH, but these are dwarfed by background signals from carbon and sulphur. The pMMO analysis was also inconclusive.

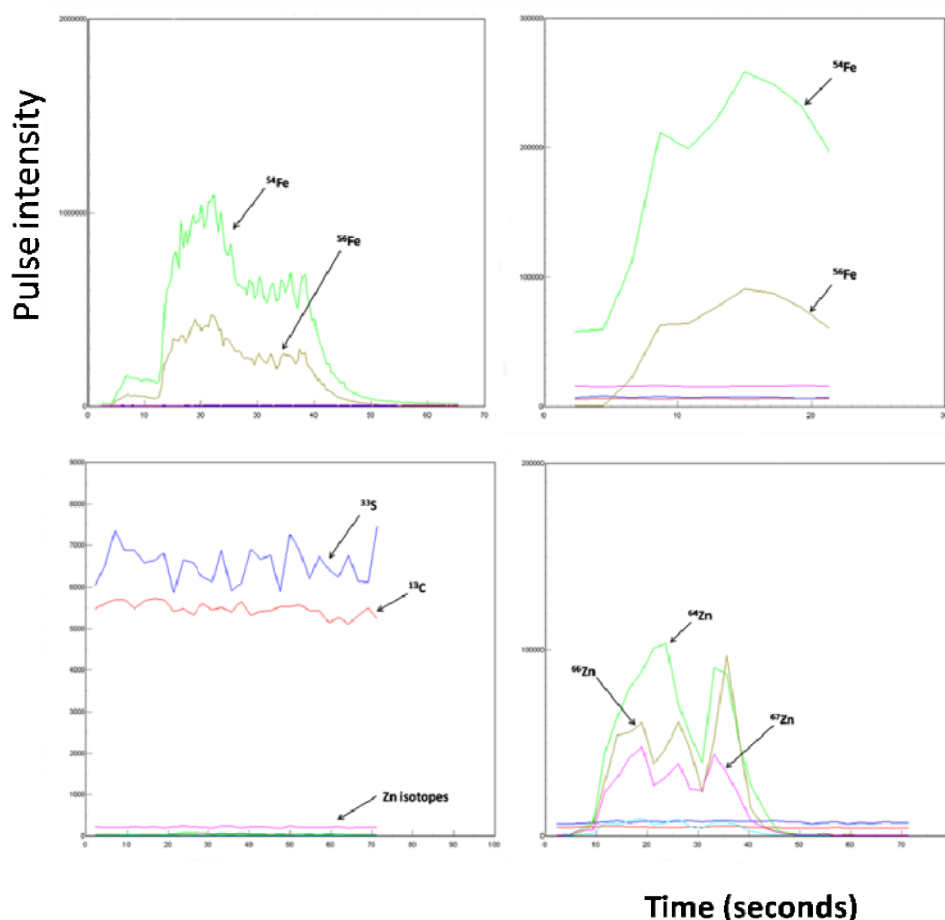


Figure 7.3 – LA-ICP-MS analysis of gels spiked with iron and zinc.

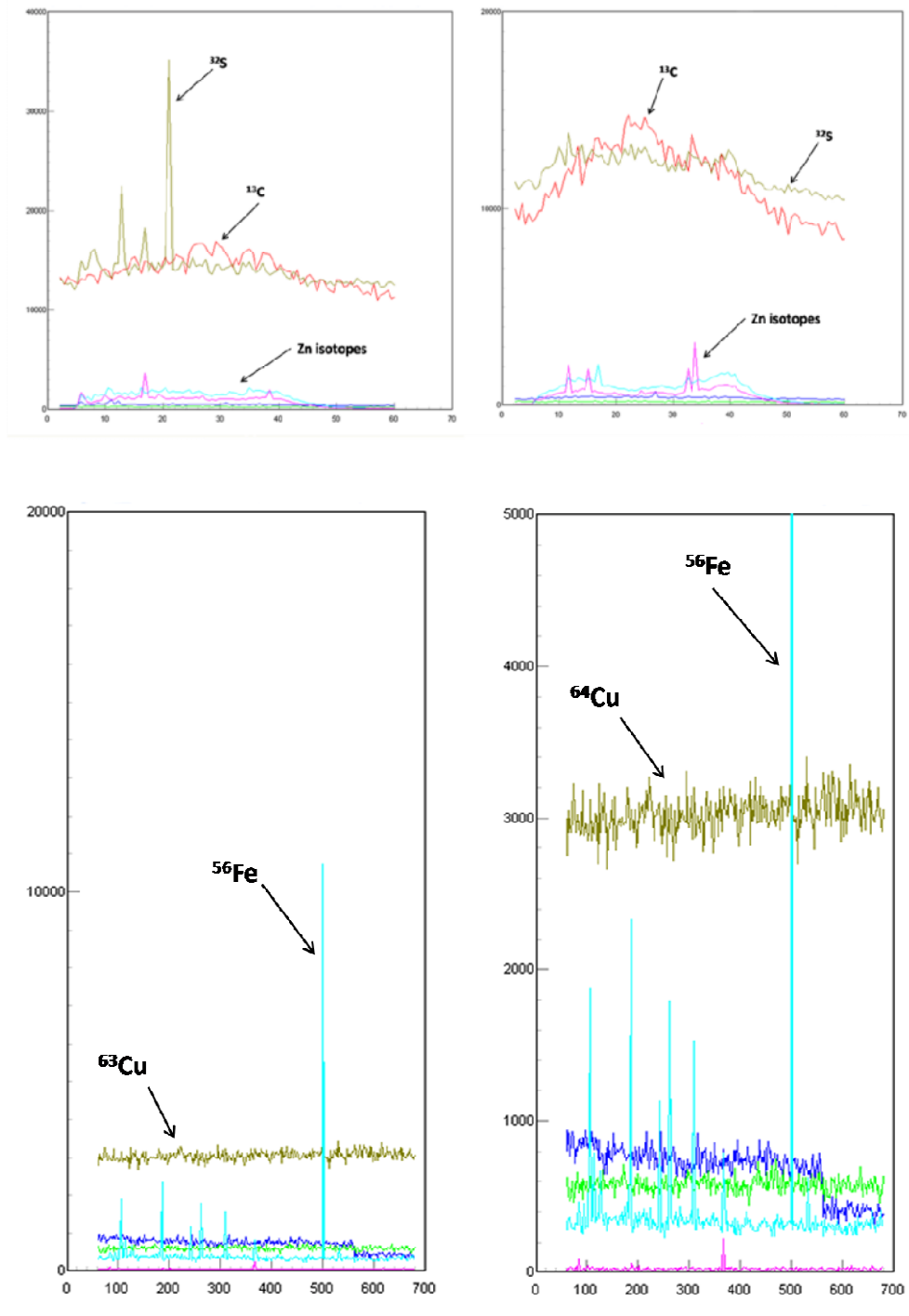


Figure 7.4 – Analysis of protein bands in wet gels; the upper panel shows readings from ADH, the lower from pMMO.

7.2 Analysis of gel-resolved proteins by digestion and in-solution ICP-MS

An alternative approach employed to obtain results from gel-resolved samples was to digest the gel bands and detect any released metals. An example of an SDS-PAGE separation of pMMO can be seen in Figure 7.5, with the constituent subunits labelled. These bands were digested and analysed for levels of copper, iron and zinc. Although there is agreement that the zinc observed in the structure of the pMMO complex is an artefact of the crystallisation process, it was monitored as a control. As shown in Figure 7.6(a), a linear response was achieved for increasing concentrations of the metals for multiple isotope calibration, when normalised to the internal standard of rhodium. The isotopes ^{63}Cu , ^{65}Cu , ^{54}Fe , ^{56}Fe , ^{57}Fe , ^{66}Zn , ^{64}Zn and ^{67}Zn were monitored. Readings for these metals have been summarised in Figure 7.6(b)-(d), and show that none of the elements appeared at any significant levels, remaining in the very low ppb range. When compared with samples from EDTA-treated pMMO, results from which are summarised in Figure 7.7, there is no difference; in fact, there is a slightly higher metal content in the treated samples, which are supposedly metal-depleted.

A fresh preparation of pMMO was analysed several weeks later in order to provide repeat data which could shed light on the unexpected results obtained. The calibration and summarised results are shown in Figure 7.8, indicating very similar results. It is possible that the SDS-PAGE caused the pMMO to denature such that the metal ions were lost, as PAGE analysis is problematic for membrane proteins (Jiménez *et al.*, 2009; Rath *et al.*, 2009). A blue native (BN) gel was therefore analysed; as the name suggests, this allows for more native conformation to be retained. Two closely positioned bands are obtained from a BN-PAGE separation of pMMO, as shown in Figure 7.9. These bands were digested and analysed, but again, no significant levels of any of the metals were detected. These results are summarised in Figure 7.10.

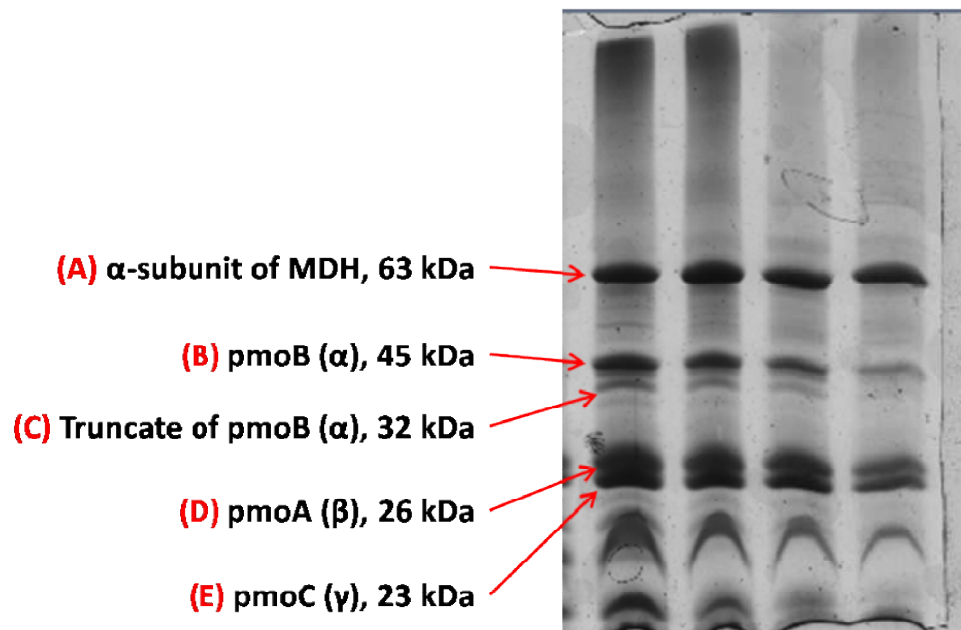


Figure 7.5 – SDS-PAGE of purified pMMO. Band identifications correspond to previous characterisation (Smith and Dalton, 1989).

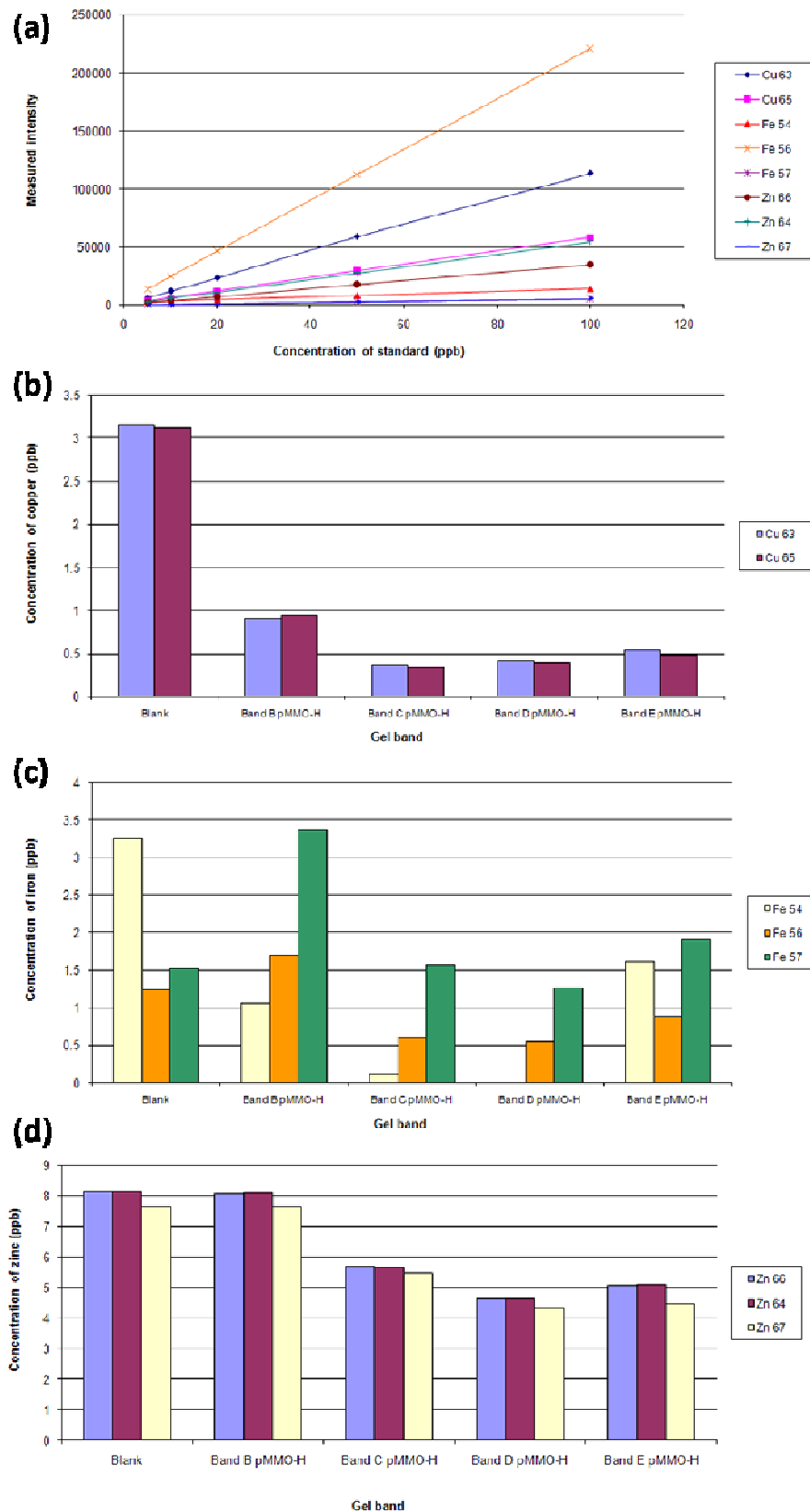
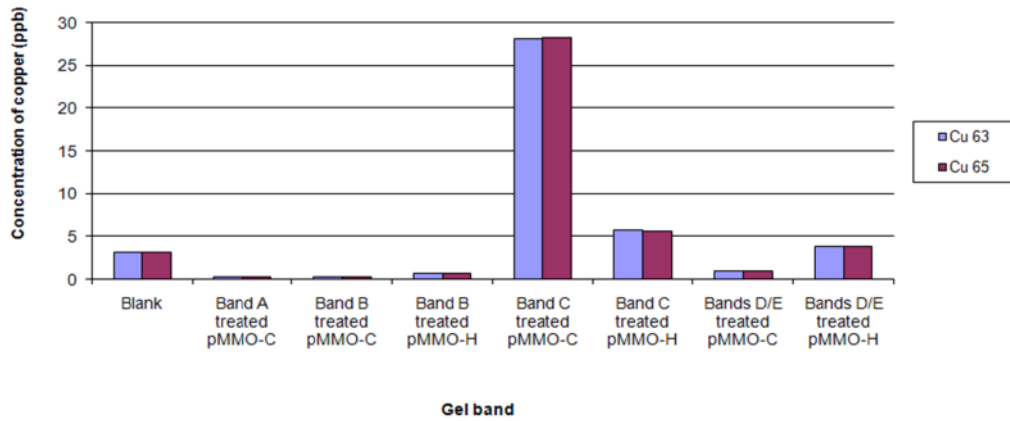
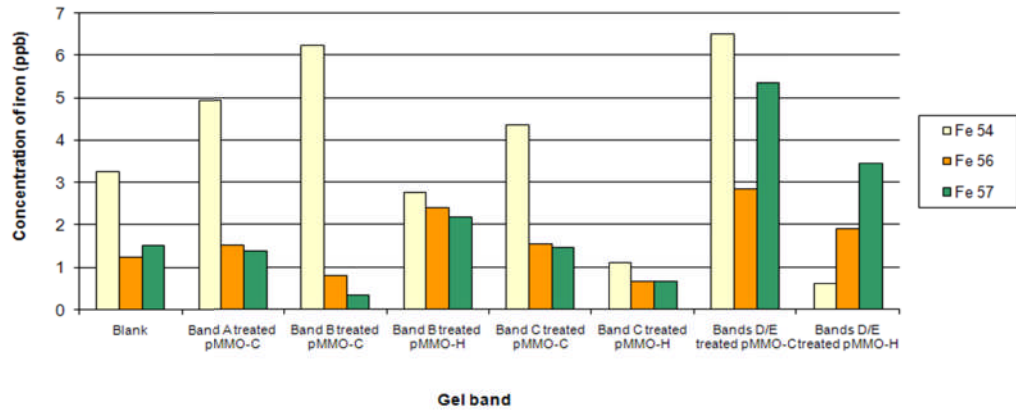


Figure 7.6 – Analysis of digested gel bands from SDS-PAGE of purified pMMO, (a) calibration curves for multiple isotopes of copper, iron and zinc, (b) copper concentrations, (c) zinc concentrations. None were at significant levels.

(a)



(b)



(c)

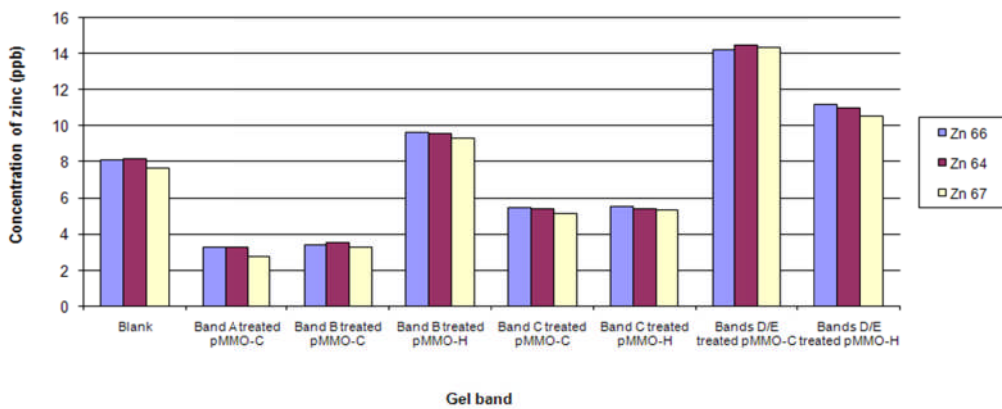


Figure 7.7 - Analysis of digested gel bands from SDS-PAGE of pMMO, treated with EDTA to deplete any naturally occurring metal, (a) copper concentrations, (b) iron concentrations, (c) zinc concentrations. None were at significant levels.

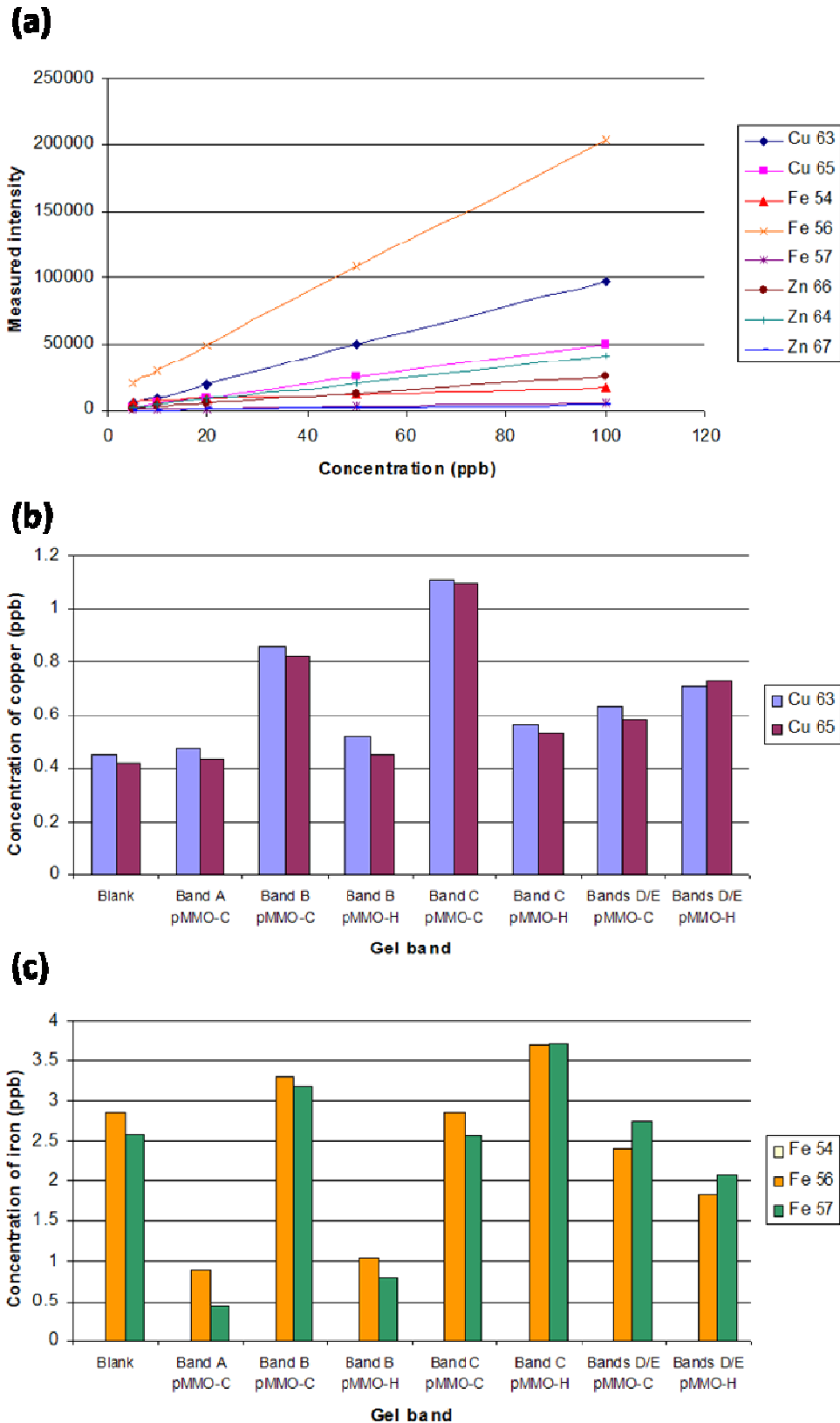


Figure 7.8 – Repeat analysis of digested gel bands from SDS-PAGE of purified pMMO, (a) calibration curve, (b) copper content, (c) iron content. None were at significant levels.

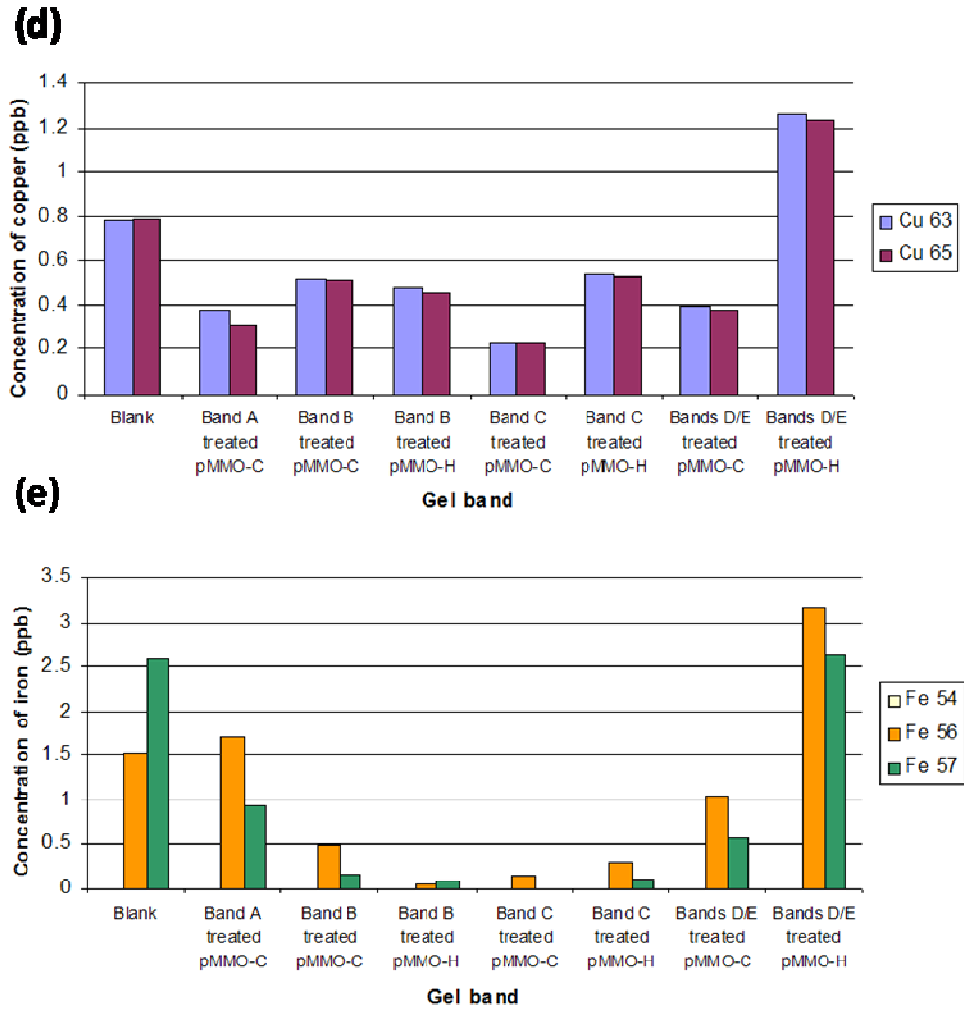


Figure 7.8 – Repeat analysis of digested gel bands from SDS-PAGE of purified pMMO, (d) copper content in EDTA-treated sample, (e) iron content in EDTA-treated sample. None were at significant levels.

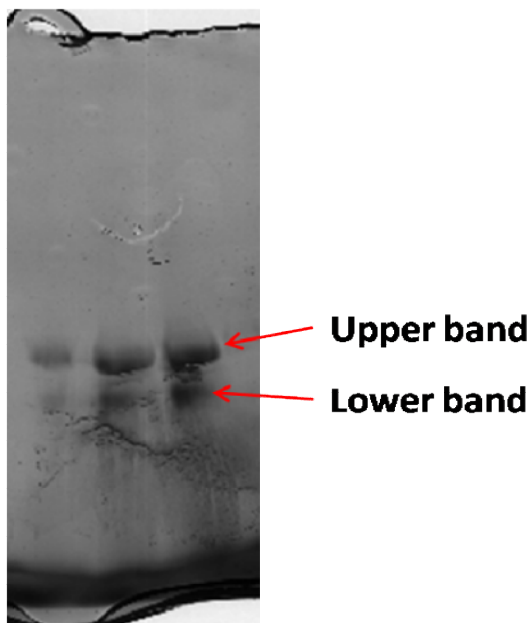


Figure 7.9 – Blue Native gel of purified pMMO

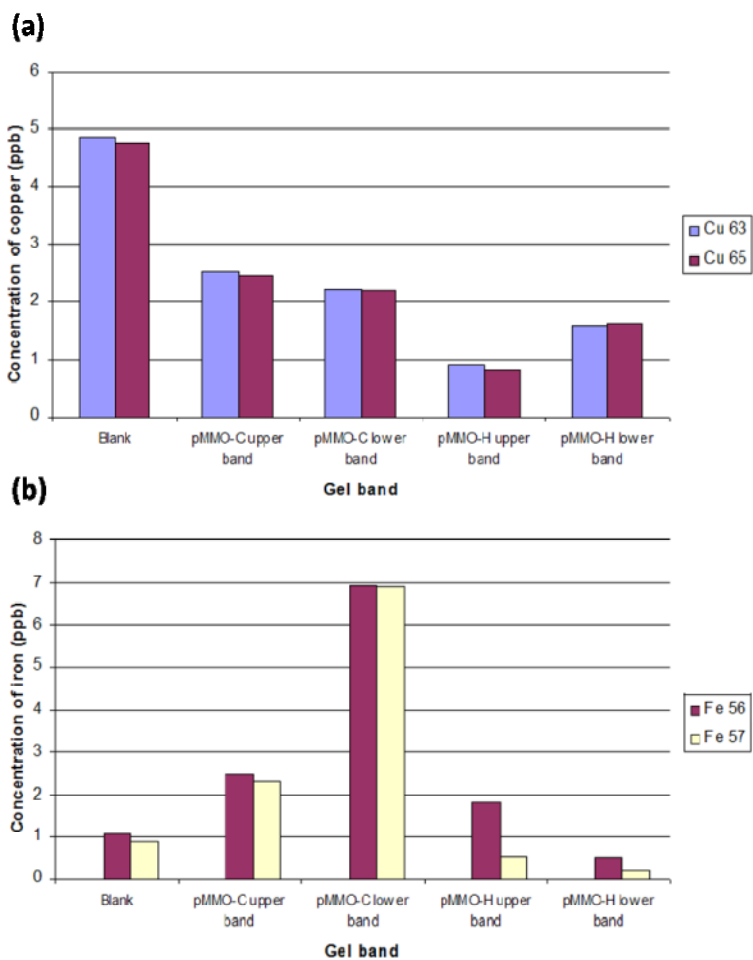


Figure 7.10 – Analysis of digested bands from blue native PAGE of purified pMMO, (a) copper concentrations, (b) iron concentrations. Neither were at significant levels.

7.3 In-solution analysis of pMMO by ICP-MS

The use of in-solution ICP-MS for the analysis of biological samples is often limited by the sample requirements, as volumes needed often reduce the concentration of the elements of interest to outside the range of sensitivity of the instrumentation. Interferences are also encountered from the buffers required to keep samples in solution. Fractions from the chromatographic separation of pMMO were analysed in the smallest possible volume to allow for replicate analyses (5 ml), and the calibration was performed with the use of 1 ppb rhodium as an internal standard to normalise all measurements against. Five fractions (I – V) were used; these correspond to elution from the column and do not necessarily correspond to a distinct separation of pMMO subunits. Half of each was treated with EDTA to deplete metal. As previously, isotopes ^{63}Cu , ^{65}Cu , ^{54}Fe , ^{56}Fe , ^{57}Fe , ^{66}Zn , ^{64}Zn and ^{67}Zn were monitored. The results are summarised in Figure 7.11 for copper and iron data; the calibration shows a good linear response, as before.

Levels of metal recorded are much higher than the gel analyses, over 100 ppb, suggesting that these readings indicate genuine detection rather than being experimental artefacts or background signals. The somewhat surprising observation is that the levels of metal appear similar or, in the case of iron, significantly higher in the EDTA-treated samples compared to the non-treated. As this has been consistently observed across analysis approaches and a number of replicate experiments it would indicate a problem with the samples. A possible explanation is that some sort of contamination is present in the column used to filter the treated samples, but confirmation of this would require further investigation.

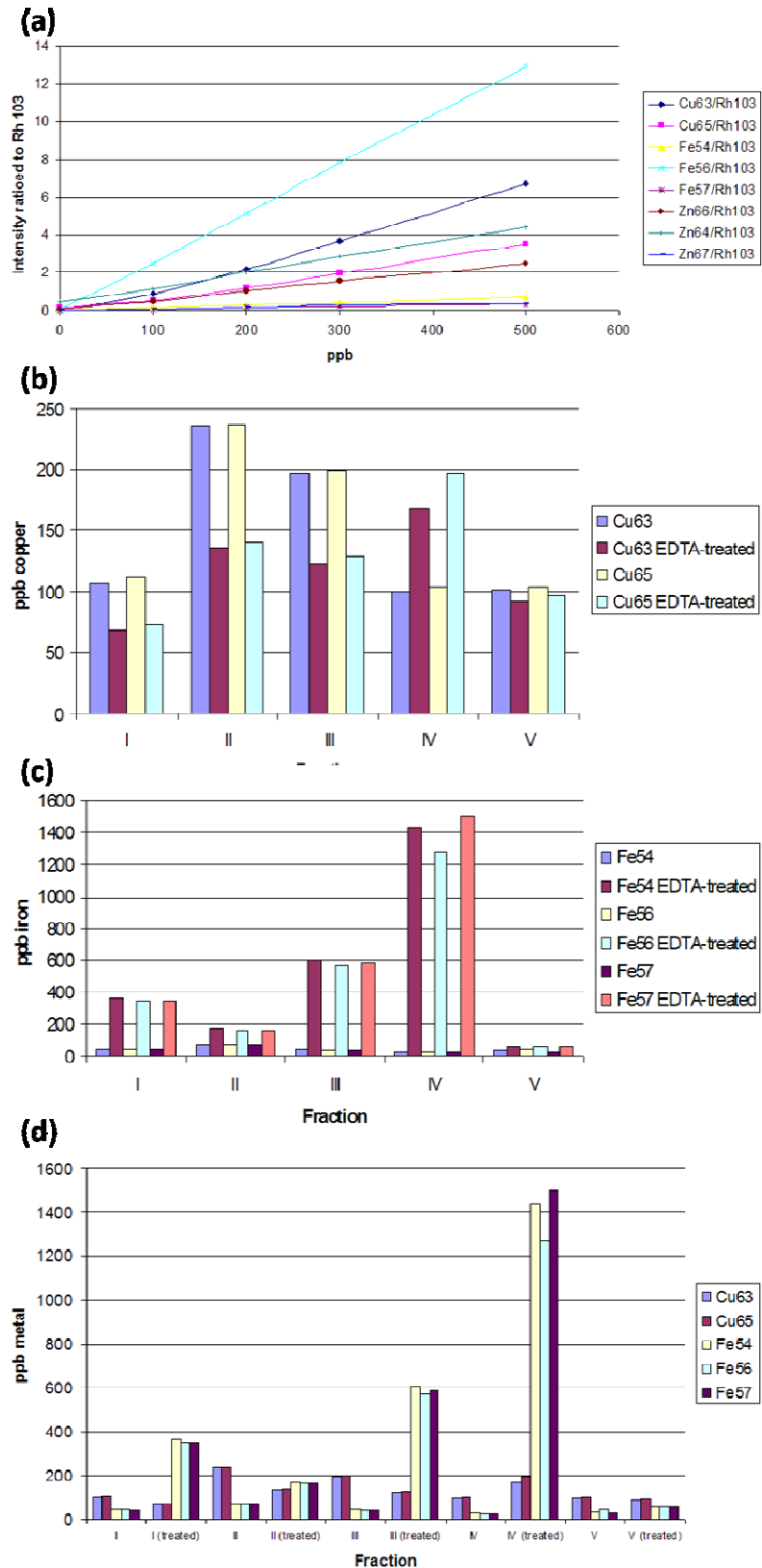


Figure 7.11 – In-solution ICP-MS analysis of pMMO, (a) calibration using known concentrations of copper, iron and zinc, (b) copper concentrations in treated and untreated samples, (c) iron concentrations in treated and untreated samples, (d) combined data for untreated samples showing considerable metal levels.

If the EDTA-treated samples are removed from the dataset, copper appears to be present at levels between 100 and 250 ppb across the fractions, and iron at levels between 25 and 60 ppb. As shown in Figure 7.12(a), Fraction II shows the highest levels of both metals, with Fraction III at slightly lower levels, and Fractions I, IV and V at comparably lower levels. The isotope levels appear almost identical, an observation made clearer by plotting the abundances of the separate isotopes present in each fraction, as in Figure 7.12(b). The data-handling software for the ICP-MS instrument accounts for natural abundances of the isotopes, shown for copper and iron in Table 7.1, and therefore each isotope should be observed at the same level. This adds confidence to the data obtained on metal content. These results appear to correlate with previous work showing presence of copper in the pMMO complex, and seem to be in agreement with the work of those groups proposing that iron is the metal in the third metal site. In order to confirm the results obtained, ICP optical emission spectrometry (ICP-OES) was performed. This technique is less sensitive, and therefore less precise than ICP-MS at very low concentrations. It can be used to monitor multiple emission lines rather than multiple isotopes. In this case, results obtained showed similar trends in metal concentration, as shown in Figure 7.13. The inorganic data obtained are not significant enough to confidently comment on any biological context.

To place any significant data in context of the overall structure of pMMO, the protein content of the fractions would need to be determined, in order to correlate the organic and inorganic structural aspects (Becker, J. Sabine *et al.*, 2005). This is problematic, however, due to the membrane-bound nature of the complex. Membrane proteins are notoriously difficult to analyse by organic mass spectrometry, as the detergents used to extract and purify them are incompatible with the ionisation methods employed (Santoni *et al.*, 2000). Some efforts were made with the available pMMO sample using methods developed at the MRC (Carroll *et al.*, 2007), Dr. Ian M. Fearnley, personal communication) but with little success. The sample availability of pMMO is a common problem in any analysis, as there has been no expression system developed to date. Suitable methanotrophs must therefore be cultivated, and naturally occurring protein extracted and purified.

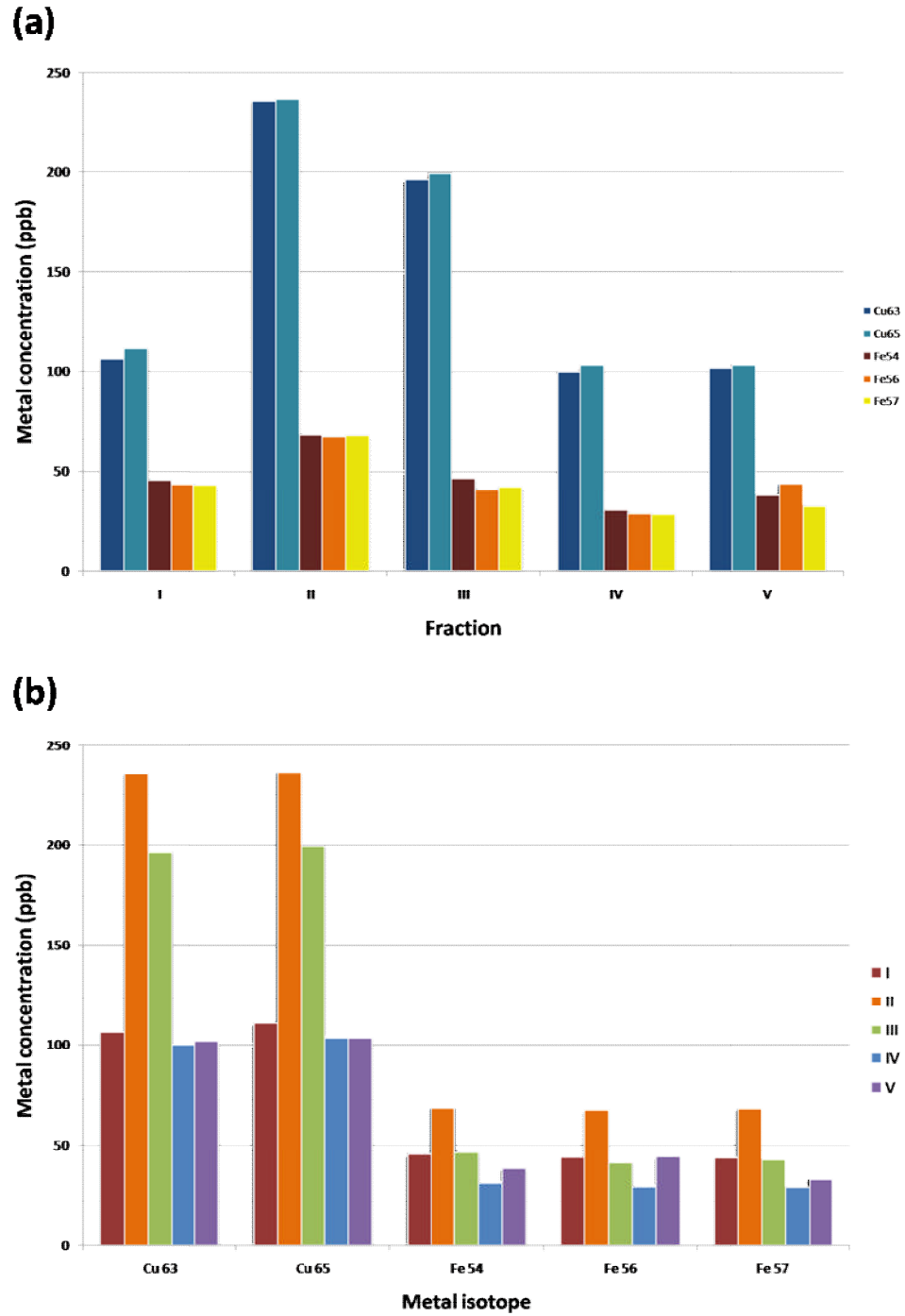


Figure 7.12 – In-solution ICP-MS of untreated pMMO, (a) concentrations of metal in each fraction, (b) concentration of individual isotopes in the five fractions showing good reproducibility.

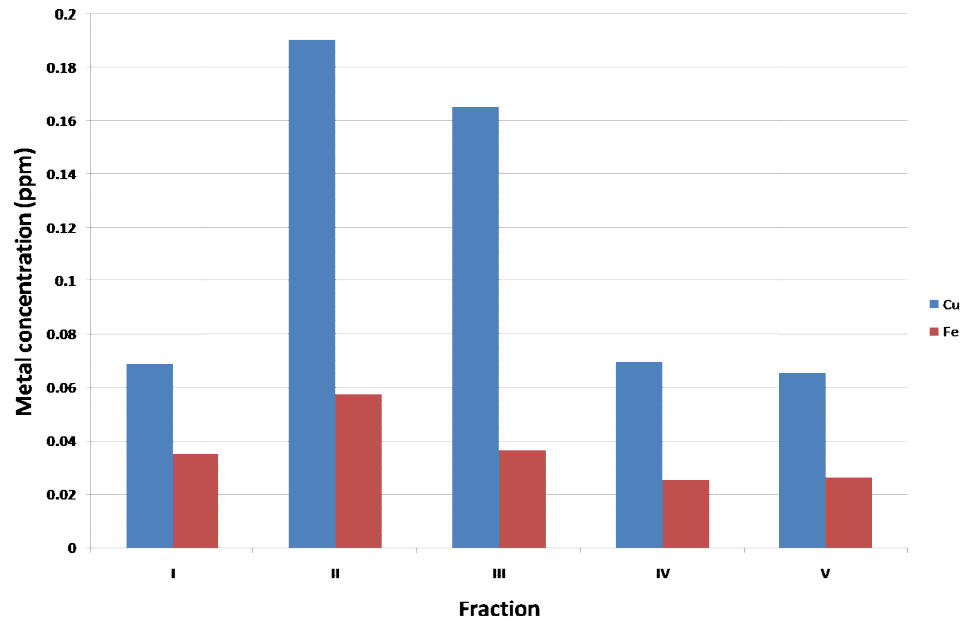


Figure 7.13 – ICP-OES analysis of untreated pMMO.

Isotope	Abundance
⁶³ Cu	69.17 %
⁶⁵ Cu	30.83 %
⁵⁴ Fe	5.8 %
⁵⁶ Fe	91.72 %
⁵⁷ Fe	2.2 %

Table 7.1 – Natural abundances of copper and iron isotopes.

7.4 Ion mobility studies of human hemoglobin

7.4.1 Instrument acquisition parameters and calibration

The instrument acquisition parameters must be optimised for each individual application of ion mobility separation. These must be tailored to each sample of interest, as optimal conditions are dependent on ionic species and mass-to-charge ratios (Tahallah *et al.*, 2001). Controlled optimisation indicated that a backing pressure of between 6.6 and 6.7 mbar was ideal for intact hemoglobin tetramer analysis. The optimal ion mobility separation of the tetramer was achieved at a travelling-wave velocity of 400 m/s and wave height of 18 V.

A calibration curve was used to allow the estimation of cross-sections for different hemoglobin molecules at different charge states. Cross sections calculated for equine myoglobin were within 2 % of absolute values obtained by drift cell experiments, seen reproducibly across the three datasets acquired. A curve plotted for one calibration dataset is shown in Figure 7.14. Arrival time at the detector is the measured parameter, which can then be related to cross-section as previously described. A correlation between arrival times on our instrument and published cross-sections can be seen.

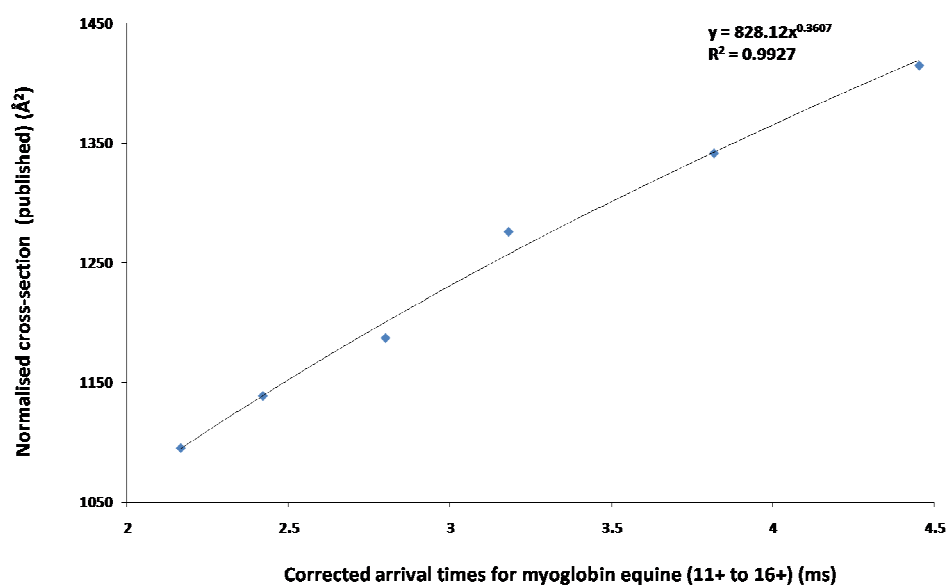


Figure 7.14 – Calibration curve for estimation of cross sections.

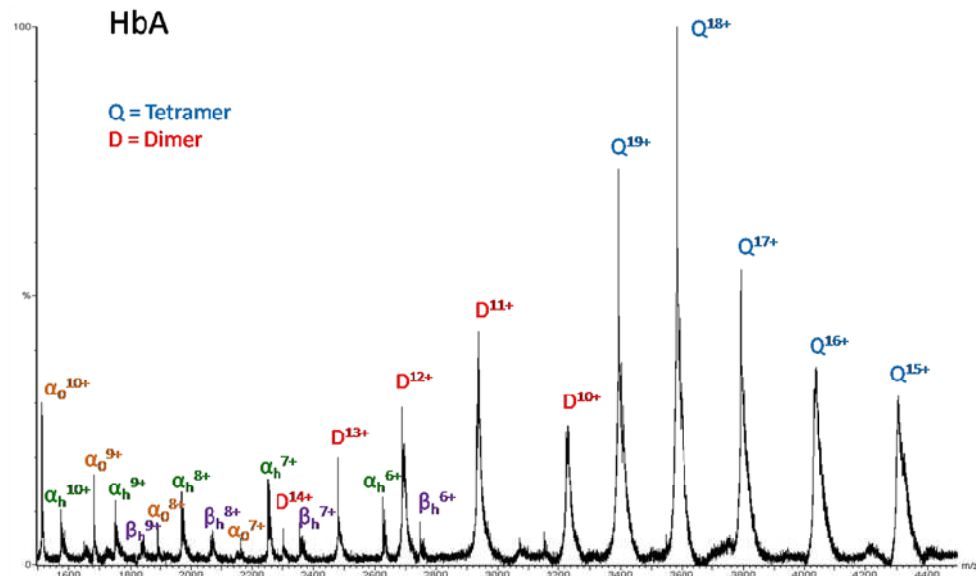
7.4.2 Hemoglobin tetramer analysis

Data for normal (HbA) and sickle (HbS) hemoglobin were analysed under non-denaturing conditions; representative spectra are shown in Figure 7.15. The data were deconvoluted to give masses of 64,454.7 Da for HbA and 64,385.8 Da for HbS, which are very close to the theoretical masses of 64,453.2 and 64,393.4, respectively (Ofori-Acquah *et al.*, 2001).

The spectra show the presence of the tetramer ($(\alpha^h\beta^h)_2$), heterodimer ($\alpha^h\beta^h$), and apo- and holo-monomer species. The trimer is not seen, as would be expected considering that the tetramer involves the non-covalent association of two $\alpha^h\beta^h$ dimers. As clinical samples were used, and carefully controlled near-physiological conditions were employed in the preparation, the absence of trimer implies that the species observed exist naturally in solution. This is consistent with results from isotope-labelling studies, which showed that non-tetrameric ions seen in the spectrum correspond to genuine in-solution species (Hossain and Konermann, 2006) rather than the products of fragmentation (Kuprowski *et al.*, 2007).

Both types of monomer (α and β) are observed in the HbA spectrum in both apo and holo form. A previous study (Griffith and Kaltashov, 2003) suggested that an α^h monomer becomes associated with a β^a monomer, to enable the β -chain to incorporate its heme group. This was based upon the fact that no β^h was observed in the spectrum. A subsequent study (Boys *et al.*, 2007) detected very small quantities of heme-deficient dimer, and found that both monomers were capable of binding heme. The discrepancies between these two studies are thought to be attributed to differences between the commercially prepared and freshly obtained samples used. In this study, which used fresh blood samples, β^h was observed in multiple charge states.

(a)



(b)

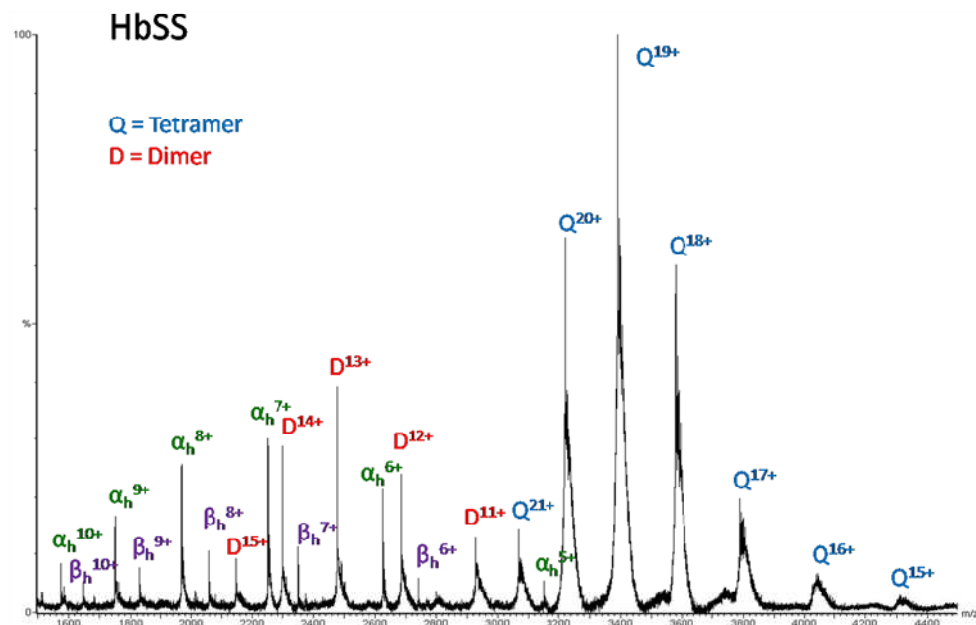


Figure 7.15 – Mass spectra of (a) normal (HbA) and (b) sickle (HbSS) hemoglobin analysed by ESI-TOF-MS. Charge states are labelled for the tetramer (Q), heterodimer (D) and apo- and mono-monomers; subscripts a and h refer to apo and holo forms, respectively.

7.4.3 Cross-section determination

It has been reported that without the attachment of a heme group, the α and β monomers adopt extensively unfolded conformations (Leutzinger and Beychok, 1981). Cross-sections for various charge states of the monomers in both apo and holo forms were estimated. These suggest that the predominant conformations of both monomers in the gas phase are similar to each other and show little change in the presence of heme, as shown in Figure 7.16. The cross-section of each of the molecules increases with an increase in charge, thought to be a result of the effects of Coulomb repulsion (Badman *et al.*, 2001; Valentine *et al.*, 1997).

The heme-deficient dimer observed in previous studies was not observed. The presence of both apo and holo forms of the monomers suggests that a β^a does not need to associate with α^h in order to recruit a heme group. Many more charge states of α^a were observed than β^a . The number of charges accepted by a protein depends upon the number of exposed basic sites on its surface; a more folded protein has fewer sites exposed than an unfolded one, and cannot accept as many charges. This may suggest, therefore, that the α -chain adopts more unfolded conformations in the gas phase than the β -chain. Alternatively, the absence of higher charge states of β^a may be due to the different desolvation behaviour of the two types of monomer. Due to its greater non-polar character, the α -chain ionises preferentially compared to the β -chain, thereby competing more effectively for charge (Kuprowski *et al.*, 2007).

The cross-section estimations of monomer, dimer and tetramer can be put together to form a picture of the hemoglobin assembly process, as shown in Figure 7.17. The 6+ charge state of α^h and β^h ($[M+6H]^{6+}$) have estimated cross-sections of 1583 and 1488 \AA^2 , respectively. If these two monomers were to come together to form a dimer, it would be expected that a dimer with a cross-section of approximately the sum of these values, i.e. around 3071 \AA^2 , would be observed. This is the case, as the ($[M+12H]^{12+}$) dimer charge state has an estimated cross-section of 3001 \AA^2 . The slightly smaller value is not unexpected, as the contact area on both of the monomers would be compacted and contribute less to the overall cross-section.

The average estimated cross-sections for the HbA and HbS, for four different charge states, are shown in Figure 7.18. The data indicate a difference in cross-section between normal and sickle-cell hemoglobin, and a variation with charge state.

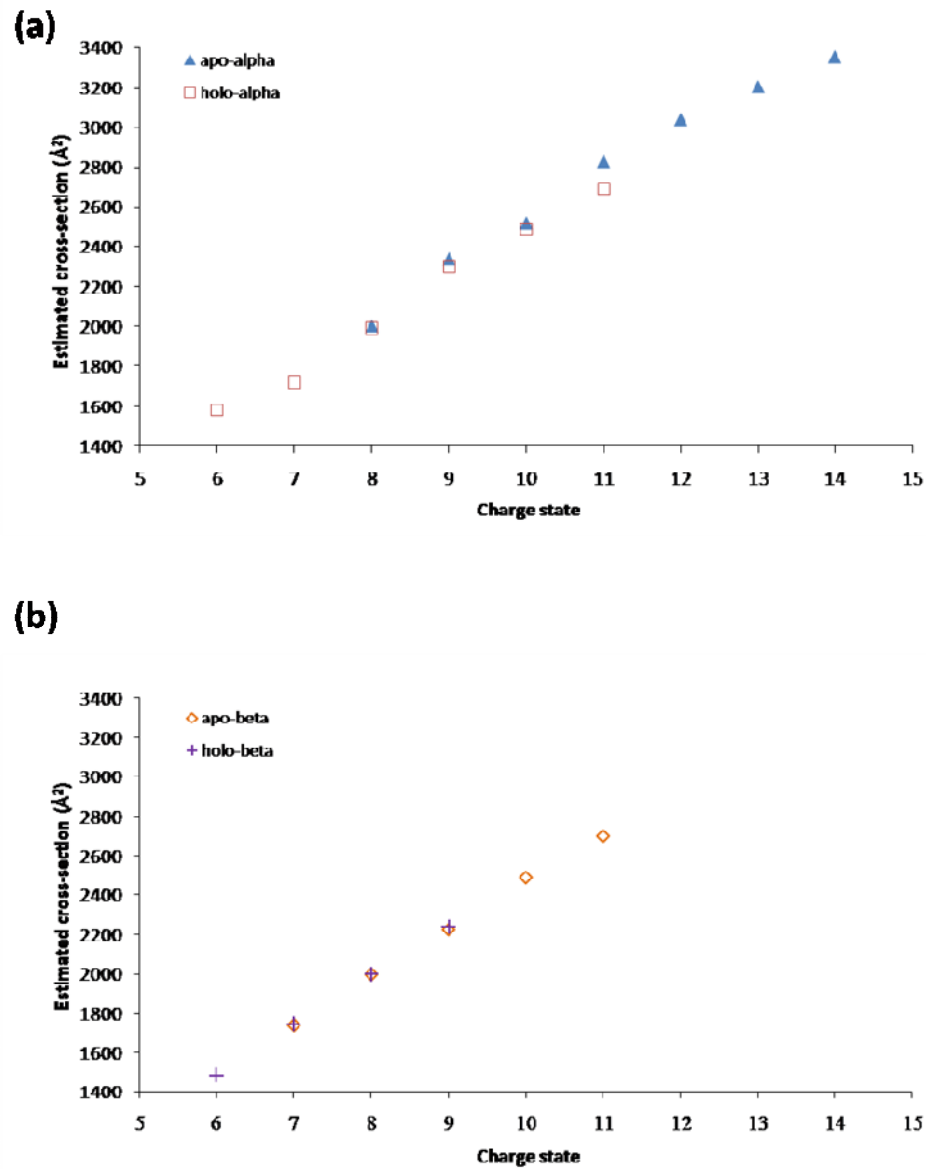


Figure 7.16 – Average estimated cross-sections for (a) the α -monomer, and (b) the β -monomer, observed across three datasets.

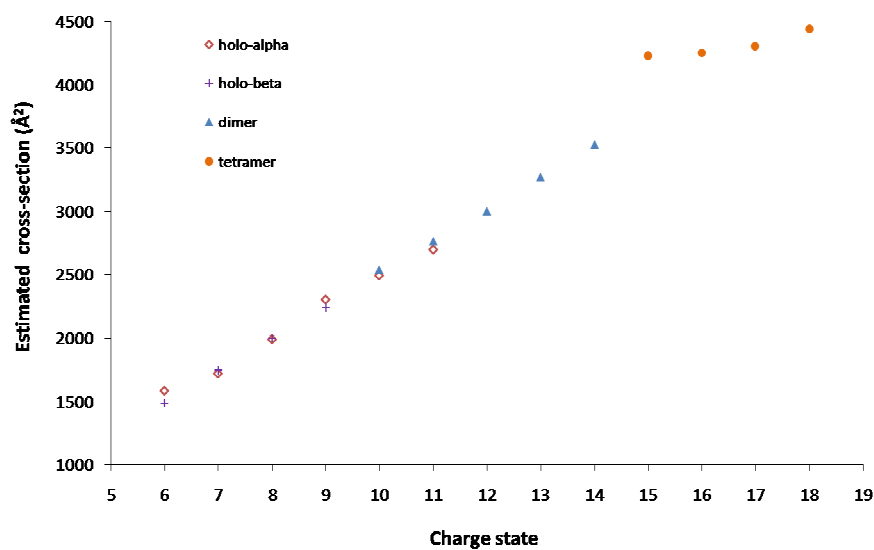


Figure 7.17 – Average estimated cross-sections for holo- α , holo- β , heterodimer and HbA tetramer, using data obtained across three datasets.

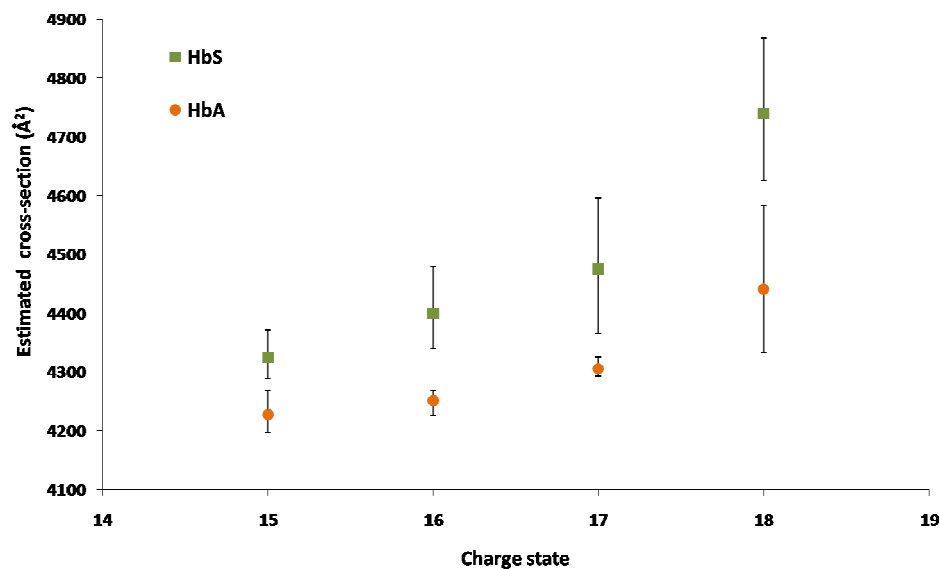


Figure 7.18 – Estimated cross-sections for HbA and HbS tetramers for four different charge states, showing averaged values from three datasets, with corresponding errors.

For the charge states studied, the cross-sections observed for HbS are somewhat larger than those of HbA. Theoretically determining what the charge state of a molecule should be, within a particular solvent at a certain pH, is difficult, as secondary, tertiary and quaternary structure have to be taken into consideration. Previous work indicates that the lowest charge states observed under near-physiological conditions are most representative of the native protein structure (Scarff *et al.*, 2008). With this in mind, the $[M+18H]^{18+}$ charge state for HbA and HbS may correspond to a tetrameric structure that is beginning to denature. The reproducibility of the cross-sections estimated for $[M+15H]^{15+}$, $[M+16H]^{16+}$ and $[M+17H]^{17+}$ charge states of HbA is $\pm 1\%$ between three replicate datasets; this is believed to be representative of the reproducibility capabilities of the experiment. The cross-sections for HbS show a larger deviation of $\pm 3\%$, which may reflect the presence of a more diverse population of conformations of the HbS molecule with similar cross-sections.

The rotationally averaged cross-sections for HbA calculated from X-ray crystallographic structures were 3133 and 4343 \AA^2 for projection approximation (PA) and exact hard sphere scattering (EHSS), respectively. The values for HbS were 3733 and 4775 \AA^2 . The experimentally obtained estimations for the $[M+15H]^{15+}$, $[M+16H]^{16+}$ and $[M+17H]^{17+}$ charge states of HbA and HbS fall between these two theoretical approximations, and agree with the X-ray observation that HbS has a larger cross-section than that of HbA. Although illustrative of solution-phase structures under controlled conditions, gas-phase conformations have been shown to be smaller than those predicted by EHSS approximations (Hoaglund-Hyzer *et al.*, 1999). A more compact conformation is thought to be adopted in the gas phase due to intramolecular interactions causing the collapse of polar side chains onto the surface of the protein (Shelimov *et al.*, 1997).

Native protein structure during and following transfer into the gas phase has been the subject of much discussion. Globular proteins, such as hemoglobin, appear to undergo a temporal evolution of structure after ESI. This process involves side-chain collapse, unfolding, and subsequent refolding into new structures, as shown in Figure 7.19. Desolvation is followed, on the picosecond (ps) scale, by interactions on the protein surface which stabilise the native fold. This structure can remain unchanged

for up to milliseconds (ms), within the time frame of the experiments performed in this study (Breuker and McLafferty, 2008).

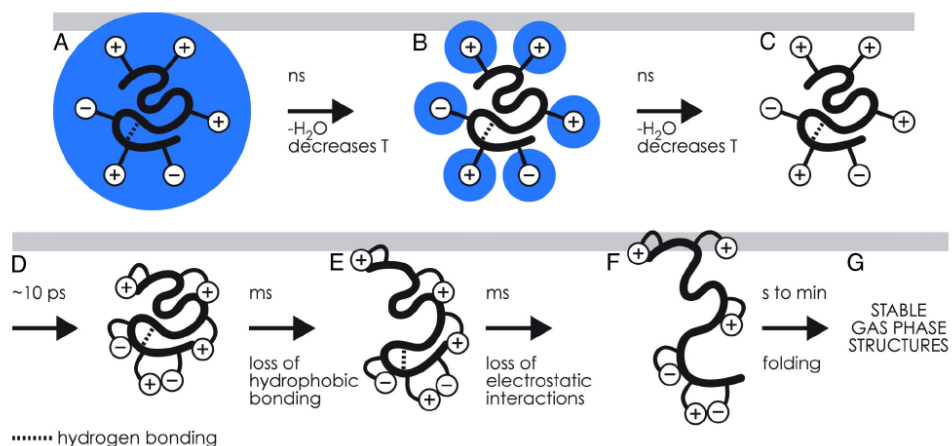


Figure 7.19 – Evolution of the structure of a globular protein following ESI, (A) native protein covered with a monolayer of water, (B) native protein with exterior ionic functionalities still hydrated (C) dry protein undergoes collapse of its exterior ionic functionalities, (D) millisecond loss of hydrophobic bonding, (E) millisecond loss of electrostatic interactions, (F) transiently unfolded ions form new noncovalent bonds in seconds, (G) more stable gaseous ion structures stabilise to energy minima conformers in minutes. From Breuker and McLafferty, 2008.

7.5 Conclusions

7.5.1 Inorganic analysis to determine metal content of pMMO

The determination of metal content by two inorganic analysis methods which both appear to indicate the presence of iron would suggest that the study of the third metal site of pMMO warrants further attention and that inorganic MS may be a useful tool. Although the LA-ICP-MS analysis proved unsuccessful in this study, it has been used to good effect elsewhere, albeit generally with sector-field (SF) instruments (Becker, J. Susanne *et al.*, 2009). It should also be considered that the laser system under use was a prototype. Use of a laser which offers better depth sampling, and with improved control on the ablation conditions could provide more positive results. The majority of work performed using ICP-MS in the field of biology has focussed on heavier elements, such as those found in anti-cancer drugs, or on systems with a higher overall metal concentration (Bettinelli, 2005).

Inorganic analysis of biological samples continues to expand, with the use of LA-ICP-MS with antibody labelling to detect proteins (Roos *et al.*, 2008) and the imaging of metals in human brain samples (Dobrowolska *et al.*, 2008). Improved methods to tackle isobaric interferences are needed, however, to fully utilise the potential of this approach. There are still a number of limitations in the sample preparation of pMMO, such as the lack of an expression system, the slow growth of the bacteria, and the incompatibility of detergents used for membrane solubilisation with MS analysis. Progress in the expression, purification and solubilisation of the complex could lead to better applicability of MS for the analysis of this complex, making use of the developments in both the organic and inorganic fields.

7.5.2 Probing hemoglobin structure

Travelling-wave ion mobility mass spectrometry has been successfully used to probe the gas-phase conformations of the three-dimensional protein structure of the non-covalent complexes of normal and sickle hemoglobin.

Non-tetrameric species were seen which correspond to apo- and holo- forms of the α - and β -monomers and $\alpha^h\beta^h$ -dimers; these are naturally present in solution, and are not products of in-source fragmentation during the ESI process. Both monomer types

have cross-sections similar to each other, suggesting that they maintain a similar folded structure. Extensively disordered monomer structures were not seen, and the similarity between the apo- and holo- forms indicates that the chains are able to retain a folded structure with or without the attachment of a heme group.

A heme-deficient dimer was not observed, and the results suggest that there is no requirement for the association of β^a with α^h in order for the β -monomer to recruit heme. The data, acquired on fresh blood samples rather than commercially prepared protein, do not support the hypothesis that a heme-deficient dimer is an essential intermediate in the assembly of the hemoglobin tetramer.

The cross-sections calculated for HbA and HbS are comparable to those estimated from published X-ray crystallographic data, supporting other work which investigated the relationship between values obtained from these different techniques. Conformational differences have been observed between the HbA and HbS molecules, a significant structural change caused by a single amino acid substitution. This result, from a relatively well-studied hemoglobin disorder, gives confidence for the use of this method for the investigation of other hemoglobin variants, and also for the study of other metalloproteins.

References

- Badman, E. R., Hoaglund-Hyzer, C. S., Clemmer, D. E.** (2001). Monitoring Structural Changes of Proteins in an Ion Trap over ~10-200 ms: Unfolding Transitions in Cytochrome c Ions. *Anal. Chem.* **73**: 6000-6007.
- Becker, J. S., Lobinski, R., Becker, J. S.** (2009). Metal imaging in non-denaturing 2D electrophoresis gels by laser ablation inductively coupled plasma mass spectrometry (LA-ICP-MS) for the detection of metalloproteins. *Metallomics* **1**: 312-316.
- Becker, J. S., Zoriy, M., Becker, J. S., Pickhardt, C., Damoc, E., Juhacz, G., Palkovits, M., Przybylski, M.** (2005). Determination of Phosphorus-, Copper-, and Zinc-Containing Human Brain Proteins by LA-ICP-MS and MALDI-FTICR-MS. *Anal. Chem.* **77**: 5851-5860.
- Bettinelli, M.** (2005). ICP-MS determination of Pt in biological fluids of patients treated with antitumor agents: evaluation of analytical uncertainty. *Microchem. J.* **79**: 357-365.
- Binet, M. R. B., Ma, R., McLeod, C. W., Poole, R. K.** (2003). Detection and characterization of zinc- and cadmium-binding proteins in Escherichia coli by gel electrophoresis and laser ablation-inductively coupled plasma-mass spectrometry. *Anal. Biochem.* **318**: 30-38.
- Boys, B. L., Kuprowski, M. C., Konermann, L.** (2007). Symmetric Behavior of Hemoglobin α - and β - Subunits during Acid-Induced Denaturation Observed by Electrospray Mass Spectrometry. *Biochemistry* **46**: 10675-10684.
- Breuker, K., McLafferty, F. W.** (2008). Stepwise evolution of protein native structure with electrospray into the gas phase, 10^{-12} to 10^2 s. *Proc. Natl. Acad. Sci. USA* **105**: 18145-18152.
- Carroll, J., Altman, M. C., Fearnley, I. M., Walker, J. E.** (2007). Identification of membrane proteins by tandem mass spectrometry of protein ions. *Proc. Natl. Acad. Sci. USA* **104**: 14330-14335.
- Dobrowolska, J., Dehnhardt, M., Matusch, A., Zoriy, M., Palomero-Gallagher, N., Koscielniak, P., Zilles, K., Becker, J. S.** (2008). Quantitative imaging of zinc, copper and lead in three distinct regions of the human brain by laser ablation inductively coupled plasma mass spectrometry. *Talanta* **74**: 717-723.
- Griffith, W. P., Kaltashov, I. A.** (2003). Highly asymmetric interactions between globin chains during hemoglobin assembly revealed by electrospray ionization mass spectrometry. *Biochemistry* **42**: 10024-10033.
- Hergenroder, R.** (2006). A model for the generation of small particles in laser ablation ICP-MS. *J. Anal. At. Spectrom.* **21**: 1016-1026.

- Hoaglund-Hyzer, C. S., Counterman, A. E., Clemmer, D. E.** (1999). Anhydrous Protein Ions. *Chem. Rev.* **99**: 3037-3080.
- Hossain, B. M., Konermann, L.** (2006). Pulsed Hydrogen/Deuterium Exchange MS/MS for Studying the Relationship between Noncovalent Protein Complexes in Solution and in the Gas Phase after Electrospray Ionization. *Anal. Chem.* **78**: 1613-1619.
- Jiménez, M., Gomez, M., Rodriguez, L., Martinez, L., Castillo, J.** (2009). Some pitfalls in PAGE-LA-ICP-MS for quantitative elemental speciation of dissolved organic matter and metalomics. *Anal. Bioanal. Chem.* **393**: 699-707.
- Kuprowski, M. C., Boys, B. L., Konermann, L.** (2007). Analysis of Protein Mixtures by Electrospray Mass Spectrometry: Effects of Conformation and Desolvation Behavior on the Signal Intensities of Hemoglobin Subunits. *J. Am. Soc. Mass Spectrom.* **18**: 1279-1285.
- Leutzinger, Y., Beychok, S.** (1981). Kinetics and mechanism of heme-induced refolding of human α -globin. *Proc. Natl. Acad. Sci. USA* **78**: 780-784.
- Ofori-Acquah, S. F., Green, B. N., Davies, S. C., Nicolaides, K. H., Serjeant, G. R., Layton, D. M.** (2001). Mass spectral analysis of asymmetric hemoglobin hybrids: Demonstration of Hb FS ($\alpha_2\gamma\beta^S$) in sickle cell disease. *Anal. Biochem.* **298**: 76-82.
- Rath, A., Glibowicka, M., Nadeau, V. G., Chen, G., Deber, C. M.** (2009). Detergent binding explains anomalous SDS-PAGE migration of membrane proteins. *Proc. Natl. Acad. Sci. USA* **106**: 1760-1765.
- Roos, P., Venkatachalam, A., Manz, A., Waentig, L., Koehler, C., Jakubowski, N.** (2008). Detection of electrophoretically separated cytochromes P450 by element-labelled monoclonal antibodies via laser ablation inductively coupled plasma mass spectrometry. *Anal. Bioanal. Chem.* **392**: 1135-1147.
- Santoni, V., Molloy, M., Rabilloud, T.** (2000). Membrane proteins and proteomics: Un amour impossible? *Electrophoresis* **21**: 1054-1070.
- Shelimov, K. B., Clemmer, D. E., Hudgins, R. R., Jarrold, M. F.** (1997). Protein Structure in Vacuo: Gas-Phase Conformations of BPTI and Cytochrome c. *J. Am. Chem. Soc.* **119**: 2240-2248.
- Smith, D. D. S., Dalton, H.** (1989). Solubilisation of methane monooxygenase from *Methylococcus capsulatus* (Bath). *Eur. J. Biochem.* **182**: 667-671.
- Tahallah, N., Pinkse, M. M., C. S., Heck, A. J. R.** (2001). The effect of the source pressure on the abundance of ions of noncovalent protein assemblies in an electrospray ionization orthogonal time-of-flight instrument. *Rapid Commun. Mass Spectrom.* **15**: 596-601.

Valentine, S. J., Anderson, J. G., Ellington, A. D., Clemmer, D. E. (1997). Disulfide-Intact and -Reduced Lysozyme in the Gas Phase: Conformations and Pathways of Folding and Unfolding. *J. Phys. Chem. B* **101**: 3891-3900.

Chapter 8

Introduction to mass spectrometry-based proteomics

8.1 Mass spectrometry-based proteomics

Proteomics can be loosely described as the large-scale study of proteins, and as a field of research it makes a crucial contribution to our understanding of biology. The term proteome was first coined by Mark Wilkins in 1995 as being the protein complement of the genome (Wilkins *et al.*, 1996). A more specific definition of proteomics is the global characterisation of protein products expressed by a given genome at a specific point in time, under particular physiological conditions. Since the mid-1990s, mass spectrometry-based strategies have been the mainstream method for determining protein identifications (Pandey and Mann, 2000). There remain, however, a number of issues to be tackled. Intrinsic characteristics of proteomes raise a number of experimental challenges. By nature, proteomes are large and complex. A single gene can often give rise to multiple, distinct proteins due to alternative splicing, sequence polymorphisms and post-translational modifications, so that protein databases generated from an organism's genome may not be a true reflection of the potential protein complement. The advent of electrospray ionisation (ESI) and matrix-assisted laser desorption-ionisation (MALDI) in the late 1980s catalysed the emergence of methods for the analysis of proteins by mass spectrometry (Domon and Aebersold, 2006). In recent years, with the development of new mass analysers and complex instruments, there has been significant progress in the area of mass spectrometry-based proteomics. The use of MS for proteomics is not the application of a single technique, but rather a collection of methods, each with its own strengths suited to particular areas of interest (Han *et al.*, 2008). Approaches in mass spectrometry-based proteomics can be grouped into two broad definitions: top-down and bottom-up.

8.1.1 Top-down proteomics

First described in 1999 (Kelleher *et al.*, 1999), top-down proteomics is the analysis of intact proteins, combining ESI with high resolution mass spectrometry. A mixture of proteins is introduced into the mass spectrometer; molecular ions of an individual component can then be mass-isolated and dissociated to yield fragment ions (Reid and McLuckey, 2002). The high mass accuracy of

instrumentation used for top-down proteomics approaches has traditionally led to their use for the characterisation of post-translational modifications such as acetylation and methylation (Lee *et al.*, 2002). The dissociation of protein ions for MS/MS analysis originally employed threshold activation methods with sequential activation events, such as energetic collisions and IR photon absorption, that cleave the weakest bonds within the protein first. More recently, the method of electron capture dissociation (ECD) (Zubarev *et al.*, 1998) and its off-shoot, electron transfer dissociation (ETD) (Syka *et al.*, 2004), add a larger amount of energy such that only backbone bonds are cleaved, generating *c* and *z* ions. The fragment masses are matched against those predicted from the protein sequence, to provide identification. The ECD/ETD mechanism preserves side-chain modifications, such as glycosylation and phosphorylation, which enables site-specific analyses.

The top-down proteomic approach requires high resolution and mass accuracy, and therefore, most top-down studies to date report the use of Fourier transform ion cyclotron resonance (FTICR) MS instrumentation. A front-end input capability for ion storage and mass separation, such as the linear trap quadrupole (LTQ), provides additional control permitting selection of specific MS and MS/MS experiments. The recently introduced Orbitrap (Makarov, 2000) shows promise in the field of top-down analysis, although its resolving power and mass accuracy do not yet match FTICR instruments. There has also been some study into the use of quadrupole time-of-flight instrumentation for top-down analysis of complex protein mixtures (Hayter *et al.*, 2003).

Top-down proteomics provides two major advantages. The first is the potential to cover the complete sequence of the protein under investigation; the second is the ability to locate and characterise post-translational modifications. This approach, however, is a young field in comparison to bottom-up proteomics; as such, it suffers from a number of limitations. The complex spectra which are generated often require significant interpretation. The need for FT instrumentation is another limitation of top-down proteomics, as they have high associated purchasing, operational and maintenance costs (Marshall, 2000). Size limitations also exist; typically top-down experiments have been used to characterise proteins up to 50 kDa (McLafferty *et al.*, 2007). The use of new methods such as

prefolding dissociation has, however, enabled the analysis of higher molecular masses (Han *et al.*, 2006). This limit is a significant concern, as around 50 % of known proteins are over 50 kDa in size. The dissociation methods of ECD and ETD often require long accumulation, activation and detection times, and the mechanisms of protein dissociation are less well understood than those of peptide fragmentation. High-throughput remains a challenge for top-down proteomics, as the approach struggles in the analysis of very complex protein mixtures.

8.1.2 Bottom-up proteomics

The bottom-up approach to proteomic study is the most widely used for identifying proteins and determining details of their sequence (Chait, 2006). Proteins are digested using an enzyme, and the resulting peptides are analysed by means of mass spectrometry.

A traditional proteomics approach is based on the separation of proteins via polyacrylamide gel electrophoresis (PAGE). Proteins are then digested within the gel, and the resulting peptides extracted for MS analysis. Two-dimensional (2D) gels were developed in the 1970s (O'Farrell, 1975), and have gained considerable popularity in the area of large-scale protein separation. A protein mixture is first separated according to isoelectric point (the pH at which the protein carries no net charge), usually with precast IPG strips (Ek *et al.*, 1983). Various gradients exist that allow different parts of the proteome to be focussed on, thereby increasing resolution. Strips with a non-linear pH gradient are also available, again to increase resolution. The second separation is according to molecular size. Proteins are visualised using a staining solution, the most commonly used being Coomassie Blue R-250, Coomassie Blue G-250, silver stain and fluorescent dyes such as SYPRO red.

Limitations associated with PAGE include limited dynamic range, insufficient resolving power to fully separate all proteins within a sample, and restricted sample throughput (Horgan, 2007). More recently, non-gel based techniques have become popular for the analysis of complex proteomic samples in order to overcome some of these limitations. Liquid phase separations of intact proteins are becoming attractive alternatives to gel-based separations. Linear separations

that are orthogonal in selectivity are combined, permitting a wide range of separation mechanisms to be applied, such as molecular size, hydrophobicity, ionic character and specific affinity interactions. Such methods offer increased selectivity compared to gels, and can be scaled up to meet sample and analysis requirements. These are often also applicable to proteins which may not be compatible with gel separations, for example those with extreme pI values or those which are hydrophobic (Millea *et al.*, 2005). Another major advantage of liquid phase separations is that they provide the option to generate a fractionated proteome of intact proteins which can then be analysed in a top-down manner as previously described.

Alternatively, analysis can be performed entirely at the peptide level. So-called ‘shotgun’ experiments are employed, where a whole proteome is digested without prior protein separation. Typically, the resulting peptides are separated by strong cation-exchange chromatography before MS analysis. This method can be favoured due to its increased proteome coverage compared to gels, although it still suffers from problems in reproducibility and dynamic range. This coupling of strong cation-exchange chromatography with reverse-phase chromatography and MS is part of what is known as ‘MudPIT’: multi-dimensional protein identification techniques (Link *et al.*, 1999). This approach is becoming increasingly utilised in proteomic studies in preference to gels (Domon and Aebersold, 2006). Once peptides have been generated, they are separated as desired, often in multiple dimensions (Issaq *et al.*, 2005). Protein identification can then take place using peptide mass fingerprinting (Henzel *et al.*, 1993) or tandem MS.

Commonly used analysers in bottom-up proteomics are quadrupole (Q), ion trap (quadrupole ion trap, QIT; linear ion trap, LIT or LTQ), and time-of-flight (ToF). These vary in their performance, and many hybrids have been designed to combine various capabilities. Each type of instrumentation has benefits and disadvantages.

Quadrupoles are low cost, robust and easy to maintain, but possess limited mass range and resolving power, and cannot perform MS/MS (El-Aneed *et al.*, 2009). Some of these issues can be overcome by attaching the quadrupole to other

analysers, and also the combination of multiple quadrupoles. The triple quadrupole was introduced in 1978, and consists of two quadrupoles and a hexapole (Yost and Enke, 1978). The first and last are mass spectrometers, the centre quadrupole is a collision cell using RF only (de Hoffmann and Stroobant, 2007).

Ion traps are often the most affordable option, but have a number of limitations. Only a finite number of ions can be trapped, and only half are transferred to the detector when released. They also suffer from limited mass accuracy and resolution, although a narrow scan range can be employed to improve these. Coupling ion traps to quadrupoles overcomes some of these limitations, as it allows the accumulation of more ions and involves the use of two detectors, minimising ion loss. The drawback, however, is a reduced mass range.

ToF analysers have the advantage of being able to detect a very high mass range, but their major limitation is the inability to perform true MS/MS. Although post-source decay (PSD) (Brown and Lennon, 2002) is an option, there is restricted ability to select ions, and the fragmentation incurred is not predictable. ToF/ToF instrumentation goes some way in tackling this problem (Bienvenut *et al.*, 2002). The first ToF is used for the ion selection process, fragmentation occurs in a collision cell and the product ions are analysed in the second ToF. Although this is a rapid technique, a relatively large amount of sample is required for sufficient ion signal. Due to the high-energy nature of this process, the dissociation which takes place is more complex than alternative techniques, with side-chain fragmentation observed. Sequencing of peptides is possible, but as a result of the complex fragmentation which takes place the data obtained are not generally compatible with available database searching programs.

The combination of the scanning capabilities of a quadrupole and the resolving power of a ToF analyser was first described in 1996 (Morris *et al.*, 1996). Q-ToF instruments usually include an additional quadrupole as an ion focussing device. One of the major advantages is the ability to interface with either ESI or MALDI, and the quadrupole offers superior selectivity for precursor ions. Other advantages include ease of operation and high resolution.

The Orbitrap has also been used in peptide-based, bottom-up proteomics studies, with results indicating good resolution, mass accuracy and sequence coverage, although the fragmentation was originally performed in a linear ion trap environment (Yates *et al.*, 2006). Improved approaches using higher energy traps have recently been developed (Olsen *et al.*, 2009; Zhang, Yi *et al.*, 2009).

A summary of these instruments and their distinguishing characteristics is given in Table 8.1.

Figure 8.1 gives an overview of the approaches employed in both top-down and bottom-up proteomics. All of the various experimental methods have the potential to provide information on different aspects of a proteome. Complete characterisation is the ultimate goal of proteomics study, and in order to achieve this it should be considered that complementary techniques need to be employed (Nemeth-Cawley *et al.*, 2003).

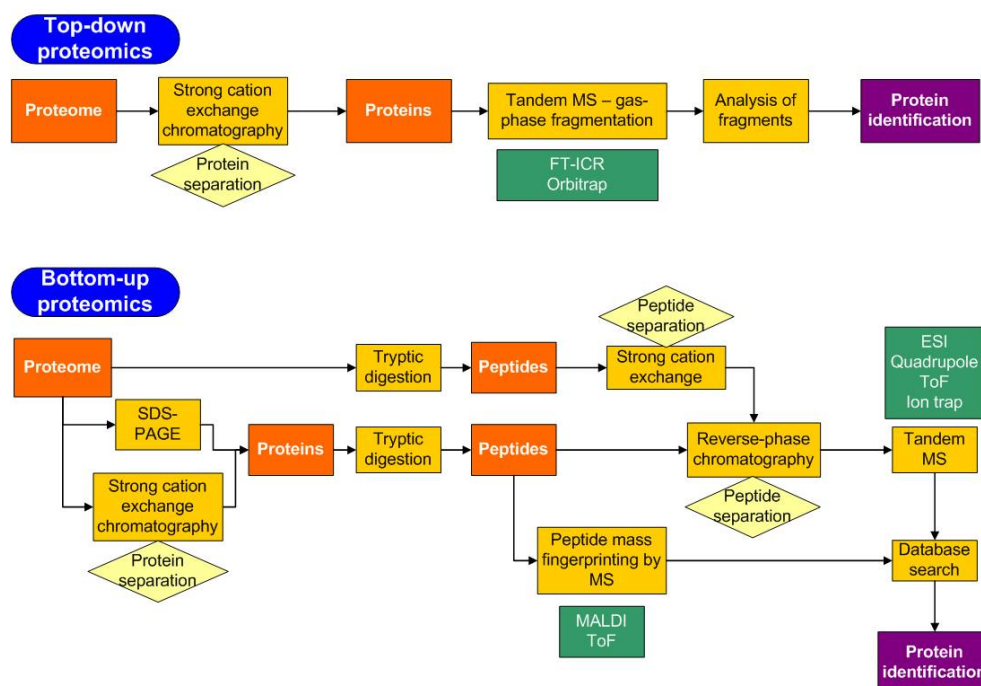


Figure 8.1 – Some typical approaches in top-down and bottom-up proteomics.

Instrument	Mass resolution	Mass accuracy (ppm)	Sensitivity	m/z range	Scan rate	Dynamic range	MS/MS capability	Ion source	Main applications
QIT	1000	100-1000	picomole	50-2000 200-4000	Moderate	10^3	MS ⁿ	ESI	Protein identification in samples of low complexity; PTM identification.
LIT	2000	100-500	femtomole	50-2000 200-4000	Fast	10^4	MS ⁿ	ESI	High throughput large-scale protein identification from complex mixtures; PTM identification.
Q-q-Q	1000	100-1000	attomole to femtomole	10-4000	Moderate	6×10^6	MS/MS	ESI	Quantification (selective reaction monitoring); PTM detection (precursor ion and neutral loss scanning).
Q-q-LIT	2000	100-500	femtomole	5-2800	Fast	4×10^6	MS ⁿ	ESI	
TOF	10,000-20,000	10-20 / <5	femtomole	No upper limit	Fast	10^4	n/a	MALDI	Protein identification from digested samples (peptide mass fingerprinting).
TOF-TOF	10,000-20,000	10-20 / <5	femtomole	No upper limit	Fast	10^4	MS/MS	MALDI	Protein identification from digested samples (peptide mass fingerprinting or CID MS/MS).
Q-q-TOF	10,000-20,000	10-20 / <5	femtomole	No upper limit	Fast	10^4	MS/MS	ESI MALDI	Protein identification from complex mixtures; Intact protein analysis; PTM identification.

FTICR	50,000- 750,000	<2	femtomole	50-2000 200-4000	Slow	10^3	MS ⁿ	ESI MALDI	Top-down proteomics; High mass accuracy PTM characterisation; Protein identification from complex mixtures
LTQ- Orbitrap	30,000- 100,000	<5	femtomole	50-2000 200-4000	Moderate to fast	4×10^3	MS ⁿ	ESI MALDI	Top-down proteomics; High mass accuracy PTM characterisation; Protein identification from complex mixtures; Quantification.

Table 8.1 – Comparison of performance characteristics of commonly used mass spectrometers for proteomics (adapted from Han et al., 2008)

8.2 Quantitative proteomics

Within the field of proteomics, in addition to a profile of what proteins are present within a system, information on the expression levels of these proteins is also often required. Since the proteome is a dynamic entity, the abundance of proteins differs depending on many factors, such as the type of cell they are expressed in, developmental stage, and changes in the environment. Information regarding protein abundance can provide insight into cellular processes, and lead to better understanding of the biological system under study. Techniques in quantitative proteomics have, therefore, also developed significantly in recent years. These quantitative techniques are generally associated with a bottom-up proteomics approach.

8.2.1 Gel-based quantification

Relative quantification can be performed to look at the levels of proteins separated by two-dimensional PAGE, using image analysis software. The gels to be analysed are scanned using a densitometer, and the captured images analysed. The software normalises and filters the images; spots are identified and manually checked, before being matched between the gels. Changes are recorded in relation to differences in spot size and intensity, and used to infer differences in protein concentration. This method suffers the effects of 2D gel limitations already mentioned, including low reproducibility. To overcome some of the problems of gel-to-gel variations, difference gel electrophoresis (DiGE) was developed (Ünlü *et al.*, 1997). Fluorescent dyes that are mass and charge matched, but spectrally resolvable, are used to label different protein samples. All samples are mixed together and resolved on the same gel, meaning the resulting images are perfectly overlaid. DiGE does not, however, omit all problems with a gel-based approach. The issue of inter-gel variation is tackled for a single experiment, but a number of gels must be run for statistical significance, making the process unsuitable for high-throughput analyses. There is also a considerable associated cost, as the dyes are expensive, and specialised equipment is required to scan the gels.

8.2.2 Peptide labelling

A number of labelling approaches can also be incorporated into MudPIT experiments. These include isotope dilution (Fairwell *et al.*, 1970), stable isotope labelled peptides (Ong *et al.*, 2002), radiolabelled amino acid incorporation (Sirlin, 1958), chemically synthesised peptide standards (Gerber *et al.*, 2003) and isotope-coded affinity tags (ICAT) (Gygi *et al.*, 1999). An approach which has become increasingly popular is the use of isobaric tags for relative and absolute quantification (iTRAQ) in which the N-terminus of peptides generated by tryptic digest are chemically tagged (Ross *et al.*, 2004). Each tagging reagent consists of a reporter group (based on N-methylpiperazine), a balance group (carbonyl) and an amine-specific peptide reactive group (N-hydroxy-succinimide ester). The system is commercially available (Applied Biosystems, USA) with four tags containing reporter groups of masses 114.1, 115.1, 116.1 and 117.1 Da, and balance groups of masses 28, 29, 30 and 31 Da. Reporter and balance groups are combined so that each of the reagents has a total mass of 145.1 Da. Upon tandem MS analysis by CID, the balance and reporter groups are lost from the peptides, with the reporter groups retaining the charge. These appear as distinct ion species in the spectrum, and are designed not to interfere with MS/MS data from the peptide fragments, as they appear in a quiet region of the spectrum where peptide ions are not observed. The intensity of reporter ions is used to quantify associated peptides. A summary of the iTRAQ system can be seen in Figure 8.2. This system has been widely used in proteomic studies (Sadowski *et al.*, 2006) and is now commercially available with eight isobaric tags (Ow *et al.*, 2008).

8.2.3 Non-labelling quantification

All labelled quantification approaches experience drawbacks: complex sample preparation, the requirement for large sample amounts, incomplete labelling. There has, therefore, been development in the area of non-labelled quantification to make an attempt at tackling some of these problems. Recently, a number of label-free approaches have emerged (Panchaud *et al.*, 2008).

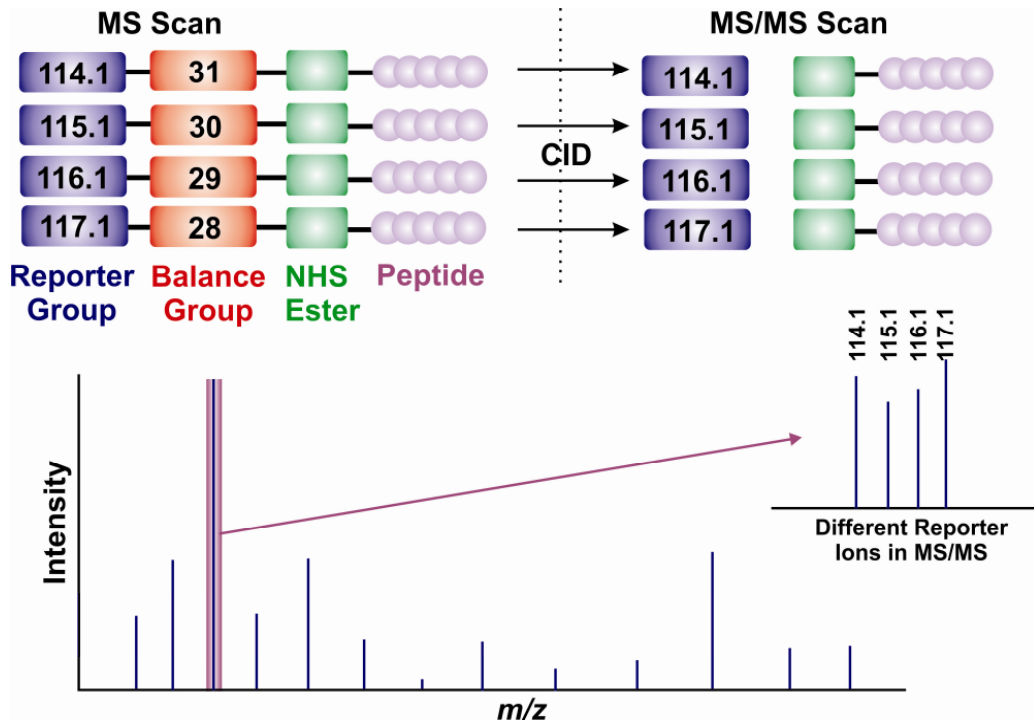


Figure 8.2 – An overview of the iTRAQ labelling quantification system.

Peptide match score summation (PMSS) uses peptide identification scores to derive a semi-quantitative abundance indicator (Allet *et al.*, 2004). A closely related method is spectrum sampling (SpS), based upon the relationship between the level of sampling observed and relative abundance of the protein, where the spectra identifying each protein are counted (Liu *et al.*, 2004). Both of these can be combined with a statistical test to detect differentially expressed proteins (Colinge *et al.*, 2005).

Another approach to estimate protein content utilises a protein abundance index (PAI), which is derived by normalising the number of observed peptides with the number of observable peptides within a sample (Rappsilber *et al.*, 2002). This uses the correlation between the number of peptides generated and the amount of protein present. There has been shown to be a linear relationship between the number of peptides and the logarithm of protein concentration, therefore PAI is converted to exponentially modified PAI (emPAI) for absolute protein quantification (Ishihama *et al.*, 2005).

The intensity of MS signal cannot be predicted from peptide sequence, although it is generally accepted that peptides containing more basic amino acids are more likely to be seen, and signal intensity for the same peptide may vary depending on the sample. This does not apply, however, when comparing chromatographic runs, although very accurate and reproducible LC and MS data is required (Chelius and Bondarenko, 2002).

A methodology for relative quantification based on LC-MS was described which utilised a reproducible chromatographic separation system along with high mass resolution and mass accuracy time-of-flight MS (Silva *et al.*, 2006a). This has since been developed into a label-free system, capable of both relative and absolute quantification. All detectable, eluting peptides and their corresponding fragments are observed via rapid switching between high and low collision energy during the LC-MS/MS experiment, giving a comprehensive list of all ions. Post-acquisition data analysis methods then extract chromatographic and MS information on the peptides, resulting in time-resolved accurate mass measurements which can be used for both identification and quantification. Absolute quantification of proteins is possible due to the discovery of a relationship between MS signal response and protein concentration, which states that the average MS signal response for the three most intense tryptic peptides per mole of protein is constant within a coefficient of less than 10 per cent (Silva *et al.*, 2006b). This analysis approach has been termed MS^E, and is available commercially as the Waters Identity^E and Expression^E systems.

Despite these advancements in the field of proteomics, the identification and quantification of proteins from large-scale, automated experiments can still be problematic. An ideal approach would enable comprehensive studies of proteomes in a high-throughput manner. Currently, the techniques involved can be complex, with high costs and may involve time-consuming data analysis. A low number of biological replicates, due to a lack of sample availability, mean that reproducibility is a big concern. In addition, any given technique may only yield information on a fraction of relevant peptides in any single analytical run (Wilkins *et al.*, 2006). Figure 8.3 summarises a bottom-up proteomics experiment, including a quantitative element, highlighting key technical limitations and intrinsic biological challenges.

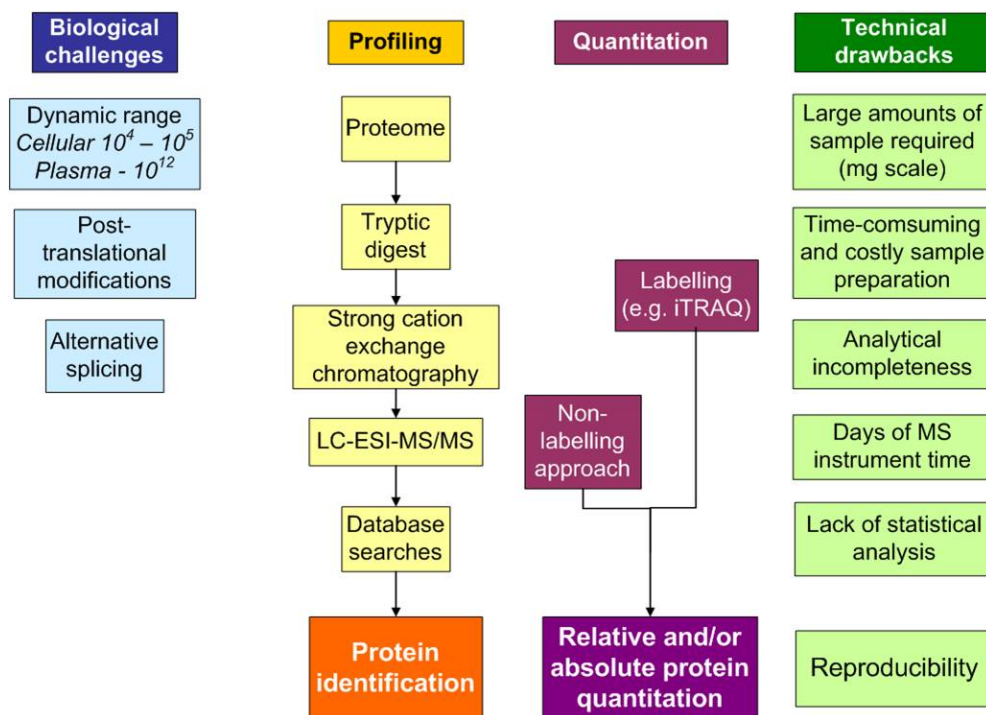


Figure 8.3 – A typical workflow for a bottom-up profiling and quantitative proteomics experiment, along with some of the associated limitations and problems.

8.3 Computational proteomics

Computational proteomics is the term used to refer to the computational methods, algorithms, databases and bioinformatics methodologies used to process, manage, analyse and interpret the data produced in proteomics experiments (Cannataro, 2008). The ultimate goal of computational proteomics is to infer knowledge models, such as the identification of proteins involved in a particular disease, from the inspection of biological samples.

Proteomic data are typically very complex, leading to a number of practical considerations when undertaking data analysis. The raw data must be processed, and protein identifications must be performed and subjected to quality control. If quantification is being performed then this information must also be extracted and combined with the profiling information to give a comprehensive dataset. Applicable statistical methods should be employed to check the validity of the data, and ideally all this should be combined into an easily accessible workflow.

As with most analytical fields, these processes are hugely dependent on computational tools. The dynamic process of software development progresses alongside the technical development of instrumentation. The resulting computational programs are often developed for specific types of mass spectrometers and will only operate with a limited number of platforms. It is, therefore, not trivial to choose an appropriate program suitable for the analysis of proteomics datasets. In addition to the limited applicability of programs to various MS platforms, it is important to consider practical factors such as file compatibility, computational requirements, ease of use, visualisation options, and variations in the sample preparation protocol. MS-based proteomics offers a high-throughput platform and the capability to analyse complex samples on a global scale, but this has led to a computational bottleneck, as available methods struggle to cope with the large datasets generated.

8.3.1 Protein identification

In protein identification, there are two basic approaches: database searching and *de novo* sequencing. Database searching methods have long been the primary method of identification (Eng *et al.*, 1994), as the fragmentation and sequence coverage required for *de novo* methods have not always been achievable with available mass spectrometers. Database searches also offer a much higher degree of automation in their setup, whether using a freely available program such as Mascot (Matrix Science, Boston, USA) or commercial software such as SEQUEST (Thermo, MA, USA) and ProteinLynx (Waters, MA, USA). The availability of a suitable protein database to search will depend upon the system under investigation. Many organisms will have specific databases, although many of those that are commonly studied are listed as part of the Universal Protein Resource (UniProt). This is a resource on protein sequence and annotation, formed in collaboration between the European Bioinformatics Institute (EBI), the Swiss Institute of Bioinformatics (SIB) and the Protein Information Resource (PIR).

One of the major issues with database searches is the large false positive rate associated with the sheer number of peptide candidates to consider for each

spectrum searched (Cargile *et al.*, 2004). As more proteomes are sequenced and protein databases further expanded, there is an increasing need for accuracy. New algorithms for matching experimental to *in silico* spectra have begun to focus on accounting for multiple sources of information, such as mass peak intensity and correlation among ions (Sun *et al.*, 2007). The performance of search algorithms will have some dependence upon the platform being used for experimentation. It is important to understand inherent limitations, and to improve the quality of the identifications by either filtering out spectra with insufficient information or removing low signal peaks ('noise' or less common fragments). An approach to reduce the raw size of a database and to remove potential false positives is the generation of proteotypic peptides, that is, peptides which are detectable by an MS-experiment. For example, certain peptides do not contain ionisable amino acids, thereby carrying only a minimal charge on the N-terminus and reducing the likelihood of their capture by MS and subsequent identification in a database search (Craig *et al.*, 2005).

Some work has focused on developing sequence-specific fragmentation models, as fragmentation patterns vary significantly depending on the sequence of the peptide (Zhang, Z, 2005). There are certain problems to this approach. The mere evaluation of a peptide's amino acid composition is not sufficient to predict fragmentation patterns on longer chains. In addition, it is not possible to build a sufficiently diverse spectral database; just a 6-mer would generate 20^6 unique peptides requiring observation. An older method to deal with insufficient sampling which has re-gained popularity is the direct use of spectral libraries. Identified spectra for specific peptides serve as models, against which newly acquired spectra are compared. A major drawback to this approach, however, is that incorrect annotations will be propagated through libraries. Additionally, peptides that undergo incomplete fragmentation are likely to be underrepresented (Webb-Robertson and Cannon, 2007).

De novo sequencing is an option which can be employed in cases where database searches cannot identify a peptide from MS/MS data. This may occur for a number of reasons: novel proteins, mutations, post-translational modifications, database sequencing errors, and meta-proteomic studies. Although, in principle, an MS/MS spectrum contains a set of ions that can be used to sequence a peptide,

it is often the case that incomplete fragmentation and low mass accuracy will only provide partial sequence information. This has seen some improvement in recent years, with the advent of newer instrumentation such as the Orbitrap and high-resolution Q-TOFs (Frank *et al.*, 2007). It is unlikely that *de novo* sequencing will become the commonly used method for protein identification, owing to the computationally compatible and user-friendly nature of database searching, but it does provide important flexibility in instances where genomes may be insufficiently sequenced or unknown. The ease of using *de novo* sequencing approaches has been assisted by the introduction of associated software, such as PEAKS (Bioinformatics Solutions, Ontario, Canada).

8.3.2 Protein quantification

There are a number of computational options available to handle proteomic data containing a quantitative aspect, from a labelling or non-labelling approach. Certain programs are freely available for the analysis of iTRAQ data, which import pre-processed MS/MS data in certain file formats. Commercial systems also offer a number of processing functionalities for more sophisticated analysis of platform-specific iTRAQ data; these include ProQuant and ProteinPilot from Applied Biosystems, and ProteinLynx from Waters.

The XPRESS software was developed for the analysis of ICAT data and is sold commercially as part of BioWorks from Thermo. ASAPRatio is a more advanced program, which includes additional downstream statistical methods (Li *et al.*, 2003); both these systems work on isotope-labelled data from Thermo's quadrupole linear ion trap (LTQ) and FT-LTQ instrumentation. SILAC experiments can be quantified by STEM, and also by MSQuant, which imports MS/MS data from Mascot searches. Commercial solutions such as Bruker's WARP-LC offer a generic platform for the quantification and visualisation of various isotopic labelling data.

Label-free quantitative studies often span a large number of LC-MS measurements; this leads to volumes of data, which require considerable computational resources to complete data analysis within a reasonable amount of time. Spectral counting methods transform frequency of peptide identification into a measure for peptide

abundance, and have most commonly been used for analysis of low to moderate mass resolution LC-MS data. Software for automated quantification via spectral counting has been developed as laboratory information management systems (LIMS), programming scripts and stand-alone software packages. The program Scaffold (Proteome Software, Oregon, USA) enables spectral counting quantification by searching a collection of MS/MS spectra against a peptide matching algorithm such as SEQUEST.

Data from signal intensity or peak area approaches can be processed using a number of computational options, several of which are freely available. These include SpecArray (part of the Trans Proteomics Pipeline), msInspect (Fred Hutchinson Cancer Research Centre, Seattle, USA), MSight (Swiss Institute of Bioinformatics), OpenMS and PEPPER (Jaffe *et al.*, 2006). Although these are accessible with no purchase costs, all carry limitations with respect to the data file types they can process, and the operating systems with which they are compatible.

A key challenge in high-throughput LC-MS analysis is how to detect and handle poor quality data, e.g. low signal response, imperfect chromatography, in large datasets. It is common to use a reference acquisition against which to normalise all other data; the use of a spiked internal standard can prove useful in this respect. The program SuperHirn implements an alternative strategy, where similarities between all LC-MS runs are computed based on common peptides of an LC-MS pair and their intensity correlation. It is compatible with mzXML-formatted QTOF, FT-LTQ and Orbitrap data, and is available for Linux and OSX platforms (Mueller *et al.*, 2007).

Commercial analysis frameworks are also available, developed and distributed under licence to be used with particular manufacturers' platforms. MassLynx is the dedicated data processing system for Waters instrumentation. The accompanying ProteinLynx software and GlobalSever search engine are essential components in the analysis of acquired measurements, with the most recent release containing in-built options to handle label-free quantification processing of MS^E data. LC-MS data from Thermo instruments, i.e. LTQ, FT-LTQ, Orbitrap, can be processed by the software package termed Statistical Iterative Exploratory Visualisation Environment (SIEVE). This performs extraction and alignment of

peptide signals across measurements and provides GUI functionality to assess the alignment and tools to perform discriminative analysis of peptide abundance changes.

Also commercially available is Expressionist from GeneData (Basel, Switzerland), which is predominantly a platform for biomarker discovery, capable of integrating different ‘omics’ data. Previously available was the Elucidator framework from Rosetta Biosoftware, which offered statistical tools for data exploration and extensive GUI functionality with a complete MS/MS processing routine, feature extraction and LC-MS alignment modules. Since a commercial takeover earlier this year, however, Elucidator has not been re-introduced.

All these commercial and platform-specific systems allow the user to assemble customised automated workflows for high-throughput data analysis. A drawback, however, is that these are closed, that is, there is no access to the source code. Workflows and data formats are implemented within the program, limiting options once the raw data has been imported into the software. In addition, commercial programs come with a cost associated with the professional setup they offer.

8.4 Quality of data

Proteomics studies typically generate large and complex sets of data; these must be statistically analysed if the results are to be meaningfully interpreted, perhaps more so in quantitative studies where regulation of protein expression is often the subject of interest. It is important to have a clear understanding of the experimental setup and the nature and quality of obtained data in order to select suitable statistical methods.

Results typically improve with increased sampling, as a greater number of biological and technical replicates allow for more stringent statistical analysis. It is not uncommon that such replicates are unavailable due to sample availability, and the cost and time which are required to perform repeat experiments. In such cases, however, biological effects may be missed as technical bias overshadows true protein regulation. Label-free approaches may offer an advantage in this area, as generally sample requirements are lower than labelling methods. The problem

of incomplete, insufficient or inconsistent labelling is also negated, which can be a significant technical issue in quantitative studies. The use of statistical testing methods also assists in determining the probability of false assignments. The performance of a particular statistical test is aided greatly by the use of reference data, another benefit of including an internal standard in label-free experiments.

The recent developments in quantitative proteomics have resulted in an exponential increase in data volume and complexity. This demands the development of appropriate statistical approaches in order to arrive at meaningful interpretations of data obtained. There is a danger that the computational methods outlined in Section 8.3 are used with little or no validation, leading to poor standards in data handling. As mass spectrometers and associated software become more widely available, there are an ever-increasing number of laboratories with little or no mass spectrometry experience that are performing analysis under the assumption that the results obtained are unquestionably correct.

It has been suggested that low quality MS/MS spectra should be filtered out of the analysis, or that machine learning techniques should be used to classify database search results and thereby improve discrimination between correct and random matches. Another approach is to send the data to more than one search engine in order to enable the cross-validation or consolidate results, so that confidence in the results obtained can be increased. A drawback of this stems from existing issues regarding the compatibility of types of data files. Another problem with direct cross-validation between different search engines is that each algorithm uses different methods for assessing the quality of its peptide assignments, meaning that comparisons can sometimes be misleading (Stead *et al.*, 2008).

A popular approach to estimate false-positive rates, and so gain an indication of information quality, is the use of decoy databases. These are composite databases containing all possible protein and peptide sequences from a given organism as well as an equivalent number of nonsense protein and peptide sequences that should not be present in the sample analysed (Huttlin *et al.*, 2006). For its ease of implementation and applicability across platforms, search engines and experimental conditions, the use of combined forward and reversed protein databases for assessment of false-positive rates has become commonplace.

8.5 Mapping proteomic data to biological pathways

A grand challenge in the post-genomic era is a complete computer representation of the cell, the organism, and the biosphere, which will enable computational prediction of higher-level complexity of cellular processes and organism behaviours from genomic and molecular information.

To take on this challenge, the Kanehisa Laboratories in the Bioinformatics Center of Kyoto University and the Human Genome Center of the University of Tokyo have developed a resource called the Kyoto Encyclopedia of Genes and Genomes (KEGG). This is a database of biological systems that integrates genomic, chemical and systemic functional information (Kanehisa *et al.*, 2008; Kanehisa and Goto, 2000). The KEGG Automatic Annotation Server (KAAS) is a tool which provides BLAST comparisons of genomic or proteomic data against the manually curated KEGG database. The result contains orthology assignments to KEGG pathways, of which there are approximately 90,000.

8.6 *Methylocella silvestris*

In this work, three bottom-up proteomics approaches have been used to identify and relatively quantify the proteins within *Methylocella silvestris*, a bacterium from the methanotroph family. Methylootrophs use one-carbon compounds as their sole carbon and energy source, and methanotrophs are the subset of these which use methane. Methane is well-known as a greenhouse gas, and methanotrophs hold a unique place in global methane cycling. They oxidise a high percentage of biogenic methane before it reaches the atmosphere, making them of significant environmental importance (King, 1997).

Methanotrophs use the enzyme methane monooxygenase to oxidise methane. This enzyme exists in two forms: soluble methane monooxygenase (sMMO) and membrane-bound, particulate methane monooxygenase (pMMO). Until recently, it was believed that all methanotrophs were unable to utilise substrates containing carbon-carbon bonds. A newly-discovered genus, *Methylocella*, has been shown to grow on multi-carbon compounds such as acetate and ethanol, whilst retaining its ability to grow on single-carbon sources (Dedysh *et al.*, 2005). Figure 8.4

shows cultures of the bacterium *Methylocella silvestris* grown using acetate as the sole carbon and energy source. Studies on *M. silvestris* show that whilst it utilises sMMO for methane oxidation as in other methanotrophs, pMMO is absent (Theisen *et al.*, 2005); this has not been observed in any other organism within the family. Identification of enzymes involved in *Methylocella* metabolism, particularly when using multi-carbon compounds, is essential to understanding the biochemical pathways in this organism.

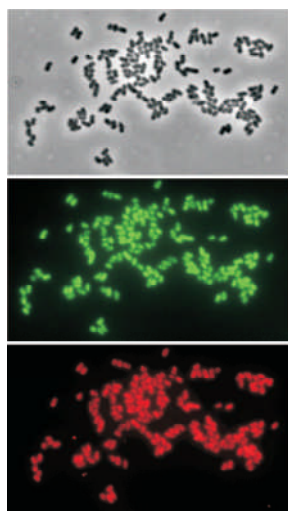


Figure 8.4 – Whole cell hybridisation in a culture of *Methylocella silvestris* grown on acetate as the sole carbon and energy source. Upper – phase contrast, middle – hybridisation with the *Methylocella* genus-specific probe Mcell-1445, lower – hybridisation with the *M. silvestris* species-specific probe Mcells-1024 (Dedysh *et al.*, 2005)

8.7 Aims and objectives

The field of proteomics is rapidly evolving; as described, techniques are constantly being developed and improved to deal with the enormous complexity that proteomes present. No single approach can provide a complete picture; the task therefore becomes to choose the most suitable experimental methods to obtain as much relevant information as possible.

The overall objectives of this project were to look at changes in protein expression under different experimental conditions in order to understand biological processes in prokaryotes, and to develop a confident, comprehensive approach for the identification and quantification of proteins that would enable this to be done.

More specifically, the aims were:

1. To characterise proteins present within the organism *Methylocella silvestris*.
2. To measure changes in expression of these proteins under varying growth conditions.
3. To relate these changes in the proteome to important biological pathways.
4. To compare existing methods with new approaches, without bias.

Samples from *M. silvestris*, grown under the substrates methane and acetate, have been analysed both qualitatively and quantitatively by three methods: one-dimensional PAGE, MudPIT incorporating iTRAQ tags, and an LC/MS^E acquisition enabling label-free quantification. In the comparison of two growth conditions, as here, relative quantification provides valuable information regarding specific protein abundance changes. This not only provides insight into the cellular processes of the organism, but also provides a comparison of the information content generated by the different approaches.

This work has been peer reviewed and published:

Patel, Thalassinos *et al.*, (2009) *A comparison of labelling and label-free mass spectrometry-based proteomics approaches*, Journal of Proteome Research **8**(7): 3752-3759

References

- Allet, N., Barrillat, N., Baussant, T., Boiteau, C., Botti, P., Bougueleret, L., Budin, N., Canet, D., Carraud, S., Chiappe, D., Christmann, N., Colinge, J., Cusin, I., Dafflon, N., Depresle, B., Fasso, I., Frauchiger, P., Gaertner, H., Gleizes, A., Gonzalez-Couto, E., Jeandenans, C., Karmime, A., Kowall, T., Lagache, S., Mahé, E., Masselot, A., Mattou, H., Moniatte, M., Niknejad, A., Paolini, M., Perret, F., Pinaud, N., Ranno, F., Raimondi, S., Reffas, S., Regamey, P.-O., Rey, P.-A., Rodriguez-Tomé, P., Rose, K., Rossellat, G., Saudrais, C., Schmidt, C., Villain, M., Zwahlen, C. (2004). In vitro and in silico processes to identify differentially expressed proteins. *Proteomics* **4**: 2333-2351.
- Bienvenut, W. V., Déon, C., Pasquarello, C., Campbell, J. M., Sanchez, J.-C., Vestal, M. L., Hochstrasser, D. F. (2002). Matrix-assisted laser desorption/ionization-tandem mass spectrometry with high resolution and sensitivity for identification and characterization of proteins. *Proteomics* **2**: 868-876.
- Brown, R. S., Lennon, J. J. (2002). Sequence-Specific Fragmentation of Matrix-Assisted Laser-Desorbed Protein/Peptide Ions. *Anal. Chem.* **67**: 3990-3999.
- Cannataro, M. (2008). Computational proteomics: management and analysis of proteomics data. *Brief Bioinform* **9**: 97-101.
- Cargile, B. J., Bundy, J. L., Stephenson, J. L. (2004). Potential for False Positive Identifications from Large Databases through Tandem Mass Spectrometry. *J Proteome Res.* **3**: 1082-1085.
- Chait, B. T. (2006). Mass Spectrometry: Bottom-Up or Top-Down? *Science* **314**: 65-66.
- Chelius, D., Bondarenko, P. V. (2002). Quantitative Profiling of Proteins in Complex Mixtures Using Liquid Chromatography and Mass Spectrometry. *J. Proteome Res.* **1**: 317-323.
- Colinge, J., Chiappe, D., Lagache, S., Moniatte, M., Bougueleret, L. (2005). Differential Proteomics via Probabilistic Peptide Identification Scores. *Anal. Chem.* **77**: 596-606.
- Craig, R., Cortens, J. P., Beavis, R. C. (2005). The use of proteotypic peptide libraries for protein identification. *Rapid Commun. Mass Spectrom.* **19**: 1844-1850.
- de Hoffmann, E., Stroobant, V. (2007). *Mass Spectrometry. Principles and Applications.* Wiley.
- Dedysh, S. N., Knief, C., Dunfield, P. F. (2005). Methylocella Species Are Facultatively Methanotrophic. *J. Bacteriol.* **187**: 4665-4670.

- Domon, B., Aebersold, R.** (2006). Mass Spectrometry and Protein Analysis. *Science* **312**: 212-217.
- Ek, K., Bjellqvist, B., Righetti, P. G.** (1983). Preparative isoelectric focusing in immobilized pH gradients. I. General principles and methodology. *J. Biochem. Biophys. Methods* **8**: 135-155.
- El-Aneed, A., Cohen, A., Banoub, J.** (2009). Mass Spectrometry, Review of the Basics: Electrospray, MALDI, and Commonly Used Mass Analyzers. *Applied Spectroscopy Reviews* **44**: 210 - 230.
- Eng, J. K., McCormack, A. L., Yates, J. R.** (1994). An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **5**: 976-989.
- Fairwell, T., Barnes, W. T., Richards, F. F., Lovins, R. E.** (1970). Sequence analysis of complex protein mixtures by isotope dilution and mass spectrometry. *Biochemistry* **9**: 2260-2267.
- Frank, A. M., Savitski, M. M., Nielsen, M. L., Zubarev, R. A., Pevzner, P. A.** (2007). De Novo Peptide Sequencing and Identification with Precision Mass Spectrometry. *J Proteome Res.* **6**: 114-123.
- Gerber, S. A., Rush, J., Stemman, O., Kirschner, M. W., Gygi, S. P.** (2003). Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proceedings of the National Academy of Sciences of the United States of America* **100**: 6940-6945.
- Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H., Aebersold, R.** (1999). Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotech* **17**: 994-999.
- Han, X., Aslanian, A., Yates III, J. R.** (2008). Mass spectrometry for proteomics. *Curr. Opin. Chem. Biol.* **12**: 483-490.
- Han, X., Jin, M., Breuker, K., McLafferty, F. W.** (2006). Extending Top-Down Mass Spectrometry to Proteins with Masses Greater Than 200 Kilodaltons. *Science* **314**: 109-112.
- Hayter, J. R., Robertson, D. H. L., Gaskell, S. J., Beynon, R. J.** (2003). Proteome Analysis of Intact Proteins in Complex Mixtures. *Mol Cell Proteomics* **2**: 85-95.
- Henzel, W. J., Billeci, T. M., Stults, J. T., Wong, S. C., Grimley, C., Watanabe, C.** (1993). Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases. *Proceedings of the National Academy of Sciences of the United States of America* **90**: 5011-5015.
- Horgan, G. W.** (2007). Sample Size and Replication in 2D Gel Electrophoresis Studies. *J Proteome Res.* **6**: 2884-2887.

- Huttlin, E. L., Hegeman, A. D., Harms, A. C., Sussman, M. R.** (2006). Prediction of Error Associated with False-Positive Rate Determination for Peptide Identification in Large-Scale Proteomics Experiments Using a Combined Reverse and Forward Peptide Sequence Database Strategy. *J Proteome Res.* **6**: 392-398.
- Ishihama, Y., Oda, Y., Tabata, T., Sato, T., Nagasu, T., Rappsilber, J., Mann, M.** (2005). Exponentially Modified Protein Abundance Index (emPAI) for Estimation of Absolute Protein Amount in Proteomics by the Number of Sequenced Peptides per Protein. *Mol Cell Proteomics* **4**: 1265-1272.
- Issaq, H. J., Chan, K. C., Janini, G. M., Conrads, T. P., Veenstra, T. D.** (2005). Multidimensional separation of peptides for effective proteomic analysis. *J. Chromatogr. B* **817**: 35-47.
- Jaffe, J. D., Mani, D. R., Leptos, K. C., Church, G. M., Gillette, M. A., Carr, S. A.** (2006). PEPPeR, a Platform for Experimental Proteomic Pattern Recognition. *Mol. Cell. Proteomics* **5**: 1927-1941.
- Kanehisa, M., Araki, M., Goto, S., Hattori, M., Hirakawa, M., Itoh, M., Katayama, T., Kawashima, S., Okuda, S., Tokimatsu, T., Yamanishi, Y.** (2008). KEGG for linking genomes to life and the environment. *Nucl. Acids Res.* **36**: D480-484.
- Kanehisa, M., Goto, S.** (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucl. Acids Res.* **28**: 27-30.
- Kelleher, N. L., Lin, H. Y., Valaskovic, G. A., Aaserud, D. J., Fridriksson, E. K., McLafferty, F. W.** (1999). Top Down versus Bottom Up Protein Characterization by Tandem High-Resolution Mass Spectrometry. *J. Am. Chem. Soc.* **121**: 806-812.
- King, G.** (1997). Responses of atmospheric methane consumption by soils to global climate change. *Global Change Biol.* **3**: 351-362.
- Lee, S.-W., Berger, S. J., Martinovi, S., Paa-Toli, L., Anderson, G. A., Shen, Y., Zhao, R., Smith, R. D.** (2002). Direct mass spectrometric analysis of intact proteins of the yeast large ribosomal subunit using capillary LC/FTICR. *Proc. Natl. Acad. Sci. USA* **99**: 5942-5947.
- Li, X.-j., Zhang, H., Ranish, J. A., Aebersold, R.** (2003). Automated Statistical Analysis of Protein Abundance Ratios from Data Generated by Stable-Isotope Dilution and Tandem Mass Spectrometry. *Anal. Chem.* **75**: 6648-6657.
- Link, A. J., Eng, J., Schieltz, D. M., Carmack, E., Mize, G. J., Morris, D. R., Garvik, B. M., Yates, J. R.** (1999). Direct analysis of protein complexes using mass spectrometry. *Nat Biotech* **17**: 676-682.
- Liu, H., Sadygov, R. G., Yates, J. R.** (2004). A Model for Random Sampling and Estimation of Relative Protein Abundance in Shotgun Proteomics. *Anal. Chem.* **76**: 4193-4201.

- Makarov, A.** (2000). Electrostatic Axially Harmonic Orbital Trapping: A High-Performance Technique of Mass Analysis. *Anal. Chem.* **72**: 1156-1162.
- Marshall, A. G.** (2000). Milestones in fourier transform ion cyclotron resonance mass spectrometry technique development. *Int. J. Mass spectrom.* **200**: 331-356.
- McLafferty, F. W., Breuker, K., Jin, M., Han, X., Infusini, G., Jiang, H., Kong, X., Begley, T. P.** (2007). Top-down MS, a powerful complement to the high capabilities of proteolysis proteomics. *FEBS J.* **274**: 6256-6268.
- Millea, K. M., Kass, I. J., Cohen, S. A., Krull, I. S., Gebler, J. C., Berger, S. J.** (2005). Evaluation of multidimensional (ion-exchange/reversed-phase) protein separations using linear and step gradients in the first dimension. *J. Chromatogr.* **1079**: 287-298.
- Morris, H. R., Dell, T. P. A., Langhorne, J., Berg, M., Bordoli, R. S., Hoyes, J., Bateman, R. H.** (1996). High Sensitivity Collisionally-activated Decomposition Tandem Mass Spectrometry on a Novel Quadrupole/Orthogonal-acceleration Time-of-flight Mass Spectrometer. *Rapid Commun. Mass Spectrom.* **10**: 889-896.
- Mueller, L. N., Rinner, O., Schmidt, A., Letarte, S., Bodenmiller, B., Brusniak, M.-Y., Vitek, O., Aebersold, R., Müller, M.** (2007). SuperHirn - a novel tool for high resolution LC-MS-based peptide/protein profiling. *Proteomics* **7**: 3470-3480.
- Nemeth-Cawley, J. F., Tangarone, B. S., Rouse, J. C.** (2003). 'Top Down' Characterization Is a Complementary Technique to Peptide Sequencing for Identifying Protein Species in Complex Mixtures. *J Proteome Res.* **2**: 495-505.
- O'Farrell, P. H.** (1975). High resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.* **250**: 4007-4021.
- Olsen, J. V., Nielsen, M. L., Damoc, N. E., Griep-Raming, J., Moehring, T., Makarov, A., Schwarz, J., Horning, S., Mann, M.** (2009). Characterization of the Velos, an Enhanced LTQ Orbitrap, for Proteomics. *Mol Cell Proteomics* **8**: S40-41.
- Ong, S.-E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A., Mann, M.** (2002). Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics. *Mol Cell Proteomics* **1**: 376-386.
- Ow, S. Y., Cardona, T., Taton, A., Magnuson, A., Lindblad, P., Stensjo, K., Wright, P. C.** (2008). Quantitative shotgun proteomics of enriched heterocysts from *Nostoc* sp. PCC 7120 using 8-plex isobaric peptide tags. *J Proteome Res.* **7**: 1615-1628.
- Panchaud, A., Affolter, M., Moreillon, P., Kussmann, M.** (2008). Experimental and computational approaches to quantitative proteomics: Status quo and outlook. *Journal of Proteomics* **71**: 19-33.

- Pandey, A., Mann, M.** (2000). Proteomics to study genes and genomes. *Nature* **405**: 837-846.
- Rappsilber, J., Ryder, U., Lamond, A. I., Mann, M.** (2002). Large-Scale Proteomic Analysis of the Human Spliceosome. *Genome Res.* **12**: 1231-1245.
- Reid, G. E., McLuckey, S. A.** (2002). 'Top down' protein characterization via tandem mass spectrometry. *J. Mass Spectrom.* **37**: 663-675.
- Ross, P. L., Huang, Y. N., Marchese, J. N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhasz, P., Martin, S., Bartlett-Jones, M., He, F., Jacobson, A., Pappin, D. J.** (2004). Multiplexed Protein Quantitation in *Saccharomyces cerevisiae* Using Amine-reactive Isobaric Tagging Reagents. *Mol Cell Proteomics* **3**: 1154-1169.
- Sadowski, P. G., Dunkley, T. P., Shadforth, I. P., Dupree, P., Bessant, C., Griffin, J. L., Lilley, K. S.** (2006). Quantitative proteomic approach to study subcellular localization of membrane proteins. *Nat Protoc* **1**: 1778-1789.
- Silva, J. C., Denny, R., Dorschel, C., Gorenstein, M. V., Li, G.-Z., Richardson, K., Wall, D., Geromanos, S. J.** (2006a). Simultaneous Qualitative and Quantitative Analysis of the *Escherichia coli* Proteome: A Sweet Tale. *Mol. Cell. Proteomics* **5**: 589-607.
- Silva, J. C., Gorenstein, M. V., Li, G.-Z., Vissers, J. P. C., Geromanos, S. J.** (2006b). Absolute Quantification of Proteins by LCMS^E: A Virtue of Parallel MS Acquisition. *Mol. Cell. Proteomics* **5**: 144-156.
- Sirlin, J. L.** (1958). On The Incorporation Of Methionine ³⁵S Into Proteins Detectable By Autoradiography. *J. Histochem. Cytochem.* **6**: 185-190.
- Stead, D. A., Paton, N. W., Missier, P., Embury, S. M., Hedeler, C., Jin, B., Brown, A. J. P., Preece, A.** (2008). Information quality in proteomics. *Brief Bioinform* **9**: 174-188.
- Sun, S., Meyer-Arendt, K., Eichelberger, B., Brown, R., Yen, C.-Y., Old, W. M., Pierce, K., Cios, K. J., Ahn, N. G., Resing, K. A.** (2007). Improved Validation of Peptide MS/MS Assignments Using Spectral Intensity Prediction. *Mol. Cell. Proteomics* **6**: 1-17.
- Syka, J. E. P., Coon, J. J., Schroeder, M. J., Shabanowitz, J., Hunt, D. F.** (2004). Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proceedings of the National Academy of Sciences of the United States of America* **101**: 9528-9533.
- Theisen, A. R., Ali, M. H., Radajewski, S., Dumont, M. G., Dunfield, P. F., McDonald, I. R., Dedysh, S. N., Miguez, C. B., Murrell, J. C.** (2005). Regulation of methane oxidation in the facultative methanotroph *Methylocella silvestris* BL2. *Mol. Microbiol.* **58**: 682-692.

- Ünlü, M., Morgan, M. E., Minden, J. S.** (1997). Difference gel electrophoresis. A single gel method for detecting changes in protein extracts. *Electrophoresis* **18**: 2071-2077.
- Webb-Robertson, B.-J. M., Cannon, W. R.** (2007). Current trends in computational inference from mass spectrometry-based proteomics. *Brief Bioinform* **8**: 304-317.
- Wilkins, M. R., Appel, R. D., Eyk, J. E. V., Chung, M. C. M., Görg, A., Hecker, M., Huber, L. A., Langen, H., Link, A. J., Paik, Y.-K., Patterson, S. D., Pennington, S. R., Rabilloud, T., Simpson, R. J., Weiss, W., Dunn, M. J.** (2006). Guidelines for the next 10 years of proteomics. *Proteomics* **6**: 4-8.
- Wilkins, M. R., Sanchez, J.-C., Gooley, A. A., Appel, R. D., Humphrey-Smith, I., Hochstrasser, D. F., Williams, K. L.** (1996). Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. *Biotechnol. Genet. Eng. Rev.* **13**: 19-50.
- Yates, J. R., Cociorva, D., Liao, L., Zabrouskov, V.** (2006). Performance of a Linear Ion Trap-Orbitrap Hybrid for Peptide Analysis. *Anal. Chem.* **78**: 493-500.
- Yost, R. A., Enke, C. G.** (1978). Selected ion fragmentation with a tandem quadrupole mass spectrometer. *J. Am. Chem. Soc.* **100**: 2274-2275.
- Zhang, Y., Ficarro, S. B., Li, S., Marto, J. A.** (2009). Optimized Orbitrap HCD for Quantitative Analysis of Phosphopeptides. *J. Am. Soc. Mass Spectrom.* **20**: 1425-1434.
- Zhang, Z.** (2005). Prediction of Low-Energy Collision-Induced Dissociation Spectra of Peptides with Three or More Charges. *Anal. Chem.* **77**: 6364-6373.
- Zubarev, R. A., Kelleher, N. L., McLafferty, F. W.** (1998). Electron Capture Dissociation of Multiply Charged Protein Cations. A Nonergodic Process. *J. Am. Chem. Soc.* **120**: 3265-3266.

Chapter 9

Materials and Methods

Mass Spectrometry-Based Proteomics

9.1 Sample preparation

9.1.1 Bacterial growth and sample preparation

This was performed with Andrew Crombie in the lab of Prof. J. Colin Murrell. *Methylocella silvestris* was grown in fermenter cultures on diluted nitrate mineral salts (NMS) medium with methane or acetate (5 mM) as previously described (Theisen *et al.*, 2005). Cells, grown to late exponential phase ($OD_{540} \sim 1.0$), were harvested by centrifugation ($17,700 \times g$, 20 min, $4^{\circ}C$), washed in growth medium, resuspended in 0.1 M PIPES buffer (piperazine-N,N'-bis[2-ethanesulfonic acid], pH 7.0), and frozen in liquid nitrogen. Subsequently, frozen cells were thawed and resuspended in PIPES buffer containing 1 mM benzamidine and broken by four passes through a French pressure cell at 125 MPa ($4^{\circ}C$) (American Instrument Co., Silver Spring, MD). Cell debris and membranes were removed by two centrifugation steps ($13,000 \times g$, 30 min, $4^{\circ}C$, followed by $140,000 \times g$, 90 min, $4^{\circ}C$), and the supernatant, containing soluble cytoplasmic proteins, used for analysis. A protein assay was conducted on the soluble extract, using a Micro BCA Protein Assay Kit (Pierce Protein Research Products, Thermo Scientific, Cramlington, UK) according to the manufacturer's protocol.

9.1.2 Protein separation by gel electrophoresis

Proteins were resolved by 1D SDS-PAGE (14 μg per lane) and stained with Coomassie Blue. 30 to 40 slices were excised from each lane, and subjected to tryptic digestion. All processing of the gel plugs was performed by a MassPrep robotic protein handling system (Waters Corporation, Manchester, UK) using the manufacturer's protocol. In brief, the gel plugs were destained, the disulfide bonds were reduced by the addition of dithiothreitol and the free cysteine residues were alkylated with iodoacetamide. The gel plugs were washed prior to a dehydration step, followed by the addition of trypsin, and incubated for 4.5 hours. The resultant tryptic peptides were extracted up to two times in total and transferred to a cooled 96-well microtitre plate; if necessary, they were stored at $-20^{\circ}C$.

9.1.3 iTRAQ labelling and strong cation exchange chromatography (SCX)

Labelled quantification was carried out using the iTRAQ 4-plex labelling kit (Applied Biosystems, Warrington, UK). Protein extracts from the two growth conditions were digested and labelled according to the manufacturer's standard protocol, and the samples pooled and lyophilised. A total of 400 µg protein from each growth condition was labelled, giving a total protein loading of 800 µg. As SCX was carried out offline, the potential for sample losses is higher. A larger initial protein loading was therefore used in order to minimise such losses and optimise the number of proteins identified by this approach. 200 µg of acetate-grown sample was labelled with the 114 reporter tag, and 200 µg with the 116 reporter tag. 200 µg of the methane-grown sample was labelled with the 115 tag and 200 µg with the 117 tag. As per the manufacturer's protocol, a maximum of 100 µg of protein was labelled per vial of iTRAQ label, i.e. two vials were used per label. The labelling of one growth condition with two different iTRAQ tags provides the means for an internal control to monitor labelling efficiency. The labelled tryptic peptides were partially resolved using a PolySULFOETHYL A SCX column, 2.1 mm × 20 cm, 5 µm particles, 300 Å pore size (PolyLC, Columbia, USA), using a stepwise gradient of KCl, adapted from Link *et al.*, (Link *et al.*, 1999) from 2.5–50% salt solution over a period of 75 minutes. In total, 64 fractions were collected.

9.1.4 In-solution tryptic digestion

100 µg of soluble protein extract was resuspended in 1 mL of 0.1% Rapigest (Waters Corporation, Milford, MA) and concentrated using a 5 kDa cut-off spin column. The solution was then heated at 80°C for 15 minutes, reduced with DTT at 60°C for 15 minutes, alkylated in the dark with iodoacetamide at ambient temperature for 30 minutes, and digested with 1:50 (w/w) sequencing grade trypsin (Promega, Southampton, UK) for 16 hours. RapiGest was hydrolysed by the addition of 2 µL 15 M HCl, centrifuged, and each sample diluted 1:1 with a 50 fmol/µl glycogen phosphorylase B standard tryptic digest to give a final protein concentration of 500 ng/µl per sample and 25 fmol/µl phosphorylase B.

9.2 Mass spectrometry

9.2.1 LC-MS/MS acquisition for gel-separated samples

Peptides extracted from the digested gel were transferred to a nanoACQUITY system (Waters Corporation). A 6.4 μl aliquot of extract was mixed with 13.6 μl of 0.1% formic acid and loaded onto a 0.5 cm LC Packings C18 5 μm 100 \AA 300 μm i.d μ -precolumn cartridge. Flushing the column with 0.1% formic acid desalted the bound peptides before a linear gradient of solvent B (0.1% formic acid in acetonitrile) at a flow rate of approximately 200 nl/min eluted the peptides for further resolution on a 15 cm LC Packings C18 5 μm 5 \AA 75 μm i.d. PepMap analytical column. The eluted peptides were analysed on a Micromass Q-ToF Global Ultima (Waters Corporation) mass spectrometer fitted with a nano-LC sprayer with an applied capillary voltage of 3.5 kV. The spectral acquisition scan rate was 1.0 s with a 0.1 s interscan delay. The instrument was calibrated against a collisionally induced dissociation (CID) spectrum of the doubly charged precursor ion of [Glu¹]-fibrinopeptide B (GFP – Sigma Aldrich, St. Louis, USA), and fitted with a GFP lockspray line. The instrument was operated in data dependent acquisition (DDA) mode over the mass/charge (m/z) range of 50-2000. During the DDA analysis, CID experiments were performed on the three most intense, multiply charged peptides as they eluted from the column at any given time. Once these data have been collected, the next three most intense peptides are selected, and this process repeated.

9.2.2 LC-MS/MS acquisition for iTRAQ samples

Fractions collected from the SCX separation of iTRAQ-labelled peptides were snap-frozen on dry ice and lyophilised to dryness. The samples were resuspended in 20 μl 0.1% formic acid and transferred to a CapLC system (Waters Corporation). A 6.4 μl aliquot of extract was mixed with 13.6 μl of 0.1% formic acid and loaded onto a 0.5 cm LC Packings C18 5 μm 100 \AA 300 μm i.d precolumn cartridge. Flushing the column with 0.1% formic acid desalted the bound peptides before a linear gradient of solvent B (0.1% formic acid in acetonitrile) at a flow rate of approximately 200 nl/min eluted the peptides for further resolution on a 15 cm LC Packings C18 5 μm 5 \AA 75 μm i.d. PepMap analytical column. The eluted peptides were analysed on a Micromass Q-ToF Global Ultima (Waters Corporation) mass spectrometer fitted with

a nano-LC sprayer with an applied capillary voltage of 3.5 kV. The spectral acquisition scan rate was 1.0 s with a 0.1 s interscan delay. The instrument was calibrated against a CID spectrum of the doubly charged precursor ion of GFP, and fitted with a GFP lockspray line. The instrument was operated in data dependent acquisition (DDA) mode as described above.

9.2.3 LC-MS configurations for label-free analysis

Nanoscale LC separations of tryptic peptides for qualitative and quantitative multiplexed LC-MS analysis were performed with a nanoACQUITY system (Waters Corporation) using a Symmetry C₁₈ trapping column (180 µm x 20 mm 5 µm) and a BEH C₁₈ analytical column (75 µm x 250 mm 1.7 µm). The composition of solvent A was 0.1% formic acid in water, and solvent B, 0.1% formic acid in acetonitrile. Each sample (total protein 0.5 µg) was applied to the trapping column and flushed with 1% solvent B for 5 minutes at a flow rate of 15 µL/min. Sample elution was performed at a flow rate of 300 nL/min by increasing the organic solvent concentration from 3 to 40% B over 90 min. All analyses were conducted in triplicate. The precursor ion masses and associated fragment ion spectra of the tryptic peptides were mass measured with a Q-ToF Premier mass spectrometer (Waters, UK) directly coupled to the chromatographic system.

The time-of-flight analyzer of the mass spectrometer was externally calibrated with NaI from m/z 50 to 1990, with the data post-acquisition lockmass-corrected using the monoisotopic mass of the doubly charged precursor of GFP, fragmented with a collision energy of 25V. The GFP was delivered at 500 fmol/µL to the mass spectrometer via a NanoLockSpray interface using the auxiliary pump of a nanoACQUITY system at a flow rate of 500 nL/min. The reference sprayer was sampled every 60 seconds.

Accurate mass data were collected in data independent mode of acquisition (LC-MS^E) by alternating the energy applied to the collision cell between a low energy and elevated energy state. The spectral acquisition scan rate was 0.6 s with a 0.1 s interscan delay. In the low energy MS mode, data were collected at constant collision energy of 4 eV. In elevated energy MS mode, the collision energy was ramped from 15 eV to 35 eV during each integration.

9.3 Data processing and database searching

9.3.1 Data processing for DDA acquisitions

The uninterpreted MS/MS data from the gel-separated and iTRAQ-labelled samples were processed using ProteinLynx Global Server (PLGS) v2.3 (Waters, UK). The data were smoothed, background subtracted, centred and deisotoped. All data were lockspray calibrated against GFP using data collected from the reference line during acquisition.

9.3.2 Data processing for label-free acquisitions

The LC-MS^E data were processed using PLGS v2.3. The ion detection, data clustering and normalisation of the data independent, alternate scanning LC-MS^E data has been explained in detail elsewhere (Geromanos *et al.*, 2009). In brief, lockmass-corrected spectra are centroided, deisotoped, and charge-state-reduced to produce a single accurately mass measured monoisotopic mass for each peptide and the associated fragment ion. The initial correlation of a precursor and a potential fragment ion is achieved by means of time alignment.

9.3.3 Database searches

All data were searched using PLGS v2.3 against a *Methylocella silvestris* database (<http://genome.ornl.gov/microbial/msil>). Fixed modification of carbamidomethyl-C was specified, and variable modifications included were acetyl N-terminus, deamidation N, deamidation Q and oxidation M. For the iTRAQ data, variable modifications for the isobaric tags were specified. One missed cleavage site was allowed. Search parameters specified were a 50 ppm tolerance against the database-generated theoretical peptide ion masses and a minimum of one matched peptide.

For the LC-MS^E data, the time-based correlation applied in data processing was followed by a further correlation process during the database search that is based on the physicochemical properties of peptides when they undergo collision induced fragmentation (Li *et al.*, 2009). The precursor and fragment ion tolerances were determined automatically. The protein identification criteria also included the

detection of at least three fragment ions per peptide, at least one peptide determined per protein and the identification of the protein in at least two out of three technical replicates. By using protein identification replication as a filter, the false positive rate is minimised, as false positive protein identifications, i.e. chemical noise, have a random nature and as such do not tend to replicate across injections. This approach rules out systematic search events errors due to the repeated ambiguity of a particular spectrum and the subsequent sequence assignment by a search algorithm, as could be the case with peptide-centric searches. An overview of the processing and searching of MS^E data is given in Figure 9.1.

9.4 Protein quantification

9.4.1 Protein quantification using iTRAQ labelling

PLGS was also used for quantitative evaluation of MS/MS data generated from the analysis of the iTRAQ-labelled peptides. A relative quantification was conducted using a merged dataset comprising the results from the database search. Concentration ratios of iTRAQ-labelled proteins were calculated based on signal intensities of reporter ions observed in peptide fragmentation spectra, with the relative areas of the peaks corresponding to proportions of the labelled peptides (Ross *et al.*, 2004).

9.4.2 Protein quantification using label-free system

Relative quantitative analysis across conditions was performed by comparing normalised peak area/intensity of each identified peptide. Normalisation of the data was conducted by the use of an internal protein digest standard. In brief, peak areas/intensities are corrected using those of the internal protein digest.

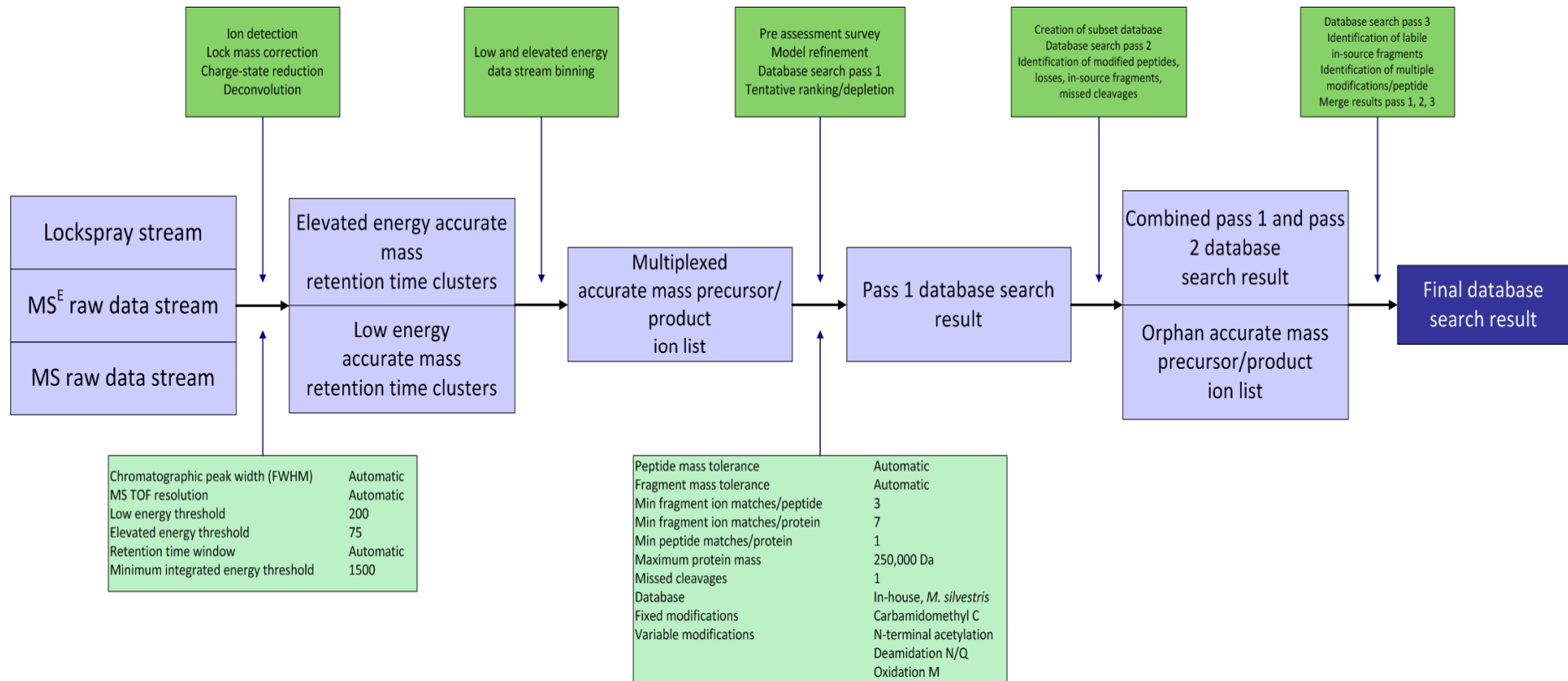


Figure 9.1 – An overview of the data processing and database searching workflow from an MS^E acquisition (adapted from Li et al., 2009).

Intensity measurements are typically further adjusted on those components, that is de-isotoped and charge-state reduced accurate mass retention time pairs, that replicate throughout the complete experiment. Next, the redundant, proteotypic quantitative measurements provided by the multiple tryptic peptide identification from each protein were used to determine an average, relative protein fold-change. The algorithm performs binary comparisons for each of the conditions to generate an average normalised intensity ratio for all matched proteins. Proteins with a likelihood of quantification smaller than 0.05 were considered to be significantly regulated. The entire data set of differentially expressed proteins was further filtered by considering only the identified peptides that replicated two out of three technical instrument replicates. A likelihood of regulation higher than 95%, as reported by the quantification algorithm, was considered.

9.5 Mapping protein identifications onto biological pathways

Proteins identified from each growth condition (acetate and methane) were placed in separate FASTA-formatted files. The FASTA files were used as an input for the KEGG Automatic Annotation Server (KAAS) program 36. KAAS performs a homology search to associate query proteins to KEGG GENES database identifiers. The list of these KEGG gene identifiers was used by an in-house Perl script to access the KEGG system, using the KEGG API (application programming interface), and map these onto KEGG biochemical pathways. The output of the script is a URL which points to an on-line interactive image of a KEGG pathway with proteins from different conditions highlighted in different colours. Since *Methylocella silvestris*-specific pathways are not comprehensively catalogued in the KEGG database, proteins were mapped onto those available.

References

- Geromanos, S. J., Vissers, J. P. C., Silva, J. C., Dorschel, C. A., Li, G.-Z., Gorenstein, M. V., Bateman, R. H., Langridge, J. I.** (2009). The detection, correlation, and comparison of peptide precursor and product ions from data independent LC-MS with data dependant LC-MS/MS. *Proteomics* **9**: 1683-1695.
- Li, G.-Z., Vissers, J. P. C., Silva, J. C., Golick, D., Gorenstein, M. V., Geromanos, S. J.** (2009). Database searching and accounting of multiplexed precursor and product ion spectra from the data independent analysis of simple and complex peptide mixtures. *Proteomics* **9**: 1696-1719.
- Link, A. J., Eng, J., Schieltz, D. M., Carmack, E., Mize, G. J., Morris, D. R., Garvik, B. M., Yates, J. R.** (1999). Direct analysis of protein complexes using mass spectrometry. *Nat Biotech* **17**: 676-682.
- Ross, P. L., Huang, Y. N., Marchese, J. N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhasz, P., Martin, S., Bartlet-Jones, M., He, F., Jacobson, A., Pappin, D. J.** (2004). Multiplexed Protein Quantitation in *Saccharomyces cerevisiae* Using Amine-reactive Isobaric Tagging Reagents. *Mol Cell Proteomics* **3**: 1154-1169.
- Theisen, A. R., Ali, M. H., Radajewski, S., Dumont, M. G., Dunfield, P. F., McDonald, I. R., Dedysh, S. N., Miguez, C. B., Murrell, J. C.** (2005). Regulation of methane oxidation in the facultative methanotroph *Methylocella silvestris* BL2. *Mol. Microbiol.* **58**: 682-692.

Chapter 10

Results

Mass Spectrometry-Based Proteomics

10.1 Protein identifications

Three distinct experimental approaches were employed in order to provide profiling and quantitative information regarding the proteome of *M. silvestris*.

The gel-based analysis identified 331 proteins and 202 protein identifications, for the acetate and methane growth conditions respectively. In total, the gel-based analysis provided 389 non-redundant identifications. The iTRAQ analysis uses a pooled sample, so both growth conditions were placed together. The total number of protein identifications was 384. As with the gel-based experiments, the MS^E method keeps the growth conditions separate. Filtering out those identifications only seen in one of the three replicates, 355 and 194 identifications were made for the acetate and methane growth conditions respectively. This corresponded to 425 non-redundant protein identifications.

The numbers of proteins identified via each approach are summarised in Table 10.1. The total number of non-redundant proteins identified is comparable for all three techniques at 389, 384 and 425 proteins respectively. Differences arise, however, when looking at the number of peptides per protein identification. There have been questions raised in the literature regarding the validity of identifications performed using a single peptide, so-called ‘one-hit wonders’, and whether they should be included in the list of proteins identified (Veenstra *et al.*, 2004). Of the gel separation identifications, 133 from the acetate-grown sample and 93 from the methane-grown sample were made using a single peptide; this translates as 40% and 46% respectively. Overall, 154 of the 389 non-redundant identifications were single-peptide assignments, which is 40%. The iTRAQ data contained 208 single-peptides from the total of 384, proportionally 54%. For the MS^E analysis, two identifications from each growth condition were identified with a single peptide; this represents 0.6% for acetate and 1% for methane. The proportions of single-peptide identifications for the gel-based and iTRAQ analyses are typical of many results in the literature (Breci *et al.*, 2005). In the label-free results, of the 425 identifications, only 4 are from a single peptide: proportionally less than 1%. As the label-free analysis is performed in triplicate, only an identification observed in at least two of the three replicates was taken to be valid; therefore, single-peptide identification in label-free data means that a single peptide was found in at least two of three data sets.

	Gel-based	iTRAQ	Label-free
Total	389	384	425
Single-peptide	154	202	4
Proportion based on single hit	40%	54%	0.9%
More than one peptide	235	178	421

Table 10.1 - Total protein identifications for the three experimental approaches

Proteins identified by each experimental setup are listed in Appendix 1, Tables 1-5. Figure 10.1 shows the overlap of protein identifications between the three approaches; 10.1(a) uses all data, including single-peptide identifications, 10.1(b) illustrates filtered data, with only identifications obtained with two or more peptides. All proteins identified are listed in Supplementary Tables 1 to 5, giving information on the molecular weight and pI of the identifications, and also the number of peptides identified. When including single peptide based identifications, there are a total of 699 proteins identified. Each of the techniques uniquely provides approximately 17% of those identifications. The remaining 49% of the identifications overlapped as shown. To overcome the uncertainty involved in the inclusion of single peptide-based identification, Figure 10.1(b) shows the data presented only including identifications made using a minimum of two peptides. This gives a total of 509 protein identifications, of which 9% were unique to the gel-based approach, 6% to iTRAQ, and 38% to label-free. This shows a significant increase in the proportion of unique identifications by the label-free method, a reduction in gel-based unique identifications and a considerable decrease in those uniquely identified by iTRAQ.

A closer inspection of the number of proteins identified with and without the inclusion of single-peptide identifications reveals some interesting observations. As one would expect, the total number of proteins identified is lower when single-peptide identifications are excluded (509 when excluded, compared to 699 when included), including those identifications common to all three methods (89 when excluded, compared to 152 when included). In contrast, the number of proteins unique to the label-free method, and those common to the label-free and gel methods, has increased.

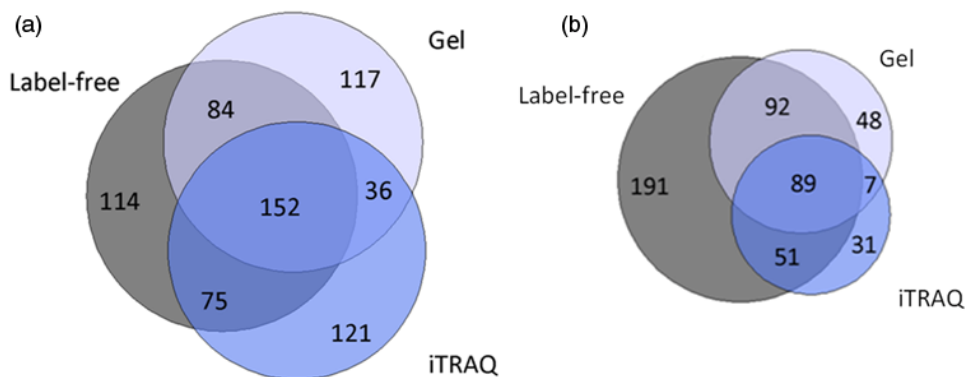


Figure 10.1 – Number of proteins identified by the various experimental approaches, (a) including single-peptide identifications, (b) based on a minimum of two peptides.

This is because all but four of the proteins identified by the label-free method were done so with two or more peptides, whereas the gel and iTRAQ methods generated a large number of single-peptide identifications. The fact that 152 proteins were independently identified by all three methods provides strong evidence that although some of these (63 in total) were identified with a single-peptide by one or more technique, they should possibly not be discarded as false-positive identifications. This raises the questions as to what should be done with protein identifications based on a single-peptide. While the majority of these are likely to correspond to false-positive identifications, there are a small number that are potentially valid and should also be included in the list of confidently-identified proteins, although this is not definitive. Further discussions of results will therefore exclude single-peptide identifications.

10.2 Relative quantification of identified proteins

10.2.1 Gel-based approach

Figure 10.2 shows the 1D SDS-PAGE separation of the soluble *M. silvestris* proteome from different growth conditions. By looking at the image, differential expression can be identified in certain protein bands, and some representative changes are highlighted. Most notably, the key methane oxidation enzyme, sMMO, can be seen as a darkly stained band in the methane-grown sample, as indicated, but

its expression is significantly reduced in the acetate-grown sample. This is very similar to previous work (Theisen *et al.*, 2005). Although the analysis of gel-separated samples provided a comparable number of protein identifications, quantitative analysis using a 1D separation is difficult. Quantitation via gel methods is more routinely performed using two-dimensional separations, which were not carried out here. Further results, focussing on differential expression, use only MudPIT and MS^E data.

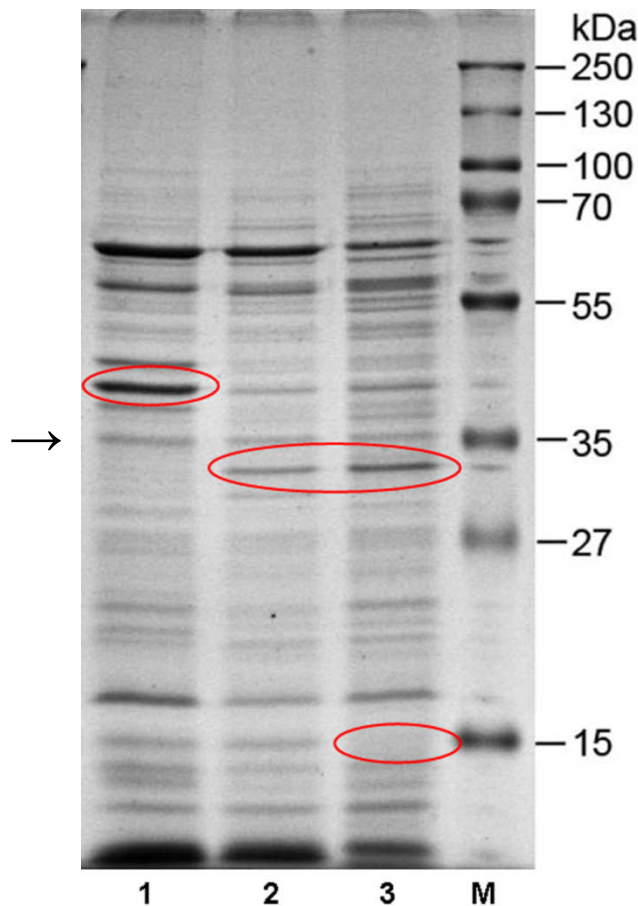


Figure 10.2 - 1D SDS-PAGE separation of the *M. silvestris* proteome under different growth conditions; 1 - methane-grown, 2- succinate-grown (further data not shown), 3 - acetate-grown, M - molecular weight markers. The areas marked in red indicate differences between the growth conditions which can be observed by eye. The band indicated with an arrow corresponds to the enzyme soluble methane monooxygenase (sMMO); corresponding bands in the succinate and acetate growth conditions are more faintly stained, implying lower levels of expression.

10.2.2 iTRAQ labelling

Figure 10.3 represents the differential expression of proteins as characterised by iTRAQ labelling, plotted on a \log_e scale; the values are included alongside protein identifications in Appendix 1, Table 3. Tags 115 and 117, which correspond to methane-grown samples, and tag 116, which corresponds to an acetate-grown sample, were normalised to tag 114, which corresponds to an acetate-grown sample. The values for the 116 sample are clustered close to a line along the x-axis as would be expected since the 114 and 116 samples should be identical. The 115 and 117 samples should also be identical and we would therefore expect good agreement between their ratios, as is observed. This experiment provides a good indication of the reproducibility of the iTRAQ approach. As can be seen by the 115 and 117 trends, distinct up- and down-regulated proteins may be identified in *M. silvestris* when grown under methane as compared to when the organism is grown under acetate. The standard deviation of all the 116:114 ratios is 0.17, providing an indication of what can be considered true up- or down-regulation. If these values are considered to be a normal distribution around a calculated mean of 0, then any proteins with 115:114 and 117:114 ratios within -0.5 and 0.5 cannot be said to be regulated, using the value of three standard deviations to provide filtering parameters. Only those identifications showing ratios outside these values have been accepted as up- or down-regulated. In the iTRAQ method, samples from different growth conditions are pooled together. Quantification depends entirely upon the isobaric tags; if insufficient data is available from the isobaric tags, the protein identification will still be provided in the overall results table, but will not appear in the quantification data. For the dataset presented here, quantification data was available for all confidently identified proteins.

10.2.3 Label-free MS^E quantification

In the label-free system, samples from differing growth conditions are kept separate, so a distinct set of protein identifications is generated for each sample. 231 proteins were identified unique to the acetate-grown sample, 70 were unique to the methane-grown sample and 124 were common to both conditions.

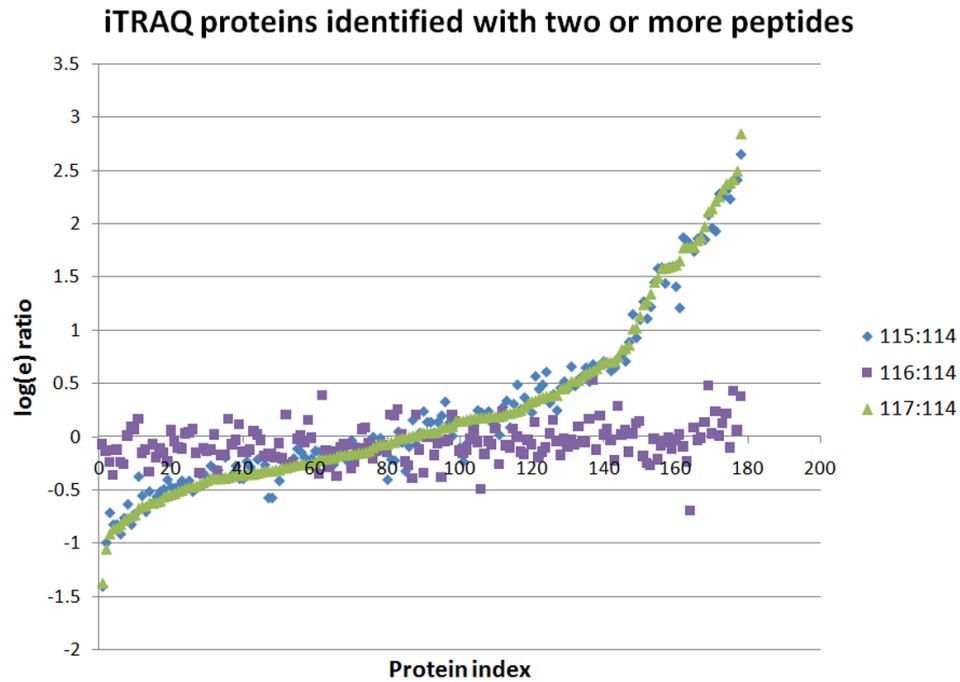


Figure 10.3 - Differential expression as determined by iTRAQ labelling; all tags have been normalised to the 114 label.

Data relating to these 124 proteins were then processed to provide information on relative expression between the samples.

Figure 10.4 shows the relative protein expression for the regulated proteins (common to acetate and methane substrates) identified using the label-free approach; this is the output from the relative quantification software, which generates peptide signal intensity measurements, using all the peptides identified for any particular protein identification. These represent deisotoped, charge-state reduced and accurately mass measured ion lists, which are used for both qualitative identification and relative quantification (Silva *et al.*, 2006). $\text{Log}_{(e)}$ values used as the quantitative measurement can be found in Appendix 1, Table 6, including indication of proteins assigned to only one of the two growth conditions. Error measurements are automatically generated as standard deviation values, which have been plotted. For an MS^E acquisition, the technical variation with respect to signal intensity has been shown to be 10-15% with highly consistent reproducibility (Silva *et al.*, 2005; Vissers *et al.*, 2007). For the label-free quantitative data, the significance level of regulation was determined at 30% fold change, which is an average relative fold

change between -0.3 and 0.3 on a natural log scale (Vissers *et al.*, 2007). This is typically 2-3 times higher than the estimated error on the intensity measurement (Chambery *et al.*, 2009; Silva *et al.*, 2006; Vissers *et al.*, 2007). Those identifications with relative expression values between -0.3 and 0.3 cannot be taken as regulated; only those identifications outside these values can be said to be regulated.

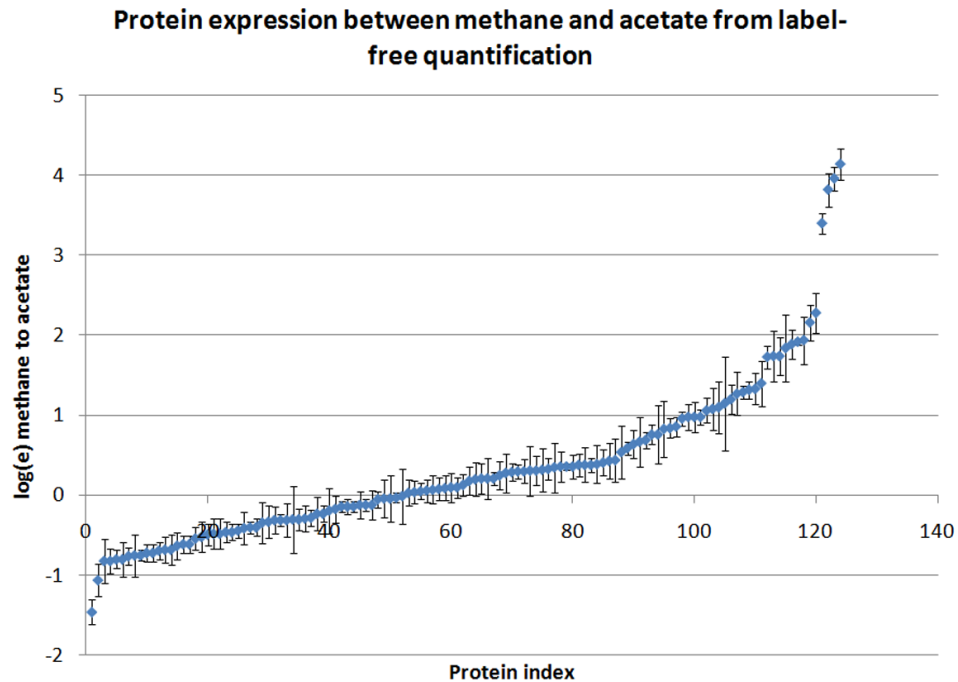


Figure 10.4 - Automated protein-level quantification of regulated proteins using the label-free system; error bars correspond to the automatically generated standard deviation values.

10.2.4 Agreement of quantification approaches

Both iTRAQ and label-free allow profiling and relative quantitative data to be concurrently collected. The ability to do this, particularly in a high throughput manner, is desirable but often difficult. In total, 79 confident identifications (i.e. more than one peptide) are common between the two methods, which is a much larger overlap than previous studies comparing methods of quantification (Ross *et al.*, 2004; Usaite *et al.*, 2008a). A scatter plot comparing the regulation as assessed by the two methods is shown in Figure 10.5a. There is reasonable correlation, with an

R^2 value of 0.69, with one distinct outlier. If the overall trend of regulation is compared, all of the common identifications for which quantification data is available are in agreement, bar the outlier. 21 proteins are indicated to be up-regulated in the methane sample compared to acetate and 6 are indicated to be down-regulated; the remaining proteins show no distinct differential expression when filters for both datasets are applied. If the one distinct outlier is removed from the dataset then the correlation improves significantly (R^2 value 0.80), as shown in Figure 10.5b.

The outlier, corresponding to the enzyme citrate synthase, presented down-regulation in the methane-grown sample according to the label-free analysis, but up-regulation according to iTRAQ. Interrogation of the raw data showed good correlation between all three replicates of the label-free acquisition in both growth conditions. In the iTRAQ data, however, there was a disparity in the data from the isobaric tags. Five peptides were used for identification, with quantification data available for four of these. Three peptides showed down-regulation in the methane growth condition; the one peptide which indicated up-regulation was the shortest of the five (four residues), the others matching at least eight residues within the assigned MS/MS spectrum. If the short peptide is removed, there is down-regulation of citrate synthase within the filtering parameters, and in-line with the label-free data, suggesting that this was a mis-assignment by the software. MS/MS spectra of the matched sequences and isobaric tags are shown in Figure 10.6. Although this is only one anomalous data point, it indicates potential problems if single-peptide identifications are used to provide quantitative data from an iTRAQ experiment.

The label-free approach differs from iTRAQ in that each growth condition is analysed independently, while in iTRAQ samples from different conditions are pooled together. Of the 425 non-redundant identifications obtained by the label-free method, 231 were unique to the acetate-grown sample and 70 unique to the methane-grown. From these 301 proteins, 54 were also identified by iTRAQ, and were compared with the iTRAQ quantification list. Of these, 25 were distinctly regulated and all showed agreement, i.e. were shown by iTRAQ to be up-regulated in whichever growth condition the label-free method had exclusively assigned. This has been represented as a comparative table in Appendix 1, Table 7. The 29

identifications which fall outside the iTRAQ filtering parameters for accepted regulation levels, as described earlier, have been highlighted.

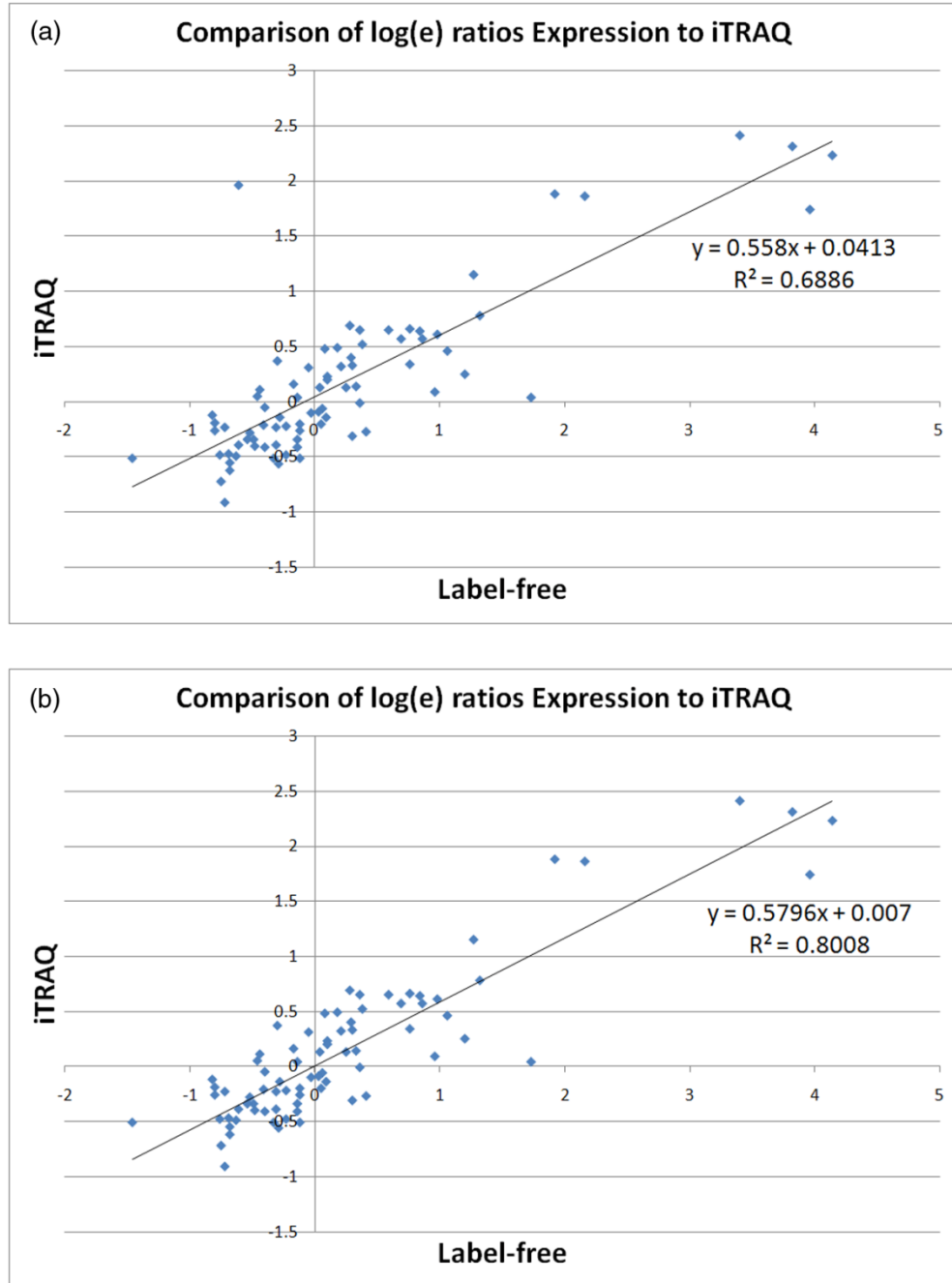


Figure 10.5 - (a) Correlation of quantification data from iTRAQ and the label-free method for identifications using two or more peptides; **(b)** Correlation when the outlier corresponding to citrate synthase is removed.

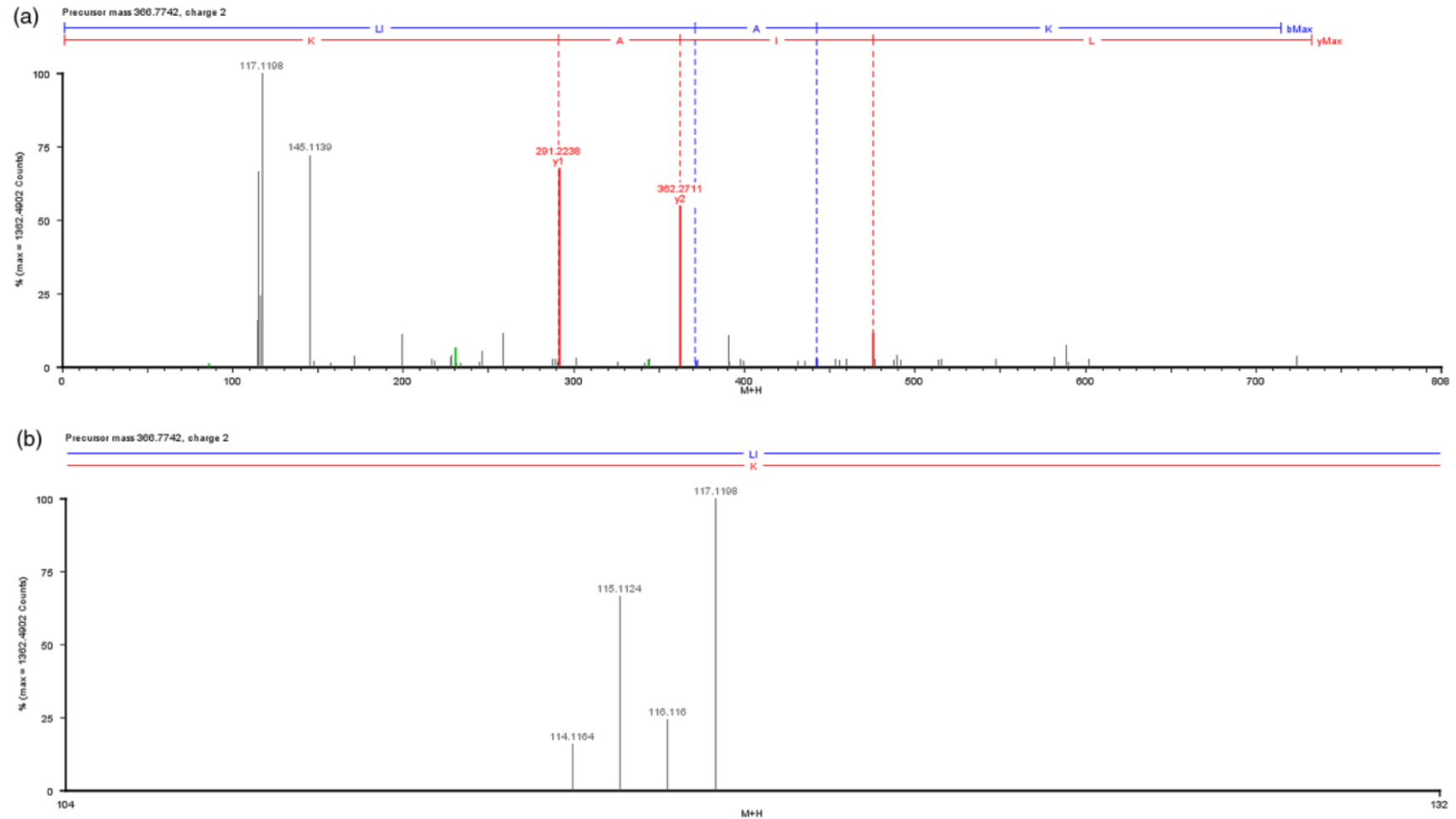


Figure 10.6 - Peptide sequences for precursors (all doubly-charged ions) and corresponding isobaric reporter ions for outlying protein identification citrate synthase **(a)** Mis-assigned peptide, precursor m/z 366, **(b)** reporter ions for precursor 366

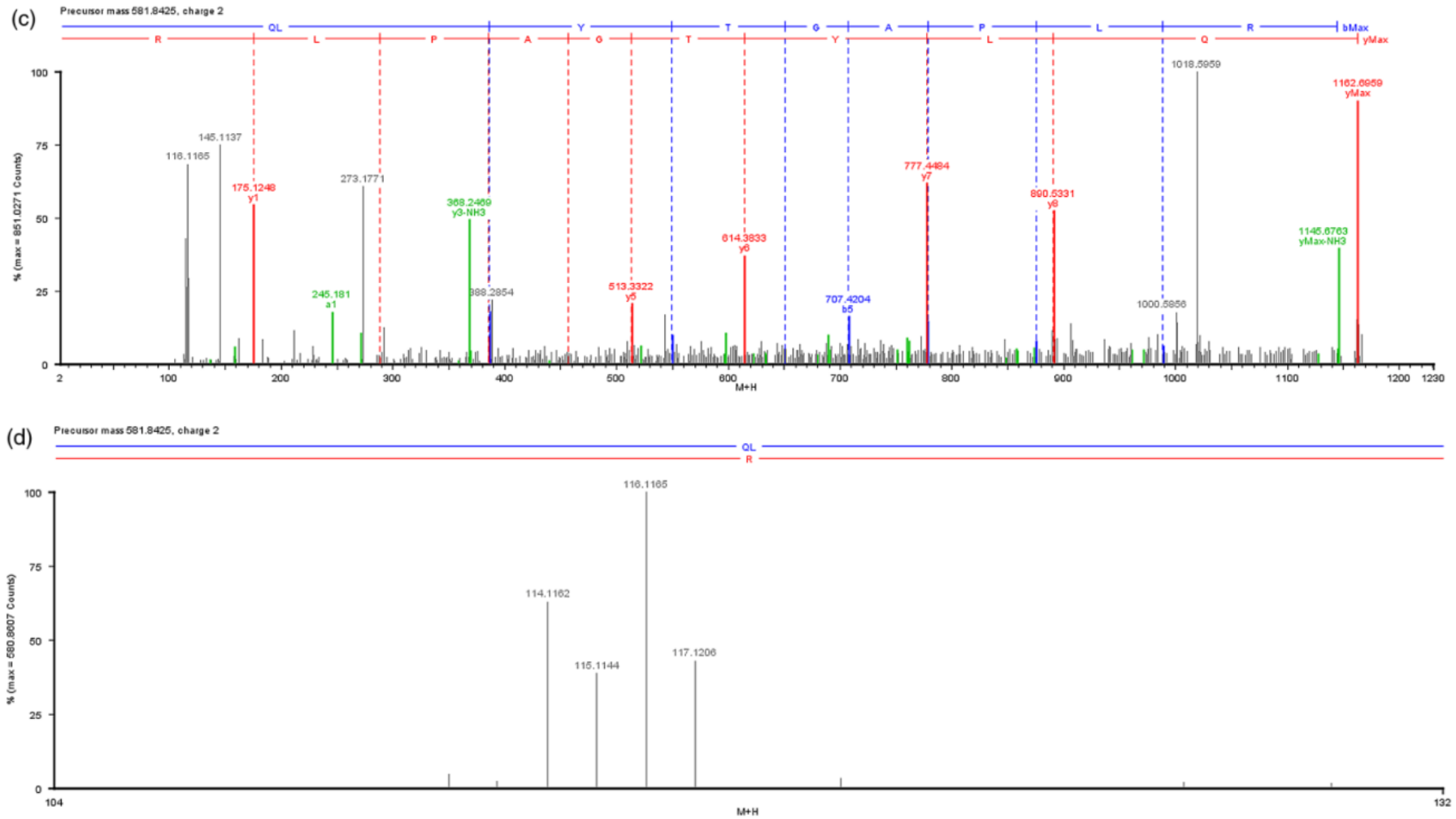
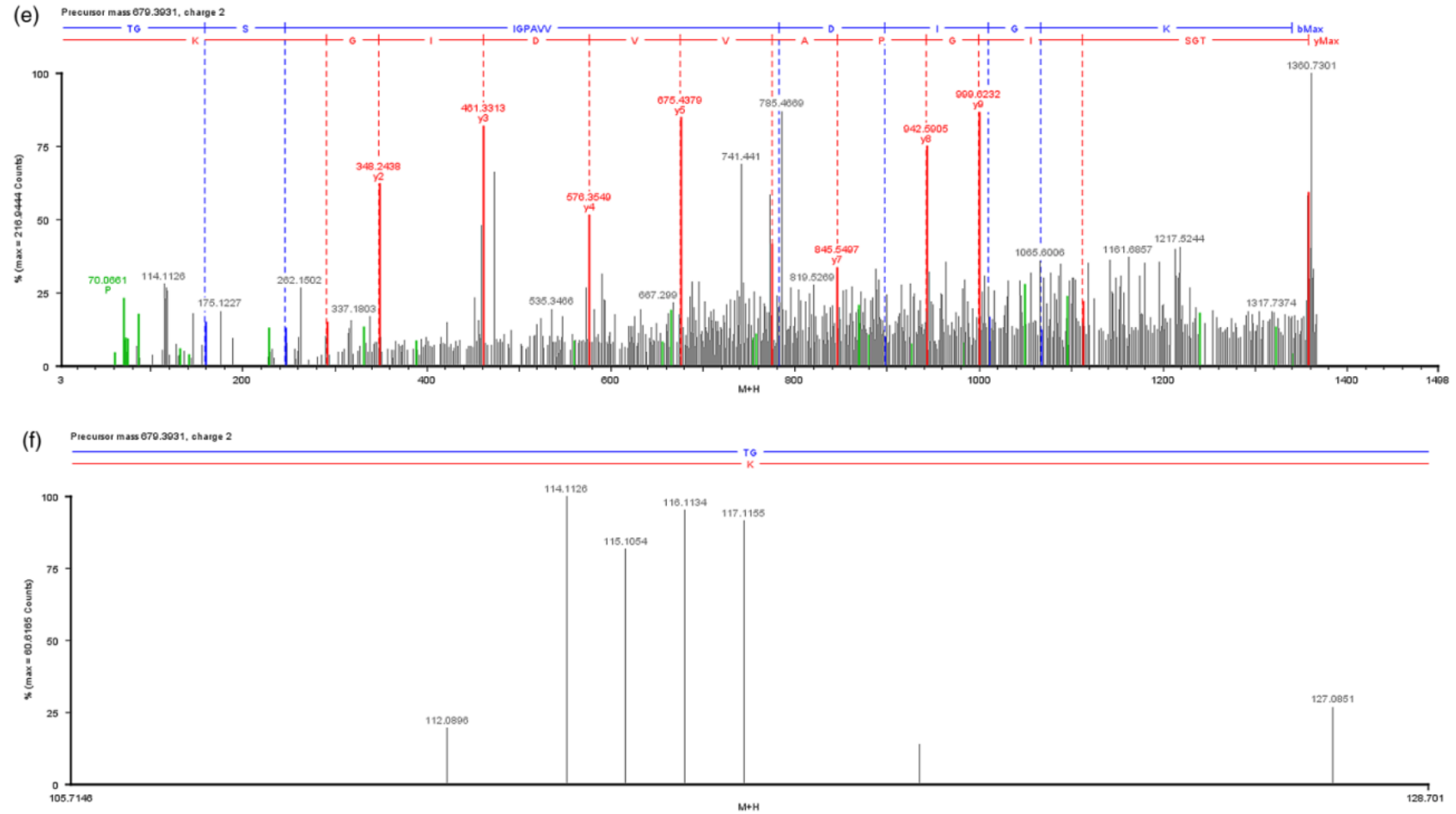


Figure 10.6 - (c) precursor m/z 581, (d) reporter ions for precursor 581



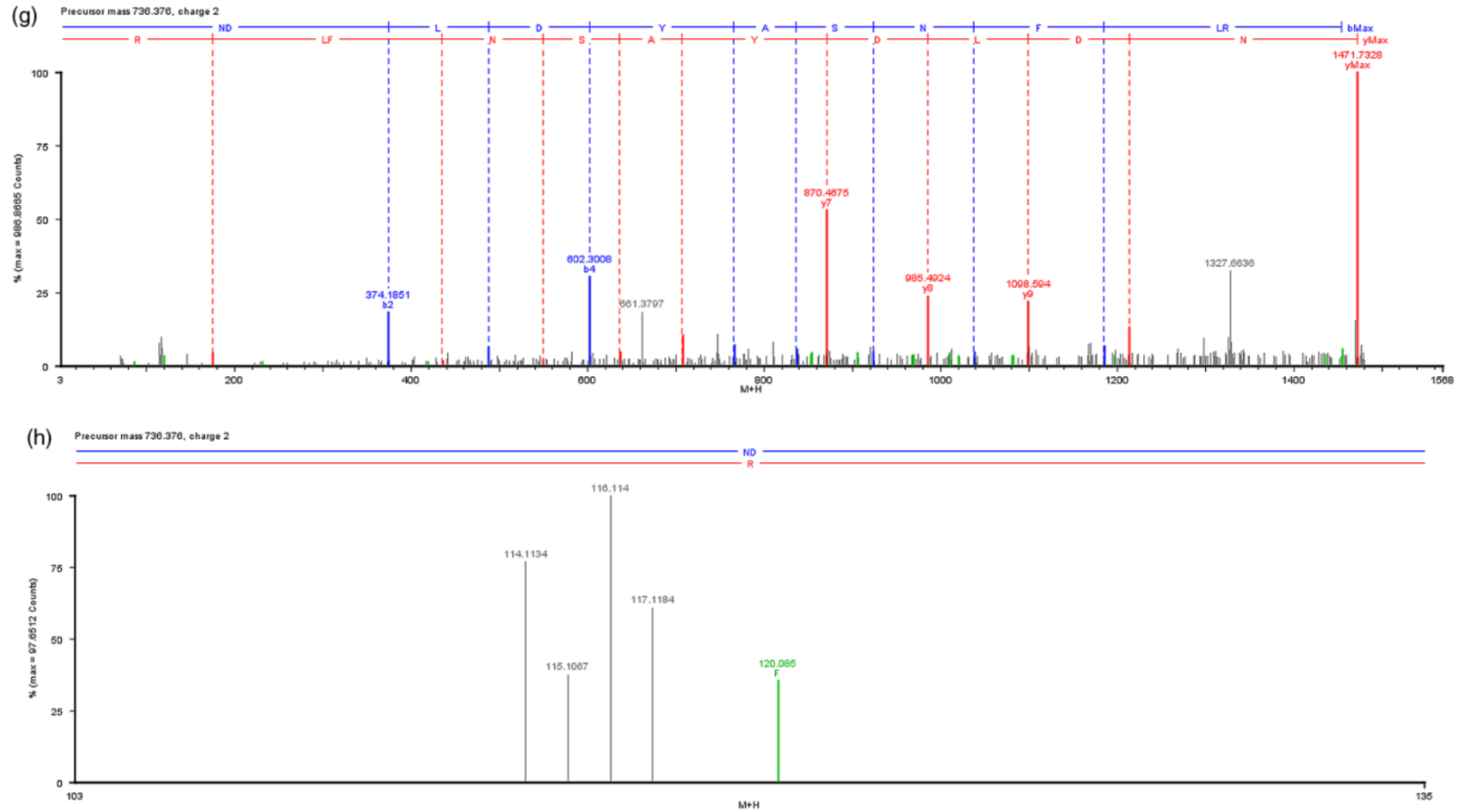


Figure 10.6 - (g) precursor m/z 736, (h) reporter ions for precursor 736.

10.3 Comparison of experimental approaches

A summary of the methodology for all three experimental systems and the results obtained from each can be seen in Table 10.2. There is a stark difference in the total amount of protein required for the three setups: up to 1 mg for iTRAQ, 14 µg for the gel-based method, and less than 1 µg for label-free. Although the injection amount for the LC-MS/MS analysis is comparable between all three techniques, this does not relate to the total amount of protein required for an adequate dataset. In the gel-based and iTRAQ approaches, the amount indicated is necessary to generate enough peptides over 30-60 fractions for MS analysis. With the employed label-free method, the amount loaded directly for LC-MS analysis is sufficient for a full qualitative and quantitative dataset. Sample requirement can be an important consideration when performing proteomic studies, as it can be a challenge to generate a suitable amount from biological systems. If less sample is required for a single experiment, additional analyses can be carried out, which will add confidence to the results obtained (Karp and Lilley, 2007). It has previously been shown that even three replicate MudPIT experiments may not provide full coverage of all the proteins within a sample (Durr *et al.*, 2004).

An ideal method for proteomic analysis would enable comprehensive and high-throughput studies, making experimental and instrumentation time an important factor when considering which approach to utilise. Both the gel-based and iTRAQ setups require up to 60 hours of MS data acquisition time, based upon our chosen number of bands cut from the gel or fractions from the strong cation exchange chromatography, and upon the gradient setup in the reversed-phase chromatography. The analytical time could be shortened by choosing fewer fractions, or reducing the reversed-phase gradient, but this may also reduce peptide recovery and/or separation. The label-free experiments require 6 hours of instrument time (2 hours per replicate). In addition to this, preparing samples for iTRAQ requires a number of days, including overnight steps. This issue can make the approach less suitable for a routine analysis setup when compared to the label-free method.

The average number of peptides identified per confident protein assignment for the gel-based and iTRAQ analyses is 5, compared to an average of 12 for the label-free method. The gel-based approach gives an average sequence coverage of 15%, higher

than the iTRAQ average of 11% which is slightly lower than previous work (Domon and Aebersold, 2006). The average sequence coverage for the label-free data is 45%. An increased number of peptides and higher sequence coverage can confer more confidence in identifications obtained.

	1D-SDS-PAGE	iTRAQ	Label-free
Protein loading	14 µg	100 µg per iTRAQ labeling vial; 800 µg total loading	0.5 µg for each of 3 technical replicates
Number of overnight steps	2	5	1
Samples to analyse by MS	30-40 fractions	30-60 fractions	1 per growth condition
Reverse-phase LC and MS acquisition	30-40 hours	30-60 hours	2 hours
Total analysis time	4 days	6 days	Less than 3 days
Total instrument time	30-40 hours	30-60 hours	6 hours per sample
Size of data file	300 MB x 40 (1.2 GB)	300 MB x 40 (1.2 GB)	6 GB x 3 (18GB)
Number of proteins confidently identified	235	178	421
Average number of peptides per protein (when including single-peptide identifications)	5 (3)	5 (3)	12 (12)
Average sequence coverage (when including single-peptide identifications)	15 % (10 %)	11 % (7 %)	45 % (45 %)

Table 10.2 - A comparison of the experimental requirements for each of the approaches, and the information obtained from the data generated.

10.4 Biological significance of results obtained

10.4.1 Placing results in the context of methane oxidation

Bacteria from the methanotroph family utilise a common pathway to process methane in order to use it as a carbon and energy source, an overview of which is given in Figure 10.7. As the *Methylocella* genus has been recently identified and is relatively uncharacterised, it is difficult to make predictions about potential biochemical changes which would be seen on a growth substrate other than methane. It could, however, be suggested that some down-regulation of the enzymes in the methane oxidation pathway would be seen. Our study identified the key enzymes methane monooxygenase (MMO) and methanol dehydrogenase, with quantitative data from both iTRAQ and the label-free approach indicating a significant down-regulation when *M. silvestris* was grown on acetate. MMO is a multimeric protein with subunits α , β and γ (Colby and Dalton, 1976). The α and β subunits show up-regulation in the methane growth samples, as does the accessory MMO Protein B. The γ subunit shows significant up-regulation on methane when analysed by iTRAQ; using our data filtering (more than one peptide, more than one replicate for label-free) this subunit is only seen in the methane growth condition for the gel and label-free analyses, as is the accessory MMO Protein C. There is also up-regulation of the alpha and beta subunits of methanol dehydrogenase in the methane-grown samples, which is the second enzyme in the methanotroph methane oxidation pathway.



Figure 10.7 - The pathway of methane oxidation in methanotrophic bacteria.

The quantification data relating to these enzyme identifications has been shown in Figure 10.8. DNA-directed RNA polymerase has also been included as a housekeeping protein, and as such should not display up- or down-regulation regardless of growth substrate. Such proteins can provide a check, on the biological level, for the significance of differential proteomic data.

10.4.2 Mapping protein identifications onto biological pathways

Having identified the proteins expressed by the *Methylocella silvestris* organism under the different growth conditions and obtained the relative expression levels for the regulated proteins it is important to probe the biochemical pathways employed. Such pathway mapping is becoming increasingly popular in the fields of genomics and proteomics. Information regarding gene expression or levels of proteins only provides part of the picture; in order to further understand the biochemistry within a system, this information must be put into the context of cell activity.

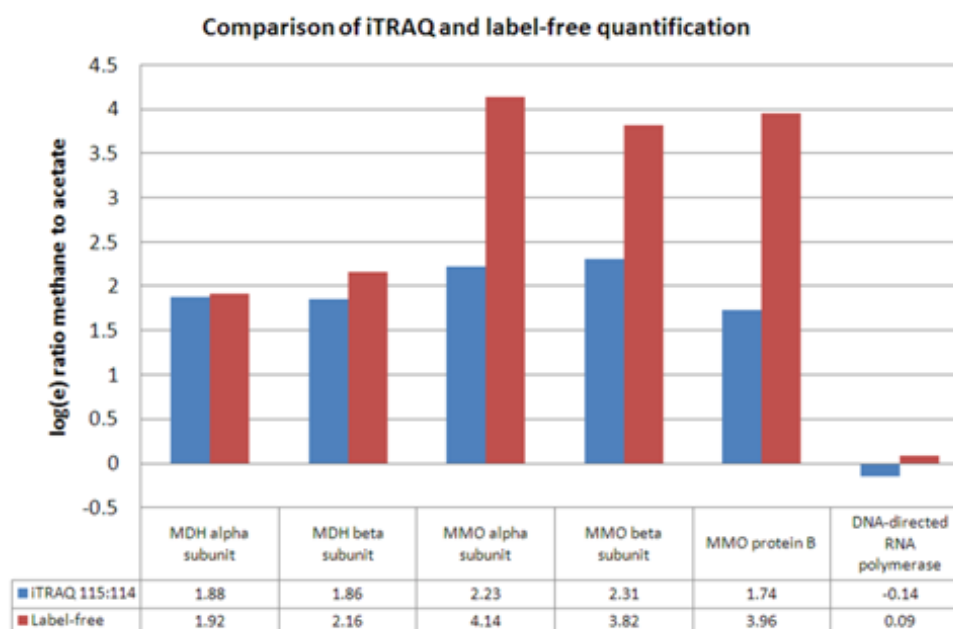


Figure 10.8 - *iTRAQ* and label-free quantification data for proteins identified as key metabolic enzymes within *M. silvestris*; a housekeeping protein has been included as indication of a biological marker.

In order to begin to classify the pathways involved in *M. silvestris* data was uploaded to KAAS. Although the system works best when a complete set of genes in a genome is known, it is not necessary. As *M. silvestris* has only recently been identified, sequencing of its genome is still in progress; the gene and protein databases are, therefore, incomplete and not publically available. Pathways specific

to this organism were therefore not able to be searched, but assignments to pathways could be made on the basis of homology. The output of the KAAS program indicated that the majority of proteins identified were involved in metabolism (e.g. TCA cycle) and genetic information processing, as would be expected for this type of organism. Figure 10.9 shows an example of the output obtained with proteins from each growth condition mapped onto the citric acid cycle; significant specific differences can be seen between the various substrates.

Methylocella assimilates carbon from the formaldehyde produced during methane oxidation via the serine cycle. The serine cycle requires the conversion of acetyl CoA into glyoxylate, catalysed by isocitrate lyase, but a lot of methanotrophs do not possess this enzyme. Interestingly, isocitrate lyase was only detected in the acetate growth condition. This might suggest that it was used during metabolism of acetate but not during metabolism of methane, and would need to be investigated further.

The conventional idea of acetate metabolism is that first acetate is converted into acetyl CoA, by an enzyme such as acetate CoA ligase, and that the metabolism of acetyl CoA requires the glyoxylate bypass, which consists of the two enzymes isocitrate lyase and malate synthase, to bypass the decarboxylation steps of the TCA cycle, and allow assimilation of biomass from acetyl CoA. Malate synthase was only detected in acetate, in keeping with its role in the glyoxylate bypass, and acetate CoA ligase was highly upregulated during growth on acetate as expected.

These data, showing differences in expression of key proteins between the two growth conditions, indicate that a proteomics-based approach has the potential to be expanded in the future to probe these metabolic pathways in more detail. Combined with other methods (e.g. enzyme assays), a great deal of valuable information could be obtained providing crucial insight into the biochemistry of *Methylocella*.

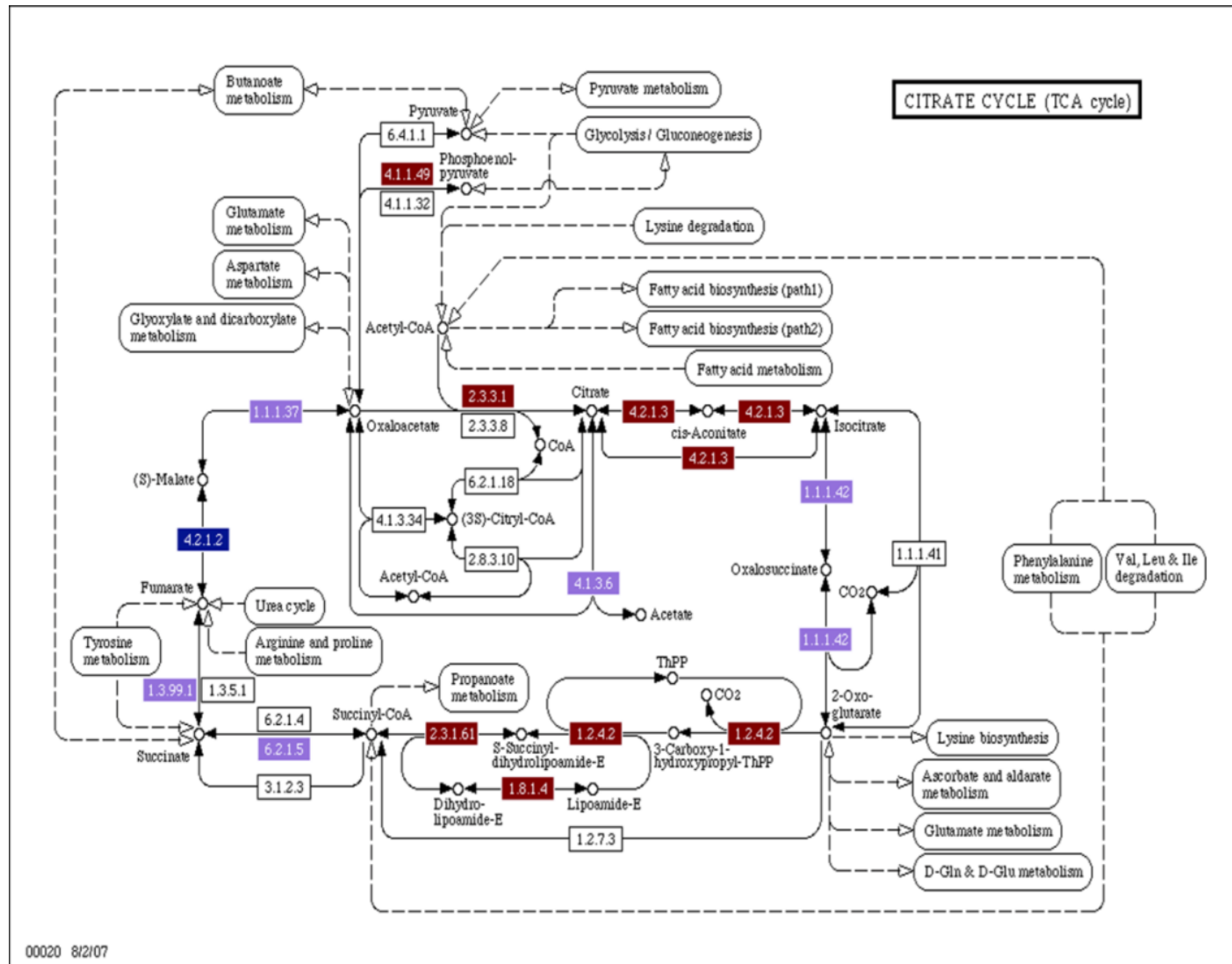


Figure 10.9 - Extracted pathway analysis of the citric acid cycle from KEGG database; red - unique when acetate used as substrate, blue - unique when methane used as substrate, purple - present under both growth conditions.

10.5 Conclusions

For any integrated proteomics experiment, a number of important issues need to be considered. These include the need for qualitative (profiling) and quantitative information. Confidence in identification and quantification, reproducibility, sample size, instrument time, sample preparation, cost, and sequence coverage are all important factors that need to be taken into account. The ability to place any changes observed into the context of the biological pathways involved remains a crucial aspect of the research. This study has evaluated the potential applicability of a number of common approaches to profiling and differential proteomics. The experiments have been restricted to a proteomics study of cytosolic proteins, and comparable technology platforms were employed. Good agreement was obtained between the commonly utilised iTRAQ labelled experiment, a gel based study and that based on a label-free LC-MS approach. At the profiling level, when considering all identifications, including those based on single peptides, the number of identified proteins was comparable for all three methods. When requiring more than one peptide for identification, the label-free approach gave superior information particularly when coverage was taken into account. Both the iTRAQ experiment and the label-free approach provided relative quantification datasets, and the agreement between the approaches was better than previously observed in comparisons between different quantitative methods (Usaite *et al.*, 2008b). This is most likely due to the use of comparable instrumentation, as each method employed high-performance liquid chromatography coupled to a Q-TOF tandem MS acquisition. The label-free experiment does, however, have advantages in terms of sample requirement, sample preparation and instrumental time requirements. A preliminary screen of the protein regulation results for biological significance shows agreement with previous analysis of the regulation of methane monooxygenase in *Methylocella*. This, together with the significant number of identifications provided by all three approaches, and the excellent agreement of two quantitative datasets, indicates the potential for further proteomic studies on this methanotroph.

References

- Breci, L., Hatstrup, E., Keeler, M., Letarte, J., Johnson, R., Haynes, P. A.** (2005). Comprehensive proteomics in yeast using chromatographic fractionation, gas phase fractionation, protein gel electrophoresis, and isoelectric focusing. *Proteomics* **5**: 2018-2028.
- Chambery, A., Vissers, J. P. C., Langridge, J. I., Lonardo, E., Minchiotti, G., Ruvo, M., Parente, A.** (2009). Qualitative and Quantitative Proteomic Profiling of Cripto^{-/-} Embryonic Stem Cells by Means of Accurate Mass LC-MS Analysis. *J Proteome Res.* **8**: 1047-1058.
- Colby, J., Dalton, H.** (1976). Some properties of a soluble methane mono-oxygenase from *Methylococcus capsulatus* strain Bath. *Biochemical Journal* **157**: 495-497.
- Domon, B., Aebersold, R.** (2006). Mass Spectrometry and Protein Analysis. *Science* **312**: 212-217.
- Durr, E., Yu, J., Krasinska, K. M., Carver, L. A., Yates, J. R., Testa, J. E., Oh, P., Schnitzer, J. E.** (2004). Direct proteomic mapping of the lung microvascular endothelial cell surface in vivo and in cell culture. *Nat Biotech* **22**: 985-992.
- Karp, N. A., Lilley, K. S.** (2007). Design and analysis issues in quantitative proteomics studies. *Proteomics* **7 Suppl 1**: 42-50.
- Ross, P. L., Huang, Y. N., Marchese, J. N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhasz, P., Martin, S., Bartlet-Jones, M., He, F., Jacobson, A., Pappin, D. J.** (2004). Multiplexed Protein Quantitation in *Saccharomyces cerevisiae* Using Amine-reactive Isobaric Tagging Reagents. *Mol Cell Proteomics* **3**: 1154-1169.
- Silva, J. C., Denny, R., Dorschel, C. A., Gorenstein, M., Kass, I. J., Li, G. Z., McKenna, T., Nold, M. J., Richardson, K., Young, P., Geromanos, S.** (2005). Quantitative Proteomic Analysis by Accurate Mass Retention Time Pairs. *Anal. Chem.* **77**: 2187-2200.
- Silva, J. C., Gorenstein, M. V., Li, G.-Z., Vissers, J. P. C., Geromanos, S. J.** (2006). Absolute Quantification of Proteins by LCMS^E: A Virtue of Parallel MS Acquisition. *Mol. Cell. Proteomics* **5**: 144-156.
- Theisen, A. R., Ali, M. H., Radajewski, S., Dumont, M. G., Dunfield, P. F., McDonald, I. R., Dedysh, S. N., Miguez, C. B., Murrell, J. C.** (2005). Regulation of methane oxidation in the facultative methanotroph *Methylocella silvestris* BL2. *Mol. Microbiol.* **58**: 682-692.
- Usaite, R., Wohlschlegel, J., Venable, J. D., Park, S. K., Nielsen, J., Olsson, L., Yates Iii, J. R.** (2008a). Characterization of Global Yeast Quantitative Proteome Data Generated from the Wild-Type and Glucose Repression *Saccharomyces cerevisiae* Strains: The Comparison of Two Quantitative Methods. *J. Proteome Res.* **7**: 266-275.

- Usaité, R., Wohlschlegel, J., Venable, J. D., Park, S. K., Nielsen, J., Olsson, L., Yates III, J. R.** (2008b). Characterization of Global Yeast Quantitative Proteome Data Generated from the Wild-Type and Glucose Repression *Saccharomyces cerevisiae* Strains: The Comparison of Two Quantitative Methods. *Journal of Proteome Research* **7**: 266-275.
- Veenstra, T. D., Conrads, T. P., Issaq, H. J.** (2004). What to do with one-hit wonders? *Electrophoresis* **25**: 1278-1279.
- Vissers, J. P. C., Langridge, J. I., Aerts, J. M. F. G.** (2007). Analysis and Quantification of Diagnostic Serum Markers and Protein Signatures for Gaucher Disease. *Mol Cell Proteomics* **6**: 755-766.

Concluding remarks

The work presented here describes the use mass spectrometry for the study of a number of biological systems.

New ambient ionisation techniques, coupled to different mass analysers, were employed for the rapid screening of pharmaceutical formulations. Active ingredients were identified, potential fragmentation pathways were elucidated and drug metabolites were also successfully identified from biological samples. Since the work was conducted, there has been huge growth in the field of ambient ionisation, with a number of novel and complementary techniques emerging for various applications. There remains much to determine regarding aspects such as fundamental mechanisms of ionisation, detection limits for compounds of interest, and the utilisation of these methods' non-invasive and high-throughput capabilities for practical applications. Studies are continuing in our research group on the investigation of pharmaceutical formulations using a variety of methods, and also on natural products used as alternative therapies and supplements.

Inorganic mass spectrometry was employed to probe the metal centres of the enzyme, particulate methane monooxygenase, a methane-oxidising complex found in certain bacteria. The data obtained was largely inconclusive, but this protein remains of significant interest. The mechanism by which it oxidises methane is still to be fully understood, but if this is to be performed by mass spectrometry, then considerable progress will first have to be made in improving the compatibility of sample preparation methods with the instrumentation used.

It was shown that protein cross-sections, obtained using a commercial ion mobility mass spectrometry instrument, can be used to probe the conformation of hemoglobin in the gas phase. This initial work has been followed by investigations of a number of other clinical blood samples, and also the study of other metal containing proteins, such as copper-bound prion and zinc finger complexes. There is still much to optimise in terms of calibrating travelling wave-based ion mobility for different types of sample, and also the correlation of data obtained with complementary techniques to gain as full a picture as possible regarding protein conformation.

The comparison of three proteomics techniques to study a novel organism, *Methylocella silvestris* showed promising results, both in terms of the technical application of methods involved and the biology of the organism. Since the publication of this work, a fully-annotated genome has been released. This will allow for the data obtained to be re-analysed, in order to create a clearer picture of differences in the organism between growth on a one-carbon and on a multi-carbon compound. Work is now being carried out to investigate the technical merits of a multi-dimensional separation of the proteome prior to mass spectrometric analysis; alongside this, the growth of *M. silvestris* on a number of multi-carbon compounds is being investigated.

Appendix A

Supplementary Proteomic Data

Supplementary Table 1

Protein identifications – Gel-based separation – Methane growth condition

Entry	Description	Number of peptides	Sequence coverage (%)	Molecular weight (Da)	pI
0008_MSIL	msil 20jun08 Contig14 revised geneMsil0008	3	8.2317	34715	6.1877
0061_MSIL	msil 20jun08 Contig14 revised geneMsil0061	3	13.6546	25530	6.1859
0082_MSIL	msil 20jun08 Contig14 revised geneMsil0082	2	3.9773	39510	5.7627
0096_MSIL	msil 20jun08 Contig14 revised geneMsil0096	6	28.5714	21825	5.5093
0097_MSIL	msil 20jun08 Contig14 revised geneMsil0097	2	8.6066	27751	6.0683
0098_MSIL	msil 20jun08 Contig14 revised geneMsil0098	2	8.7805	23289	4.9658
0216_MSIL	msil 20jun08 Contig14 revised geneMsil0216	9	11.8421	92754	6.0967
0268_MSIL	msil 20jun08 Contig14 revised geneMsil0268	4	6.8627	77238	5.161
0340_MSIL	msil 20jun08 Contig14 revised geneMsil0340	2	11.7021	20083	7.8691
0341_MSIL	msil 20jun08 Contig14 revised geneMsil0341	4	8.6444	54708	5.7219
0343_MSIL	msil 20jun08 Contig14 revised geneMsil0343	9	22.1532	51312	4.745
0367_MSIL	msil 20jun08 Contig14 revised geneMsil0367	2	3.6957	49285	5.0629
0426_MSIL	msil 20jun08 Contig14 revised geneMsil0426	8	20.4611	38813	5.7072
0427_MSIL	msil 20jun08 Contig14 revised geneMsil0427	4	8.9552	52033	5.2363
0471_MSIL	msil 20jun08 Contig14 revised geneMsil0471	29	48.32	68485	5.8453
0472_MSIL	msil 20jun08 Contig14 revised geneMsil0472	5	15.4386	31041	5.1363
0473_MSIL	msil 20jun08 Contig14 revised geneMsil0473	4	26.2376	22047	5.814
0475_MSIL	msil 20jun08 Contig14 revised geneMsil0475	2	4.6377	38805	5.5419
0515_MSIL	msil 20jun08 Contig14 revised geneMsil0515	7	19.4379	44925	4.5795
0577_MSIL	msil 20jun08 Contig14 revised geneMsil0577	2	5.7348	30494	11.35
0579_MSIL	msil 20jun08 Contig14 revised geneMsil0579	2	7.767	22125	10.897

0580_MSIL	msil 20jun08 Contig14 revised geneMsil0580	3	13.4694	26030	10.3667
0582_MSIL	msil 20jun08 Contig14 revised geneMsil0582	12	30.8081	43107	5.3617
0583_MSIL	msil 20jun08 Contig14 revised geneMsil0583	4	4.9204	76241	5.0898
0584_MSIL	msil 20jun08 Contig14 revised geneMsil0584	4	29.4872	17839	10.3623
0620_MSIL	msil 20jun08 Contig14 revised geneMsil0620	3	14.2857	28612	4.7858
0643_MSIL	msil 20jun08 Contig14 revised geneMsil0643	9	22.1477	47756	6.0291
0795_MSIL	msil 20jun08 Contig14 revised geneMsil0795	13	32.5411	57572	4.9437
0811_MSIL	msil 20jun08 Contig14 revised geneMsil0811	8	26.8199	27389	5.9431
0832_MSIL	msil 20jun08 Contig14 revised geneMsil0832	3	10.25	41752	6.078
0928_MSIL	msil 20jun08 Contig14 revised geneMsil0928	2	5.8981	38838	5.5254
0968_MSIL	msil 20jun08 Contig14 revised geneMsil0968	2	15	15026	7.7732
0971_MSIL	msil 20jun08 Contig14 revised geneMsil0971	2	8.1522	20591	4.5989
1012_MSIL	msil 20jun08 Contig14 revised geneMsil1012	2	8.5911	31519	5.4069
1043_MSIL	msil 20jun08 Contig14 revised geneMsil1043	2	8.5561	19103	4.5654
1125_MSIL	msil 20jun08 Contig14 revised geneMsil1125	4	8.5714	54389	5.9583
1140_MSIL	msil 20jun08 Contig14 revised geneMsil1140	2	8.2645	26506	4.8895
1193_MSIL	msil 20jun08 Contig14 revised geneMsil1193	2	2.5487	70184	5.9489
1226_MSIL	msil 20jun08 Contig14 revised geneMsil1226	5	7.4419	70867	5.6898
1237_MSIL	msil 20jun08 Contig14 revised geneMsil1237	4	15.5116	33325	6.8214
1262_MSIL	msil 20jun08 Contig14 revised geneMsil1262	11	23.5741	59751	5.7755
1263_MSIL	msil 20jun08 Contig14 revised geneMsil1263	18	40.1535	44905	5.7274
1265_MSIL	msil 20jun08 Contig14 revised geneMsil1265	6	28.8235	19537	9.3867
1267_MSIL	msil 20jun08 Contig14 revised geneMsil1267	3	10.5714	38492	4.8748
1325_MSIL	msil 20jun08 Contig14 revised geneMsil1325	2	2.6352	79052	6.0073
1360_MSIL	msil 20jun08 Contig14 revised geneMsil1360	12	10.7236	126072	5.9681
1382_MSIL	msil 20jun08 Contig14 revised geneMsil1382	2	3.6058	46188	5.373

1395_MSIL	msil 20jun08 Contig14 revised geneMsil1395	2	2.9371	75621	5.2117
1574_MSIL	msil 20jun08 Contig14 revised geneMsil1574	2	11.8519	28560	5.7598
1575_MSIL	msil 20jun08 Contig14 revised geneMsil1575	4	14.0794	29145	6.791
1607_MSIL	msil 20jun08 Contig14 revised geneMsil1607	4	17.2691	27020	5.248
1681_MSIL	msil 20jun08 Contig14 revised geneMsil1681	14	27.972	47690	6.3444
1691_MSIL	msil 20jun08 Contig14 revised geneMsil1691	3	13.4783	24353	4.483
1706_MSIL	msil 20jun08 Contig14 revised geneMsil1706	2	4.5356	49832	5.1275
1712_MSIL	msil 20jun08 Contig14 revised geneMsil1712	2	3.5842	59899	6.6418
1713_MSIL	msil 20jun08 Contig14 revised geneMsil1713	9	36.4217	33824	5.2581
1714_MSIL	msil 20jun08 Contig14 revised geneMsil1714	9	27.2727	42582	6.893
1716_MSIL	msil 20jun08 Contig14 revised geneMsil1716	6	19.2982	42881	5.2007
1718_MSIL	msil 20jun08 Contig14 revised geneMsil1718	3	2.8292	102443	6.1371
1758_MSIL	msil 20jun08 Contig14 revised geneMsil1758	3	2.5758	141391	5.892
1808_MSIL	msil 20jun08 Contig14 revised geneMsil1808	13	38.4342	30823	5.7768
1819_MSIL	msil 20jun08 Contig14 revised geneMsil1819	3	3.0336	101793	5.4351
1821_MSIL	msil 20jun08 Contig14 revised geneMsil1821	2	5.5556	35710	5.2853
1841_MSIL	msil 20jun08 Contig14 revised geneMsil1841	2	7.8947	28110	9.3347
1860_MSIL	msil 20jun08 Contig14 revised geneMsil1860	2	5.1383	52530	6.5993
1872_MSIL	msil 20jun08 Contig14 revised geneMsil1872	2	13.2075	17446	10.975
1907_MSIL	msil 20jun08 Contig14 revised geneMsil1907	2	2.8935	95453	5.6468
2004_MSIL	msil 20jun08 Contig14 revised geneMsil2004	8	31.3953	28160	5.4965
2007_MSIL	msil 20jun08 Contig14 revised geneMsil2007	23	59.2593	43429	7.1534
2074_MSIL	msil 20jun08 Contig14 revised geneMsil2074	2	4.5977	45551	5.0458
2091_MSIL	msil 20jun08 Contig14 revised geneMsil2091	6	15.6627	35103	5.5523
2110_MSIL	msil 20jun08 Contig14 revised geneMsil2110	8	17.9724	45996	6.7786
2111_MSIL	msil 20jun08 Contig14 revised geneMsil2111	3	10.8824	33339	5.5651

2114_MSIL	msil 20jun08 Contig14 revised geneMsil2114	2	5.2	27896	5.5937
2282_MSIL	msil 20jun08 Contig14 revised geneMsil2282	7	21.194	35502	6.9589
2301_MSIL	msil 20jun08 Contig14 revised geneMsil2301	6	17.2962	55865	5.3183
2385_MSIL	msil 20jun08 Contig14 revised geneMsil2385	7	17.2881	30948	4.9871
2387_MSIL	msil 20jun08 Contig14 revised geneMsil2387	4	15.528	33939	5.3212
2390_MSIL	msil 20jun08 Contig14 revised geneMsil2390	5	31.4286	18558	5.8748
2400_MSIL	msil 20jun08 Contig14 revised geneMsil2400	4	12.1951	39186	4.9005
2402_MSIL	msil 20jun08 Contig14 revised geneMsil2402	2	7.947	31864	6.1088
2447_MSIL	msil 20jun08 Contig14 revised geneMsil2447	4	15.7895	36002	6.5391
2460_MSIL	msil 20jun08 Contig14 revised geneMsil2460	2	2.3256	76831	6.9587
2501_MSIL	msil 20jun08 Contig14 revised geneMsil2501	7	18.0685	33442	5.088
2503_MSIL	msil 20jun08 Contig14 revised geneMsil2503	2	8.8435	30052	5.3925
2523_MSIL	msil 20jun08 Contig14 revised geneMsil2523	11	30.1242	34028	5.8129
2601_MSIL	msil 20jun08 Contig14 revised geneMsil2601	2	7.2829	38162	5.0845
2753_MSIL	msil 20jun08 Contig14 revised geneMsil2753	2	8.0508	23938	5.599
2955_MSIL	msil 20jun08 Contig14 revised geneMsil2955	9	18.9274	67769	4.823
2970_MSIL	msil 20jun08 Contig14 revised geneMsil2970	2	10	23277	4.7113
2996_MSIL	msil 20jun08 Contig14 revised geneMsil2996	5	14.6154	40184	6.267
2997_MSIL	msil 20jun08 Contig14 revised geneMsil2997	6	24.8963	25472	8.4214
3002_MSIL	msil 20jun08 Contig14 revised geneMsil3002	3	8.6567	35366	5.6917
3011_MSIL	msil 20jun08 Contig14 revised geneMsil3011	5	12.9794	36619	5.714
3157_MSIL	msil 20jun08 Contig14 revised geneMsil3157	4	7.3801	59779	5.8885
3226_MSIL	msil 20jun08 Contig14 revised geneMsil3226	2	7.7441	32668	6.0456
3309_MSIL	msil 20jun08 Contig14 revised geneMsil3309	3	13.806	28857	4.6071
3510_MSIL	msil 20jun08 Contig14 revised geneMsil3510	2	6.4865	40188	4.9473
3530_MSIL	msil 20jun08 Contig14 revised geneMsil3530	6	16.2162	45231	6.7689

3552_MSIL	msil 20jun08 Contig14 revised geneMsil3552	5	23.1183	19778	9.7436
3626_MSIL	msil 20jun08 Contig14 revised geneMsil3626	4	25.1613	17620	5.0894
3632_MSIL	msil 20jun08 Contig14 revised geneMsil3632	3	5.802	31524	4.7128
3774_MSIL	msil 20jun08 Contig14 revised geneMsil3774	5	5.7811	86684	5.2244
3818_MSIL	msil 20jun08 Contig14 revised geneMsil3818	2	5.6537	31221	5.0032
3865_MSIL	msil 20jun08 Contig14 revised geneMsil3865	3	12.987	23929	9.9968
3866_MSIL	msil 20jun08 Contig14 revised geneMsil3866	2	12.2807	17717	9.906
3868_MSIL	msil 20jun08 Contig14 revised geneMsil3868	6	4.2878	153612	5.0548
3874_MSIL	msil 20jun08 Contig14 revised geneMsil3874	2	5.0847	51686	6.0818
3875_MSIL	msil 20jun08 Contig14 revised geneMsil3875	20	14.1124	170158	5.7504

Supplementary Table 2

Protein identifications – Gel-based separation – Acetate growth condition

Entry	Description	Number of peptides	Sequence coverage (%)	Molecular weight (Da)	pI
0058_MSIL	msil 20jun08 Contig14 revised geneMsil0058	6	27.9245	27691	9.0434
0061_MSIL	msil 20jun08 Contig14 revised geneMsil0061	3	13.6546	25530	6.1859
0074_MSIL	msil 20jun08 Contig14 revised geneMsil0074	2	10.5882	27817	6.3344
0081_MSIL	msil 20jun08 Contig14 revised geneMsil0081	5	18.0685	35081	5.3448
0082_MSIL	msil 20jun08 Contig14 revised geneMsil0082	7	19.3182	39510	5.7627
0096_MSIL	msil 20jun08 Contig14 revised geneMsil0096	4	12.7551	21825	5.5093
0097_MSIL	msil 20jun08 Contig14 revised geneMsil0097	5	20.9016	27751	6.0683
0098_MSIL	msil 20jun08 Contig14 revised geneMsil0098	5	19.0244	23289	4.9658
0105_MSIL	msil 20jun08 Contig14 revised geneMsil0105	2	8.7805	23649	10.3174
0106_MSIL	msil 20jun08 Contig14 revised geneMsil0106	2	4.1016	54236	5.7662
0152_MSIL	msil 20jun08 Contig14 revised geneMsil0152	2	5.9701	35747	10.205
0178_MSIL	msil 20jun08 Contig14 revised geneMsil0178	6	26.284	34870	6.0123
0194_MSIL	msil 20jun08 Contig14 revised geneMsil0194	7	6.7568	96924	5.1328
0195_MSIL	msil 20jun08 Contig14 revised geneMsil0195	3	22.9508	19801	5.4911
0202_MSIL	msil 20jun08 Contig14 revised geneMsil0202	3	9.9432	37207	6.2
0209_MSIL	msil 20jun08 Contig14 revised geneMsil0209	5	9.3098	70016	5.3441
0216_MSIL	msil 20jun08 Contig14 revised geneMsil0216	4	6.1005	92754	6.0967
0255_MSIL	msil 20jun08 Contig14 revised geneMsil0255	2	9.9567	26309	5.1048
0268_MSIL	msil 20jun08 Contig14 revised geneMsil0268	15	19.888	77238	5.161
0286_MSIL	msil 20jun08 Contig14 revised geneMsil0286	3	10.9489	29864	4.5688
0318_MSIL	msil 20jun08 Contig14 revised geneMsil0318	2	5.7229	37242	5.4631

0336_MSIL	msil 20jun08 Contig14 revised geneMsil0336	3	3.4247	97923	5.6922
0340_MSIL	msil 20jun08 Contig14 revised geneMsil0340	3	18.0851	20083	7.8691
0341_MSIL	msil 20jun08 Contig14 revised geneMsil0341	12	26.7191	54708	5.7219
0342_MSIL	msil 20jun08 Contig14 revised geneMsil0342	5	16.8385	31610	9.8547
0343_MSIL	msil 20jun08 Contig14 revised geneMsil0343	8	19.0476	51312	4.745
0351_MSIL	msil 20jun08 Contig14 revised geneMsil0351	2	10.7884	25862	5.4586
0367_MSIL	msil 20jun08 Contig14 revised geneMsil0367	2	6.087	49285	5.0629
0380_MSIL	msil 20jun08 Contig14 revised geneMsil0380	4	16.2264	28019	6.1848
0412_MSIL	msil 20jun08 Contig14 revised geneMsil0412	2	6.4257	26499	7.2621
0426_MSIL	msil 20jun08 Contig14 revised geneMsil0426	2	6.9164	38813	5.7072
0427_MSIL	msil 20jun08 Contig14 revised geneMsil0427	9	24.0938	52033	5.2363
0471_MSIL	msil 20jun08 Contig14 revised geneMsil0471	26	42.08	68485	5.8453
0472_MSIL	msil 20jun08 Contig14 revised geneMsil0472	9	32.6316	31041	5.1363
0473_MSIL	msil 20jun08 Contig14 revised geneMsil0473	4	29.2079	22047	5.814
0499_MSIL	msil 20jun08 Contig14 revised geneMsil0499	3	10.2041	35928	5.3007
0515_MSIL	msil 20jun08 Contig14 revised geneMsil0515	5	12.8806	44925	4.5795
0519_MSIL	msil 20jun08 Contig14 revised geneMsil0519	2	5.814	38123	6.9263
0520_MSIL	msil 20jun08 Contig14 revised geneMsil0520	2	5.6522	48749	4.5293
0559_MSIL	msil 20jun08 Contig14 revised geneMsil0559	10	48.7047	20525	7.5602
0565_MSIL	msil 20jun08 Contig14 revised geneMsil0565	2	11.2994	18780	10.478
0579_MSIL	msil 20jun08 Contig14 revised geneMsil0579	2	7.767	22125	10.897
0582_MSIL	msil 20jun08 Contig14 revised geneMsil0582	9	21.2121	43107	5.3617
0583_MSIL	msil 20jun08 Contig14 revised geneMsil0583	10	15.0507	76241	5.0898
0619_MSIL	msil 20jun08 Contig14 revised geneMsil0619	3	16.1111	19836	5.2982
0620_MSIL	msil 20jun08 Contig14 revised geneMsil0620	5	22.6481	28612	4.7858
0629_MSIL	msil 20jun08 Contig14 revised geneMsil0629	2	4.8309	44413	9.3138

0643_MSIL	msil 20jun08 Contig14 revised geneMsil0643	4	9.396	47756	6.0291
0672_MSIL	msil 20jun08 Contig14 revised geneMsil0672	3	2.3487	118521	4.981
0673_MSIL	msil 20jun08 Contig14 revised geneMsil0673	2	11.8421	16433	4.869
0730_MSIL	msil 20jun08 Contig14 revised geneMsil0730	3	8.0769	27821	6.2047
0764_MSIL	msil 20jun08 Contig14 revised geneMsil0764	2	9.375	31170	5.9786
0786_MSIL	msil 20jun08 Contig14 revised geneMsil0786	4	15.9159	35835	5.38
0795_MSIL	msil 20jun08 Contig14 revised geneMsil0795	24	44.0585	57572	4.9437
0811_MSIL	msil 20jun08 Contig14 revised geneMsil0811	8	24.5211	27389	5.9431
0822_MSIL	msil 20jun08 Contig14 revised geneMsil0822	5	15.1515	36012	5.0724
0823_MSIL	msil 20jun08 Contig14 revised geneMsil0823	3	24.6377	14691	4.8723
0848_MSIL	msil 20jun08 Contig14 revised geneMsil0848	5	16.443	31524	6.3311
0895_MSIL	msil 20jun08 Contig14 revised geneMsil0895	2	8.9796	25252	8.4752
0931_MSIL	msil 20jun08 Contig14 revised geneMsil0931	12	22.8632	51377	5.7431
0971_MSIL	msil 20jun08 Contig14 revised geneMsil0971	3	12.5	20591	4.5989
1012_MSIL	msil 20jun08 Contig14 revised geneMsil1012	5	18.9003	31519	5.4069
1025_MSIL	msil 20jun08 Contig14 revised geneMsil1025	8	38.5475	20004	5.2809
1043_MSIL	msil 20jun08 Contig14 revised geneMsil1043	2	8.5561	19103	4.5654
1068_MSIL	msil 20jun08 Contig14 revised geneMsil1068	4	12.8472	30007	6.239
1097_MSIL	msil 20jun08 Contig14 revised geneMsil1097	2	9.8361	19610	4.7765
1140_MSIL	msil 20jun08 Contig14 revised geneMsil1140	3	12.3967	26506	4.8895
1160_MSIL	msil 20jun08 Contig14 revised geneMsil1160	2	8.0986	31741	5.3829
1165_MSIL	msil 20jun08 Contig14 revised geneMsil1165	4	11.9403	35440	4.5306
1184_MSIL	msil 20jun08 Contig14 revised geneMsil1184	3	3.6735	77073	4.8657
1191_MSIL	msil 20jun08 Contig14 revised geneMsil1191	6	24.4318	37950	5.2592
1193_MSIL	msil 20jun08 Contig14 revised geneMsil1193	3	4.4978	70184	5.9489
1194_MSIL	msil 20jun08 Contig14 revised geneMsil1194	3	10.5263	39071	5.571

1214_MSIL	msil 20jun08 Contig14 revised geneMsil1214	4	13.2296	54892	5.2304
1215_MSIL	msil 20jun08 Contig14 revised geneMsil1215	5	12.1076	47959	5.3205
1218_MSIL	msil 20jun08 Contig14 revised geneMsil1218	3	12.7962	22113	5.4562
1226_MSIL	msil 20jun08 Contig14 revised geneMsil1226	17	25.5814	70867	5.6898
1237_MSIL	msil 20jun08 Contig14 revised geneMsil1237	2	9.901	33325	6.8214
1325_MSIL	msil 20jun08 Contig14 revised geneMsil1325	10	13.4535	79052	6.0073
1328_MSIL	msil 20jun08 Contig14 revised geneMsil1328	2	4.4355	52701	6.206
1375_MSIL	msil 20jun08 Contig14 revised geneMsil1375	10	23.694	58198	6.0253
1382_MSIL	msil 20jun08 Contig14 revised geneMsil1382	6	11.7788	46188	5.373
1388_MSIL	msil 20jun08 Contig14 revised geneMsil1388	8	32.5359	22288	5.8751
1395_MSIL	msil 20jun08 Contig14 revised geneMsil1395	10	10.9091	75621	5.2117
1519_MSIL	msil 20jun08 Contig14 revised geneMsil1519	5	15.3846	51487	5.7548
1575_MSIL	msil 20jun08 Contig14 revised geneMsil1575	7	25.2708	29145	6.791
1587_MSIL	msil 20jun08 Contig14 revised geneMsil1587	12	23.7937	65094	5.8834
1603_MSIL	msil 20jun08 Contig14 revised geneMsil1603	5	25.7642	24460	8.2592
1624_MSIL	msil 20jun08 Contig14 revised geneMsil1624	2	5.0378	42515	4.3233
1676_MSIL	msil 20jun08 Contig14 revised geneMsil1676	3	13.5678	22228	5.6911
1677_MSIL	msil 20jun08 Contig14 revised geneMsil1677	3	14.3541	22918	8.5001
1681_MSIL	msil 20jun08 Contig14 revised geneMsil1681	11	26.5734	47690	6.3444
1688_MSIL	msil 20jun08 Contig14 revised geneMsil1688	2	2.5499	100109	5.3073
1691_MSIL	msil 20jun08 Contig14 revised geneMsil1691	3	14.7826	24353	4.483
1693_MSIL	msil 20jun08 Contig14 revised geneMsil1693	4	12.2581	31980	5.3806
1694_MSIL	msil 20jun08 Contig14 revised geneMsil1694	2	4.5845	37679	5.5659
1705_MSIL	msil 20jun08 Contig14 revised geneMsil1705	4	14.7059	37163	5.33
1706_MSIL	msil 20jun08 Contig14 revised geneMsil1706	2	4.7516	49832	5.1275
1713_MSIL	msil 20jun08 Contig14 revised geneMsil1713	3	13.738	33824	5.2581

1714_MSIL	msil 20jun08 Contig14 revised geneMsil1714	8	26.5152	42582	6.893
1716_MSIL	msil 20jun08 Contig14 revised geneMsil1716	4	9.7744	42881	5.2007
1718_MSIL	msil 20jun08 Contig14 revised geneMsil1718	2	1.741	102443	6.1371
1759_MSIL	msil 20jun08 Contig14 revised geneMsil1759	2	5.3846	41902	7.7274
1760_MSIL	msil 20jun08 Contig14 revised geneMsil1760	4	10.3578	55688	5.8566
1808_MSIL	msil 20jun08 Contig14 revised geneMsil1808	11	31.6726	30823	5.7768
1851_MSIL	msil 20jun08 Contig14 revised geneMsil1851	2	8.2988	24865	4.1404
1907_MSIL	msil 20jun08 Contig14 revised geneMsil1907	12	14.9306	95453	5.6468
1910_MSIL	msil 20jun08 Contig14 revised geneMsil1910	6	9.8462	71093	4.9285
2091_MSIL	msil 20jun08 Contig14 revised geneMsil2091	7	29.5181	35103	5.5523
2110_MSIL	msil 20jun08 Contig14 revised geneMsil2110	7	21.1982	45996	6.7786
2111_MSIL	msil 20jun08 Contig14 revised geneMsil2111	5	18.8235	33339	5.5651
2114_MSIL	msil 20jun08 Contig14 revised geneMsil2114	6	24	27896	5.5937
2221_MSIL	msil 20jun08 Contig14 revised geneMsil2221	5	10.5751	55691	5.0689
2224_MSIL	msil 20jun08 Contig14 revised geneMsil2224	3	10.4762	34241	9.8112
2245_MSIL	msil 20jun08 Contig14 revised geneMsil2245	4	21.3198	21627	9.588
2282_MSIL	msil 20jun08 Contig14 revised geneMsil2282	9	24.4776	35502	6.9589
2283_MSIL	msil 20jun08 Contig14 revised geneMsil2283	4	10.5651	43185	5.7662
2342_MSIL	msil 20jun08 Contig14 revised geneMsil2342	5	11.0672	55624	5.5428
2344_MSIL	msil 20jun08 Contig14 revised geneMsil2344	2	10.4651	28660	4.5842
2360_MSIL	msil 20jun08 Contig14 revised geneMsil2360	4	21.8579	19910	4.6688
2361_MSIL	msil 20jun08 Contig14 revised geneMsil2361	3	8.9286	56514	5.4291
2362_MSIL	msil 20jun08 Contig14 revised geneMsil2362	3	18.0124	17271	6.3389
2385_MSIL	msil 20jun08 Contig14 revised geneMsil2385	3	9.4915	30948	4.9871
2387_MSIL	msil 20jun08 Contig14 revised geneMsil2387	2	7.1429	33939	5.3212
2390_MSIL	msil 20jun08 Contig14 revised geneMsil2390	3	25.7143	18558	5.8748

2400_MSIL	msil 20jun08 Contig14 revised geneMsil2400	7	17.0732	39186	4.9005
2403_MSIL	msil 20jun08 Contig14 revised geneMsil2403	3	12.6866	27342	6.1179
2447_MSIL	msil 20jun08 Contig14 revised geneMsil2447	6	22.2222	36002	6.5391
2485_MSIL	msil 20jun08 Contig14 revised geneMsil2485	2	6.0423	36190	8.7427
2501_MSIL	msil 20jun08 Contig14 revised geneMsil2501	13	44.8598	33442	5.088
2503_MSIL	msil 20jun08 Contig14 revised geneMsil2503	11	37.7551	30052	5.3925
2504_MSIL	msil 20jun08 Contig14 revised geneMsil2504	12	8.5462	113040	6.3444
2505_MSIL	msil 20jun08 Contig14 revised geneMsil2505	4	8.8785	44809	6.1965
2506_MSIL	msil 20jun08 Contig14 revised geneMsil2506	4	8.0851	49375	5.8597
2523_MSIL	msil 20jun08 Contig14 revised geneMsil2523	12	40.0621	34028	5.8129
2532_MSIL	msil 20jun08 Contig14 revised geneMsil2532	3	7.109	47245	5.8951
2576_MSIL	msil 20jun08 Contig14 revised geneMsil2576	3	4.8387	69037	5.3423
2592_MSIL	msil 20jun08 Contig14 revised geneMsil2592	3	14.2857	18866	4.6838
2598_MSIL	msil 20jun08 Contig14 revised geneMsil2598	2	6.7797	31211	5.5496
2608_MSIL	msil 20jun08 Contig14 revised geneMsil2608	4	12.6685	39929	5.7501
2641_MSIL	msil 20jun08 Contig14 revised geneMsil2641	2	8.5837	24484	9.2
2647_MSIL	msil 20jun08 Contig14 revised geneMsil2647	6	22.2727	24375	5.6852
2655_MSIL	msil 20jun08 Contig14 revised geneMsil2655	13	18.1446	79319	4.9565
2733_MSIL	msil 20jun08 Contig14 revised geneMsil2733	2	3.2134	87555	5.8548
2810_MSIL	msil 20jun08 Contig14 revised geneMsil2810	7	15.8317	53968	5.5942
2816_MSIL	msil 20jun08 Contig14 revised geneMsil2816	2	4.6794	61891	6.2639
2912_MSIL	msil 20jun08 Contig14 revised geneMsil2912	3	9.0909	50614	4.5209
2913_MSIL	msil 20jun08 Contig14 revised geneMsil2913	4	24.6512	24061	5.6364
2914_MSIL	msil 20jun08 Contig14 revised geneMsil2914	2	4.7506	46241	5.4924
2924_MSIL	msil 20jun08 Contig14 revised geneMsil2924	2	2.7656	73697	5.5728
2955_MSIL	msil 20jun08 Contig14 revised geneMsil2955	26	41.3249	67769	4.823

2961_MSIL	msil 20jun08 Contig14 revised geneMsil2961	4	9.75	43501	5.7889
2969_MSIL	msil 20jun08 Contig14 revised geneMsil2969	2	9.7872	25234	6.6251
2970_MSIL	msil 20jun08 Contig14 revised geneMsil2970	2	9.5455	23277	4.7113
2977_MSIL	msil 20jun08 Contig14 revised geneMsil2977	2	5.3191	49334	5.8105
2991_MSIL	msil 20jun08 Contig14 revised geneMsil2991	3	8.8757	37111	4.9232
2993_MSIL	msil 20jun08 Contig14 revised geneMsil2993	2	3.7815	49645	6.4279
2996_MSIL	msil 20jun08 Contig14 revised geneMsil2996	11	43.8462	40184	6.267
2997_MSIL	msil 20jun08 Contig14 revised geneMsil2997	11	43.5685	25472	8.4214
3011_MSIL	msil 20jun08 Contig14 revised geneMsil3011	7	19.174	36619	5.714
3157_MSIL	msil 20jun08 Contig14 revised geneMsil3157	10	17.8967	59779	5.8885
3166_MSIL	msil 20jun08 Contig14 revised geneMsil3166	2	8.042	32038	5.525
3172_MSIL	msil 20jun08 Contig14 revised geneMsil3172	5	6.1728	88327	5.6957
3203_MSIL	msil 20jun08 Contig14 revised geneMsil3203	2	7.8049	21612	9.2192
3210_MSIL	msil 20jun08 Contig14 revised geneMsil3210	2	8.7719	37778	5.5278
3226_MSIL	msil 20jun08 Contig14 revised geneMsil3226	3	10.4377	32668	6.0456
3255_MSIL	msil 20jun08 Contig14 revised geneMsil3255	4	10.0446	46035	4.9156
3285_MSIL	msil 20jun08 Contig14 revised geneMsil3285	6	23.8298	25547	8.4968
3309_MSIL	msil 20jun08 Contig14 revised geneMsil3309	10	37.6866	28857	4.6071
3336_MSIL	msil 20jun08 Contig14 revised geneMsil3336	6	19.9387	36137	6.5101
3509_MSIL	msil 20jun08 Contig14 revised geneMsil3509	4	4.3956	98931	6.9772
3510_MSIL	msil 20jun08 Contig14 revised geneMsil3510	3	9.1892	40188	4.9473
3523_MSIL	msil 20jun08 Contig14 revised geneMsil3523	4	4.3575	96361	5.417
3530_MSIL	msil 20jun08 Contig14 revised geneMsil3530	12	31.4496	45231	6.7689
3538_MSIL	msil 20jun08 Contig14 revised geneMsil3538	2	3.0864	73034	5.8524
3545_MSIL	msil 20jun08 Contig14 revised geneMsil3545	13	8.996	136026	5.0726
3550_MSIL	msil 20jun08 Contig14 revised geneMsil3550	4	8.9855	37436	6.0612

3552_MSIL	msil 20jun08 Contig14 revised geneMsil3552	3	13.4409	19778	9.7436
3553_MSIL	msil 20jun08 Contig14 revised geneMsil3553	4	29.8013	16533	7.5068
3562_MSIL	msil 20jun08 Contig14 revised geneMsil3562	2	7.722	29297	7.5939
3563_MSIL	msil 20jun08 Contig14 revised geneMsil3563	2	3.7582	66456	6.2254
3597_MSIL	msil 20jun08 Contig14 revised geneMsil3597	6	9.366	75934	5.6633
3674_MSIL	msil 20jun08 Contig14 revised geneMsil3674	3	8.0882	28127	9.8057
3687_MSIL	msil 20jun08 Contig14 revised geneMsil3687	4	5.5152	76585	5.163
3697_MSIL	msil 20jun08 Contig14 revised geneMsil3697	2	8.1356	31834	5.9965
3698_MSIL	msil 20jun08 Contig14 revised geneMsil3698	5	18.4066	39171	5.0233
3705_MSIL	msil 20jun08 Contig14 revised geneMsil3705	4	10.4839	40354	5.1559
3712_MSIL	msil 20jun08 Contig14 revised geneMsil3712	2	11.2676	16557	9.3783
3739_MSIL	msil 20jun08 Contig14 revised geneMsil3739	2	11.0526	21191	5.7189
3801_MSIL	msil 20jun08 Contig14 revised geneMsil3801	2	13.2948	18317	6.4805
3809_MSIL	msil 20jun08 Contig14 revised geneMsil3809	4	24.2718	22722	6.3755
3820_MSIL	msil 20jun08 Contig14 revised geneMsil3820	3	16.9399	19338	5.9685
3828_MSIL	msil 20jun08 Contig14 revised geneMsil3828	4	25.1337	20392	8.2513
3865_MSIL	msil 20jun08 Contig14 revised geneMsil3865	3	13.8528	23929	9.9968
3866_MSIL	msil 20jun08 Contig14 revised geneMsil3866	3	21.6374	17717	9.906
3868_MSIL	msil 20jun08 Contig14 revised geneMsil3868	18	12.5727	153612	5.0548
3869_MSIL	msil 20jun08 Contig14 revised geneMsil3869	12	8.0888	154345	8.0233
3881_MSIL	msil 20jun08 Contig14 revised geneMsil3881	3	6.1144	53327	5.4987

Supplementary Table 3

Protein identifications – iTRAQ labelling

Entry	Description	Number of peptides	Sequence coverage (%)	Molecular weight (Da)	pI	Isobaric tags 115:114 Log(e)	Isobaric tags 116:114 Log(e)	Isobaric tags 117:114 Log(e)
0008_MSIL	msil_20jun08_Contig14_revised_geneMsil0008	2	5.4878	34715	6.1877	-0.11	-0.13	-0.08
0011_MSIL	msil_20jun08_Contig14_revised_geneMsil0011	5	21.2121	13835	8.7636	-0.31	-0.32	-0.39
0061_MSIL	msil_20jun08_Contig14_revised_geneMsil0061	7	18.4739	25530	6.1859	0.65	-0.05	0.59
0082_MSIL	msil_20jun08_Contig14_revised_geneMsil0082	5	7.1023	39510	5.7627	0.05	0.26	-0.02
0096_MSIL	msil_20jun08_Contig14_revised_geneMsil0096	2	8.1633	21825	5.5093	-0.56	-0.19	-0.61
0098_MSIL	msil_20jun08_Contig14_revised_geneMsil0098	2	7.3171	23289	4.9658	-0.23	-0.07	-0.17
0110_MSIL	msil_20jun08_Contig14_revised_geneMsil0110	14	18.6567	43565	5.5878	2.26	0.13	2.32
0112_MSIL	msil_20jun08_Contig14_revised_geneMsil0112	3	13.125	18205	5.0261	0.54	0.1	0.53
0162_MSIL	msil_20jun08_Contig14_revised_geneMsil0162	3	4.0892	58608	5.7361	0.25	-0.04	0.39
0191_MSIL	msil_20jun08_Contig14_revised_geneMsil0191	2	0.8172	107892	7.0968	0.02	-0.26	0.19
0195_MSIL	msil_20jun08_Contig14_revised_geneMsil0195	5	19.1257	19801	5.4911	0.23	-0.07	0.33
0209_MSIL	msil_20jun08_Contig14_revised_geneMsil0209	2	3.3708	70016	5.3441	-0.46	-0.1	-0.51
0240_MSIL	msil_20jun08_Contig14_revised_geneMsil0240	2	0.4322	189740	5.3062	0.13	0.15	0.15
0268_MSIL	msil_20jun08_Contig14_revised_geneMsil0268	5	4.3417	77238	5.161	-0.28	-0.26	-0.28
0286_MSIL	msil_20jun08_Contig14_revised_geneMsil0286	2	5.4745	29864	4.5688	0.13	-0.03	0.09
0341_MSIL	msil_20jun08_Contig14_revised_geneMsil0341	11	16.8959	54708	5.7219	-0.47	0.06	-0.54
0343_MSIL	msil_20jun08_Contig14_revised_geneMsil0343	5	13.2505	51312	4.745	-0.26	-0.18	-0.32
0380_MSIL	msil_20jun08_Contig14_revised_geneMsil0380	4	10.1887	28019	6.1848	0.4	0.16	0.39
0410_MSIL	msil_20jun08_Contig14_revised_geneMsil0410	18	26.0246	52382	6.4645	1.59	-0.01	1.59

0427_MSIL	msil_20jun08_Contig14_revised_geneMsil0427	2	3.1983	52033	5.2363	0.2	-0.38	0.08
0471_MSIL	msil_20jun08_Contig14_revised_geneMsil0471	52	34.08	68485	5.8453	1.88	-0.01	1.87
0472_MSIL	msil_20jun08_Contig14_revised_geneMsil0472	7	13.6842	31041	5.1363	0.78	0.02	0.82
0474_MSIL	msil_20jun08_Contig14_revised_geneMsil0474	14	31.6832	10891	9.4294	1.86	-0.03	1.84
0475_MSIL	msil_20jun08_Contig14_revised_geneMsil0475	2	4.6377	38805	5.5419	0.23	-0.49	0.18
0515_MSIL	msil_20jun08_Contig14_revised_geneMsil0515	3	6.7916	44925	4.5795	-0.34	-0.03	-0.33
0521_MSIL	msil_20jun08_Contig14_revised_geneMsil0521	2	2.7027	45767	6.8047	0.45	-0.09	0.48
0522_MSIL	msil_20jun08_Contig14_revised_geneMsil0522	2	2.7083	50915	6.6623	-0.21	0.02	-0.34
0538_MSIL	msil_20jun08_Contig14_revised_geneMsil0538	2	8.2949	22785	9.704	-0.63	0.01	-0.76
0563_MSIL	msil_20jun08_Contig14_revised_geneMsil0563	2	7.9365	20307	11.7014	-0.19	-0.17	-0.38
0582_MSIL	msil_20jun08_Contig14_revised_geneMsil0582	10	14.3939	43107	5.3617	-0.39	0.12	-0.36
0583_MSIL	msil_20jun08_Contig14_revised_geneMsil0583	3	4.631	76241	5.0898	-0.91	-0.24	-0.82
0620_MSIL	msil_20jun08_Contig14_revised_geneMsil0620	4	16.7247	28612	4.7858	-0.1	-0.11	-0.15
0643_MSIL	msil_20jun08_Contig14_revised_geneMsil0643	2	4.2506	47756	6.0291	-0.48	-0.04	-0.53
0673_MSIL	msil_20jun08_Contig14_revised_geneMsil0673	3	17.1053	16433	4.869	-0.45	0.03	-0.48
0675_MSIL	msil_20jun08_Contig14_revised_geneMsil0675	2	2.4055	64094	4.9874	-0.37	0.17	-0.67
0678_MSIL	msil_20jun08_Contig14_revised_geneMsil0678	2	8.1818	12021	6.1296	0.89	-0.14	0.86
0729_MSIL	msil_20jun08_Contig14_revised_geneMsil0729	2	1.7593	117473	4.5879	-0.57	-0.16	-0.32
0730_MSIL	msil_20jun08_Contig14_revised_geneMsil0730	2	5	27821	6.2047	-0.57	-0.16	-0.32
0736_MSIL	msil_20jun08_Contig14_revised_geneMsil0736	2	6.7114	14673	12.2175	-0.23	-0.2	-0.29
0767_MSIL	msil_20jun08_Contig14_revised_geneMsil0767	8	14.2857	34409	8.9361	1.84	-0.23	1.78
0770_MSIL	msil_20jun08_Contig14_revised_geneMsil0770	5	10.1266	42305	6.6301	0.71	0.06	0.83
0794_MSIL	msil_20jun08_Contig14_revised_geneMsil0794	8	45.2632	10218	5.2236	0.31	-0.03	0.31
0795_MSIL	msil_20jun08_Contig14_revised_geneMsil0795	34	41.3163	57572	4.9437	-0.01	-0.09	-0.07
0806_MSIL	msil_20jun08_Contig14_revised_geneMsil0806	2	6.7669	26953	3.9626	0.14	0	0.04
0823_MSIL	msil_20jun08_Contig14_revised_geneMsil0823	2	8.6957	14691	4.8723	0.14	0	0.04

0848_MSIL	msil_20jun08_Contig14_revised_geneMsil0848	3	9.0604	31524	6.3311	0.04	-0.17	0.04
0893_MSIL	msil_20jun08_Contig14_revised_geneMsil0893	2	1.6706	44084	6.0714	-0.38	-0.17	-0.39
0941_MSIL	msil_20jun08_Contig14_revised_geneMsil0941	3	28.125	6581	9.934	2.65	0.38	2.84
0960_MSIL	msil_20jun08_Contig14_revised_geneMsil0960	2	4.1667	25419	4.7461	-0.76	-0.26	-0.78
0971_MSIL	msil_20jun08_Contig14_revised_geneMsil0971	2	8.6957	20591	4.5989	0.45	-0.19	0.36
1000_MSIL	msil_20jun08_Contig14_revised_geneMsil1000	2	7.8049	23094	5.0301	-0.7	-0.12	-0.65
1012_MSIL	msil_20jun08_Contig14_revised_geneMsil1012	2	4.4674	31519	5.4069	-0.35	-0.34	-0.44
1025_MSIL	msil_20jun08_Contig14_revised_geneMsil1025	5	18.4358	20004	5.2809	0.49	-0.14	0.37
1117_MSIL	msil_20jun08_Contig14_revised_geneMsil1117	2	1.3675	63830	5.5768	-0.51	-0.33	-0.62
1171_MSIL	msil_20jun08_Contig14_revised_geneMsil1171	2	10.1266	17679	5.6594	-0.26	0.21	-0.29
1188_MSIL	msil_20jun08_Contig14_revised_geneMsil1188	2	5.3648	49713	5.6631	0.13	-0.07	0.19
1193_MSIL	msil_20jun08_Contig14_revised_geneMsil1193	4	5.0975	70184	5.9489	-0.1	-0.13	0.15
1226_MSIL	msil_20jun08_Contig14_revised_geneMsil1226	5	4.9612	70867	5.6898	-0.4	-0.14	-0.07
1262_MSIL	msil_20jun08_Contig14_revised_geneMsil1262	29	27.3764	59751	5.7755	-0.05	0.02	-0.01
1263_MSIL	msil_20jun08_Contig14_revised_geneMsil1263	35	38.1074	44905	5.7274	2.23	-0.1	2.38
1264_MSIL	msil_20jun08_Contig14_revised_geneMsil1264	10	26.2774	15296	4.5163	2.31	0.22	2.37
1265_MSIL	msil_20jun08_Contig14_revised_geneMsil1265	19	53.5294	19537	9.3867	1.74	0.09	1.78
1312_MSIL	msil_20jun08_Contig14_revised_geneMsil1312	15	36.2745	32521	5.7444	2.08	0.48	2.11
1321_MSIL	msil_20jun08_Contig14_revised_geneMsil1321	4	9.1413	38462	5.1075	1.45	-0.02	1.45
1328_MSIL	msil_20jun08_Contig14_revised_geneMsil1328	3	5.2419	52701	6.206	1.27	-0.18	1.24
1365_MSIL	msil_20jun08_Contig14_revised_geneMsil1365	2	11.1111	8500	6.1018	-0.11	-0.06	-0.13
1375_MSIL	msil_20jun08_Contig14_revised_geneMsil1375	4	6.3433	58198	6.0253	0.28	0.24	0.2
1382_MSIL	msil_20jun08_Contig14_revised_geneMsil1382	3	5.0481	46188	5.373	-1.4	-0.07	-1.37
1405_MSIL	msil_20jun08_Contig14_revised_geneMsil1405	3	5.5385	34207	4.7607	-0.23	0.39	-0.21
1576_MSIL	msil_20jun08_Contig14_revised_geneMsil1576	3	28.9855	6993	7.7192	0.31	0.08	0.23
1595_MSIL	msil_20jun08_Contig14_revised_geneMsil1595	2	1.9149	51573	5.4084	0.01	0.21	0.1

1611_MSIL	msil_20jun08_Contig14_revised_geneMsil1611	2	10.582	20276	4.8536	-0.11	-0.02	-0.26
1619_MSIL	msil_20jun08_Contig14_revised_geneMsil1619	2	4.3478	17417	9.4969	-0.13	-0.27	-0.24
1624_MSIL	msil_20jun08_Contig14_revised_geneMsil1624	2	4.0302	42515	4.3233	-0.14	0.07	-0.15
1676_MSIL	msil_20jun08_Contig14_revised_geneMsil1676	4	15.0754	22228	5.6911	1.15	0.03	1.01
1681_MSIL	msil_20jun08_Contig14_revised_geneMsil1681	5	8.6247	47690	6.3444	-0.41	-0.11	-0.5
1693_MSIL	msil_20jun08_Contig14_revised_geneMsil1693	3	5.1613	31980	5.3806	1.96	0.03	2.14
1706_MSIL	msil_20jun08_Contig14_revised_geneMsil1706	4	6.0475	49832	5.1275	-0.12	0.09	-0.14
1712_MSIL	msil_20jun08_Contig14_revised_geneMsil1712	2	2.509	59899	6.6418	0.34	-0.08	0.22
1713_MSIL	msil_20jun08_Contig14_revised_geneMsil1713	2	7.3482	33824	5.2581	-0.03	-0.3	-0.16
1714_MSIL	msil_20jun08_Contig14_revised_geneMsil1714	2	2.7778	42582	6.893	0	-0.2	-0.1
1716_MSIL	msil_20jun08_Contig14_revised_geneMsil1716	6	8.7719	42881	5.2007	-0.39	0.17	-0.38
1743_MSIL	msil_20jun08_Contig14_revised_geneMsil1743	2	4.2328	38453	5.4631	0.14	-0.06	0.05
1745_MSIL	msil_20jun08_Contig14_revised_geneMsil1745	4	20.3822	16460	11.1802	-0.09	-0.27	0
1806_MSIL	msil_20jun08_Contig14_revised_geneMsil1806	3	35.0365	14138	5.1848	-0.23	0.16	-0.26
1808_MSIL	msil_20jun08_Contig14_revised_geneMsil1808	8	8.8968	30823	5.7768	-0.32	-0.23	-0.01
1834_MSIL	msil_20jun08_Contig14_revised_geneMsil1834	2	17.5926	11686	10.4636	-0.82	0.1	-0.74
1861_MSIL	msil_20jun08_Contig14_revised_geneMsil1861	3	14.2857	12014	9.1807	-0.13	-0.07	-0.17
1881_MSIL	msil_20jun08_Contig14_revised_geneMsil1881	2	5.0314	32781	8.8033	0.69	0.07	0.7
1907_MSIL	msil_20jun08_Contig14_revised_geneMsil1907	4	4.0509	95453	5.6468	-0.08	-0.23	-0.16
1910_MSIL	msil_20jun08_Contig14_revised_geneMsil1910	6	10	71093	4.9285	-0.3	0.02	-0.39
1941_MSIL	msil_20jun08_Contig14_revised_geneMsil1941	3	8.0645	7213	6.5078	-0.06	-0.12	0.16
1945_MSIL	msil_20jun08_Contig14_revised_geneMsil1945	2	0.843	107081	5.8046	1.6	-0.11	1.6
2091_MSIL	msil_20jun08_Contig14_revised_geneMsil2091	4	13.5542	35103	5.5523	2.07	2.12	2.21
2110_MSIL	msil_20jun08_Contig14_revised_geneMsil2110	14	28.1106	45996	6.7786	-0.2	-0.24	-0.27
2111_MSIL	msil_20jun08_Contig14_revised_geneMsil2111	2	7.3529	33339	5.5651	-0.51	-0.11	-0.6
2224_MSIL	msil_20jun08_Contig14_revised_geneMsil2224	2	6.0317	34241	9.8112	0.65	-0.22	0.71

2232_MSIL	msil_20jun08_Contig14_revised_geneMsil2232	2	1.726	101306	5.0079	-0.23	-0.15	0.15
2233_MSIL	msil_20jun08_Contig14_revised_geneMsil2233	3	13.6905	18690	5.883	0.24	-0.34	0.04
2246_MSIL	msil_20jun08_Contig14_revised_geneMsil2246	11	16.0458	37409	8.7356	-0.2	-0.37	-0.18
2282_MSIL	msil_20jun08_Contig14_revised_geneMsil2282	9	17.0149	35502	6.9589	2.41	0.43	2.41
2283_MSIL	msil_20jun08_Contig14_revised_geneMsil2283	3	5.4054	43185	5.7662	-0.55	-0.15	-0.65
2309_MSIL	msil_20jun08_Contig14_revised_geneMsil2309	2	23.6842	8468	5.792	-0.82	-0.12	-0.85
2329_MSIL	msil_20jun08_Contig14_revised_geneMsil2329	6	10.4784	47932	5.4802	0.15	-0.02	0.17
2342_MSIL	msil_20jun08_Contig14_revised_geneMsil2342	8	8.6957	55624	5.5428	-0.39	-0.14	-0.36
2345_MSIL	msil_20jun08_Contig14_revised_geneMsil2345	9	61.8644	12692	5.2178	-0.72	0.07	-0.73
2360_MSIL	msil_20jun08_Contig14_revised_geneMsil2360	2	8.7432	19910	4.6688	0.04	-0.12	0.03
2361_MSIL	msil_20jun08_Contig14_revised_geneMsil2361	2	3.373	56514	5.4291	0.26	-0.14	0.25
2362_MSIL	msil_20jun08_Contig14_revised_geneMsil2362	2	4.3478	17271	6.3389	0.11	0.08	0.19
2385_MSIL	msil_20jun08_Contig14_revised_geneMsil2385	4	7.7966	30948	4.9871	0.68	0.53	0.62
2390_MSIL	msil_20jun08_Contig14_revised_geneMsil2390	14	38.8571	18558	5.8748	-0.09	-0.1	-0.17
2430_MSIL	msil_20jun08_Contig14_revised_geneMsil2430	2	1.6632	52561	7.7446	0.61	-0.1	0.38
2442_MSIL	msil_20jun08_Contig14_revised_geneMsil2442	3	13.3333	13844	5.6887	0.52	0.17	0.61
2447_MSIL	msil_20jun08_Contig14_revised_geneMsil2447	2	5.848	36002	6.5391	0.46	-0.17	0.45
2464_MSIL	msil_20jun08_Contig14_revised_geneMsil2464	2	2.0513	41928	7.8556	0.57	0.14	0.34
2501_MSIL	msil_20jun08_Contig14_revised_geneMsil2501	4	7.7882	33442	5.088	0.18	-0.16	0.18
2503_MSIL	msil_20jun08_Contig14_revised_geneMsil2503	3	7.1429	30052	5.3925	-0.48	-0.15	-0.46
2504_MSIL	msil_20jun08_Contig14_revised_geneMsil2504	3	2.6523	113040	6.3444	-0.34	-0.11	-0.43
2505_MSIL	msil_20jun08_Contig14_revised_geneMsil2505	4	9.5794	44809	6.1965	-0.24	-0.13	-0.21
2506_MSIL	msil_20jun08_Contig14_revised_geneMsil2506	2	3.1915	49375	5.8597	0.48	-0.07	0.52
2523_MSIL	msil_20jun08_Contig14_revised_geneMsil2523	4	12.4224	34028	5.8129	-0.82	-0.36	-0.86
2592_MSIL	msil_20jun08_Contig14_revised_geneMsil2592	2	4.3478	18866	4.6838	-0.51	0.07	-0.47
2608_MSIL	msil_20jun08_Contig14_revised_geneMsil2608	2	5.1213	39929	5.7501	0.93	0.13	1.02

2647_MSIL	msil_20jun08_Contig14_revised_geneMsil2647	6	17.2727	24375	5.6852	-0.27	-0.12	-0.39
2655_MSIL	msil_20jun08_Contig14_revised_geneMsil2655	2	3.0014	79319	4.9565	0.57	-0.05	0.57
2800_MSIL	msil_20jun08_Contig14_revised_geneMsil2800	2	5.3476	20491	4.8182	0.37	-0.16	0.28
2809_MSIL	msil_20jun08_Contig14_revised_geneMsil2809	2	1.6746	88070	9.1743	-0.23	-0.14	-0.35
2810_MSIL	msil_20jun08_Contig14_revised_geneMsil2810	2	2.6052	53968	5.5942	0.68	0.2	0.68
2816_MSIL	msil_20jun08_Contig14_revised_geneMsil2816	3	3.9861	61891	6.2639	0.13	0.06	0.17
2876_MSIL	msil_20jun08_Contig14_revised_geneMsil2876	2	3.2751	48749	6.6044	0.24	-0.04	0.19
2884_MSIL	msil_20jun08_Contig14_revised_geneMsil2884	3	1.7316	122346	6.446	0.49	0	0.24
2912_MSIL	msil_20jun08_Contig14_revised_geneMsil2912	3	6.71	50614	4.5209	1.59	-0.08	1.58
2913_MSIL	msil_20jun08_Contig14_revised_geneMsil2913	2	6.5116	24061	5.6364	-0.41	-0.06	-0.31
2915_MSIL	msil_20jun08_Contig14_revised_geneMsil2915	3	0.9938	88961	5.9123	-0.2	-0.29	-0.2
2952_MSIL	msil_20jun08_Contig14_revised_geneMsil2952	2	24.2857	7311	5.8109	-0.07	0.21	0.01
2955_MSIL	msil_20jun08_Contig14_revised_geneMsil2955	13	13.8801	67769	4.823	0.09	-0.1	0.22
2961_MSIL	msil_20jun08_Contig14_revised_geneMsil2961	5	5.25	43501	5.7889	-0.34	0.05	-0.35
2996_MSIL	msil_20jun08_Contig14_revised_geneMsil2996	11	29.2308	40184	6.267	0.74	0.29	0.75
2997_MSIL	msil_20jun08_Contig14_revised_geneMsil2997	6	21.5768	25472	8.4214	-0.2	-0.01	-0.25
3011_MSIL	msil_20jun08_Contig14_revised_geneMsil3011	2	5.0147	36619	5.714	-0.14	-0.08	-0.07
3030_MSIL	msil_20jun08_Contig14_revised_geneMsil3030	3	6.0606	37155	4.7115	0.16	-0.39	0.01
3098_MSIL	msil_20jun08_Contig14_revised_geneMsil3098	4	0.9346	47122	10.6414	1.22	-0.27	1.34
3127_MSIL	msil_20jun08_Contig14_revised_geneMsil3127	4	13.9394	16923	7.3207	1.87	-0.09	1.77
3128_MSIL	msil_20jun08_Contig14_revised_geneMsil3128	2	10.8696	19951	5.5578	1.11	-0.25	1.27
3129_MSIL	msil_20jun08_Contig14_revised_geneMsil3129	4	23.2704	16933	9.3918	2.28	0.01	2.25
3131_MSIL	msil_20jun08_Contig14_revised_geneMsil3131	7	9.4697	88101	6.4166	1.8	-0.69	1.78
3132_MSIL	msil_20jun08_Contig14_revised_geneMsil3132	3	4.9793	52339	6.3926	1.93	0.24	2.21
3135_MSIL	msil_20jun08_Contig14_revised_geneMsil3135	4	13.9535	17983	9.2488	1.58	-0.21	1.49
3145_MSIL	msil_20jun08_Contig14_revised_geneMsil3145	2	4.2813	34932	4.9797	2.41	0.06	2.49

3157_MSIL	msil_20jun08_Contig14_revised_geneMsil3157	6	9.4096	59779	5.8885	1.21	0.02	1.65
3181_MSIL	msil_20jun08_Contig14_revised_geneMsil3181	2	2.069	63099	5.7794	-0.41	0.04	-0.47
3199_MSIL	msil_20jun08_Contig14_revised_geneMsil3199	4	17.1429	14553	5.8634	1.44	-0.05	1.58
3206_MSIL	msil_20jun08_Contig14_revised_geneMsil3206	4	11.2583	16033	5.1931	-0.32	-0.19	-0.31
3218_MSIL	msil_20jun08_Contig14_revised_geneMsil3218	2	20	11710	8.1954	0.64	-0.12	0.64
3220_MSIL	msil_20jun08_Contig14_revised_geneMsil3220	2	3.1496	55425	5.5692	-0.99	-0.13	-1.05
3226_MSIL	msil_20jun08_Contig14_revised_geneMsil3226	2	6.734	32668	6.0456	0.62	-0.03	0.7
3234_MSIL	msil_20jun08_Contig14_revised_geneMsil3234	2	2.7597	65172	5.8015	0.66	-0.03	0.52
3255_MSIL	msil_20jun08_Contig14_revised_geneMsil3255	4	9.1518	46035	4.9156	0.71	0.02	0.7
3285_MSIL	msil_20jun08_Contig14_revised_geneMsil3285	3	12.3404	25547	8.4968	-0.4	-0.23	-0.55
3309_MSIL	msil_20jun08_Contig14_revised_geneMsil3309	4	8.5821	28857	4.6071	-0.28	-0.12	-0.35
3400_MSIL	msil_20jun08_Contig14_revised_geneMsil3400	2	1.4925	43379	6.1718	-0.19	-0.05	-0.26
3416_MSIL	msil_20jun08_Contig14_revised_geneMsil3416	3	1.3158	49794	8.007	0.25	-0.05	0.18
3440_MSIL	msil_20jun08_Contig14_revised_geneMsil3440	2	1.2793	54017	5.2592	1.85	0.14	1.97
3502_MSIL	msil_20jun08_Contig14_revised_geneMsil3502	2	12.4031	13946	4.4063	-0.14	0.01	-0.26
3530_MSIL	msil_20jun08_Contig14_revised_geneMsil3530	8	18.6732	45231	6.7689	-0.27	-0.03	-0.36
3552_MSIL	msil_20jun08_Contig14_revised_geneMsil3552	2	9.6774	19778	9.7436	-0.22	0.2	-0.05
3697_MSIL	msil_20jun08_Contig14_revised_geneMsil3697	2	8.1356	31834	5.9965	-0.35	-0.07	-0.37
3698_MSIL	msil_20jun08_Contig14_revised_geneMsil3698	6	19.7802	39171	5.0233	-0.39	-0.13	-0.41
3705_MSIL	msil_20jun08_Contig14_revised_geneMsil3705	3	8.0645	40354	5.1559	0.52	-0.02	0.46
3821_MSIL	msil_20jun08_Contig14_revised_geneMsil3821	2	14.6667	7377	7.0692	0.33	-0.04	0.09
3828_MSIL	msil_20jun08_Contig14_revised_geneMsil3828	4	13.9037	20392	8.2513	-0.21	0.21	-0.05
3846_MSIL	msil_20jun08_Contig14_revised_geneMsil3846	2	1.0165	85703	9.1377	-0.62	-0.07	-0.62
3867_MSIL	msil_20jun08_Contig14_revised_geneMsil3867	2	10.3175	12939	4.7399	1.1	0.15	1.13
3868_MSIL	msil_20jun08_Contig14_revised_geneMsil3868	4	2.6163	153612	5.0548	-0.26	-0.14	-0.2
3869_MSIL	msil_20jun08_Contig14_revised_geneMsil3869	5	2.2906	154345	8.0233	-0.14	-0.35	-0.21

3875_MSIL	msil_20jun08_Contig14_revised_geneMsil3875	3	1.2771	170158	5.7504	-0.49	-0.15	-0.56
3881_MSIL	msil_20jun08_Contig14_revised_geneMsil3881	9	12.6233	53327	5.4987	1.41	-0.03	1.61
3907_MSIL	msil_20jun08_Contig14_revised_geneMsil3907	2	6.25	14650	10.5747	0.32	0.03	0.38

Supplementary Table 4

Protein identifications – label-free quantification – Methane growth condition

Description	Molecular weight (Da)	Average number of peptides	Average sequence coverage (%)	Technical replicates
msil_20jun08_Contig14_revised_geneMsil0008	34737	10	32.01	2
msil_20jun08_Contig14_revised_geneMsil0011	13843	7.666666667	75.50666667	3
msil_20jun08_Contig14_revised_geneMsil0058	27709	8.666666667	40.62666667	3
msil_20jun08_Contig14_revised_geneMsil0061	25546	6.666666667	49.66666667	3
msil_20jun08_Contig14_revised_geneMsil0082	39534	15.66666667	44.79333333	3
msil_20jun08_Contig14_revised_geneMsil0096	21839	6.5	35.46	2
msil_20jun08_Contig14_revised_geneMsil0106	54270	14.5	34.67	2
msil_20jun08_Contig14_revised_geneMsil0110	43592	21	48.34	3
msil_20jun08_Contig14_revised_geneMsil0154	14853	6.333333333	53.52	3
msil_20jun08_Contig14_revised_geneMsil0162	58645	20.5	39.87	2
msil_20jun08_Contig14_revised_geneMsil0178	34892	16.66666667	70.19	3
msil_20jun08_Contig14_revised_geneMsil0179	46560	15.33333333	46.56	3
msil_20jun08_Contig14_revised_geneMsil0181	9466	2.333333333	36.67	3
msil_20jun08_Contig14_revised_geneMsil0195	19814	8.5	55.735	2
msil_20jun08_Contig14_revised_geneMsil0239	32226	6.5	32.56	2
msil_20jun08_Contig14_revised_geneMsil0268	77286	15	24.3	2
msil_20jun08_Contig14_revised_geneMsil0286	29882	6.333333333	37.10333333	3
msil_20jun08_Contig14_revised_geneMsil0288	10907	4	42.925	2
msil_20jun08_Contig14_revised_geneMsil0341	54742	14.66666667	29.66666667	3
msil_20jun08_Contig14_revised_geneMsil0343	51344	20	58.18	3

msil_20jun08_Contig14_revised_geneMsil0352	15502	5.333333333	32.4	3
msil_20jun08_Contig14_revised_geneMsil0380	28036	12.66666667	48.05	3
msil_20jun08_Contig14_revised_geneMsil0410	52414	25.66666667	64.75333333	3
msil_20jun08_Contig14_revised_geneMsil0427	52067	15	38.59	2
msil_20jun08_Contig14_revised_geneMsil0428	12019	5.5	55.805	2
msil_20jun08_Contig14_revised_geneMsil0468	12976	4	35.45	3
msil_20jun08_Contig14_revised_geneMsil0471	68529	53	69.97333333	3
msil_20jun08_Contig14_revised_geneMsil0472	31060	17.66666667	56.72333333	3
msil_20jun08_Contig14_revised_geneMsil0473	22062	5.666666667	30.03333333	3
msil_20jun08_Contig14_revised_geneMsil0474	10898	6	23.76	3
msil_20jun08_Contig14_revised_geneMsil0499	35950	8.5	37.755	2
msil_20jun08_Contig14_revised_geneMsil0515	44953	14	35.20666667	3
msil_20jun08_Contig14_revised_geneMsil0520	48780	10.5	28.915	2
msil_20jun08_Contig14_revised_geneMsil0541	58002	17	40.27	2
msil_20jun08_Contig14_revised_geneMsil0566	14589	3.5	20.45	2
msil_20jun08_Contig14_revised_geneMsil0582	43134	16.66666667	44.36	3
msil_20jun08_Contig14_revised_geneMsil0583	76289	22.66666667	38.73666667	3
msil_20jun08_Contig14_revised_geneMsil0584	17850	4.5	30.13	2
msil_20jun08_Contig14_revised_geneMsil0612	34378	7.5	26.75	2
msil_20jun08_Contig14_revised_geneMsil0619	19849	9.666666667	63.33333333	3
msil_20jun08_Contig14_revised_geneMsil0620	28630	12.33333333	67.36333333	3
msil_20jun08_Contig14_revised_geneMsil0629	44441	9	27.535	2
msil_20jun08_Contig14_revised_geneMsil0636	50409	12.5	30.34	2
msil_20jun08_Contig14_revised_geneMsil0643	47785	11	30.875	2
msil_20jun08_Contig14_revised_geneMsil0678	12028	7.5	76.365	2
msil_20jun08_Contig14_revised_geneMsil0764	31189	8.5	28.3	2

msil_20jun08_Contig14_revised_geneMsil0767	34430	18	52.28	3
msil_20jun08_Contig14_revised_geneMsil0770	42331	15.33333333	53.33333333	3
msil_20jun08_Contig14_revised_geneMsil0774	33332	9.666666667	35.97333333	3
msil_20jun08_Contig14_revised_geneMsil0794	10224	5.666666667	56.84333333	3
msil_20jun08_Contig14_revised_geneMsil0795	57608	36.66666667	64.77666667	3
msil_20jun08_Contig14_revised_geneMsil0811	27406	6	30.265	2
msil_20jun08_Contig14_revised_geneMsil0823	14700	5.666666667	36.95666667	3
msil_20jun08_Contig14_revised_geneMsil0839	36345	9	29.395	2
msil_20jun08_Contig14_revised_geneMsil0870	19884	7.5	48.59	2
msil_20jun08_Contig14_revised_geneMsil0887	9398	4	46.51	2
msil_20jun08_Contig14_revised_geneMsil1012	31539	8.666666667	35.85666667	3
msil_20jun08_Contig14_revised_geneMsil1025	20017	7	47.765	2
msil_20jun08_Contig14_revised_geneMsil1043	19114	4.333333333	37.25666667	3
msil_20jun08_Contig14_revised_geneMsil1068	30027	7	29.05	3
msil_20jun08_Contig14_revised_geneMsil1124	63363	16	39.715	2
msil_20jun08_Contig14_revised_geneMsil1136	9819	5	42.42333333	3
msil_20jun08_Contig14_revised_geneMsil1140	26523	8.666666667	39.66666667	3
msil_20jun08_Contig14_revised_geneMsil1146	14184	4	34.09333333	3
msil_20jun08_Contig14_revised_geneMsil1149	28664	6	26.88	2
msil_20jun08_Contig14_revised_geneMsil1171	17691	8	37.76333333	3
msil_20jun08_Contig14_revised_geneMsil1178	11619	7	72.38333333	3
msil_20jun08_Contig14_revised_geneMsil1186	12265	5.5	62.5	2
msil_20jun08_Contig14_revised_geneMsil1224	10019	3	47.195	2
msil_20jun08_Contig14_revised_geneMsil1226	70912	23	45.32333333	3
msil_20jun08_Contig14_revised_geneMsil1262	59789	39.66666667	63.49666667	3
msil_20jun08_Contig14_revised_geneMsil1263	44933	35	78.17666667	3

msil_20jun08_Contig14_revised_geneMsil1264	15306	17.33333333	74.45	3
msil_20jun08_Contig14_revised_geneMsil1265	19549	12.66666667	60.78333333	3
msil_20jun08_Contig14_revised_geneMsil1267	38516	12.33333333	47.43	3
msil_20jun08_Contig14_revised_geneMsil1270	59731	14.66666667	31.96	3
msil_20jun08_Contig14_revised_geneMsil1278	9552	3.5	59.57	2
msil_20jun08_Contig14_revised_geneMsil1312	32541	17.33333333	73.09	3
msil_20jun08_Contig14_revised_geneMsil1314	12494	6	66.825	2
msil_20jun08_Contig14_revised_geneMsil1321	38486	12	45.15333333	3
msil_20jun08_Contig14_revised_geneMsil1354	43183	15.5	35.845	2
msil_20jun08_Contig14_revised_geneMsil1388	22303	6	23.205	2
msil_20jun08_Contig14_revised_geneMsil1389	16866	8	51.40666667	3
msil_20jun08_Contig14_revised_geneMsil1395	75669	14.33333333	25.78333333	3
msil_20jun08_Contig14_revised_geneMsil1411	11268	3.66666667	42.68	3
msil_20jun08_Contig14_revised_geneMsil1575	29163	5	22.745	2
msil_20jun08_Contig14_revised_geneMsil1603	24476	5.5	25.985	2
msil_20jun08_Contig14_revised_geneMsil1624	42542	6.66666667	24.85333333	3
msil_20jun08_Contig14_revised_geneMsil1628	8288	4.5	39.61	2
msil_20jun08_Contig14_revised_geneMsil1649	40230	12.66666667	41.36333333	3
msil_20jun08_Contig14_revised_geneMsil1651	64278	15	35.42	2
msil_20jun08_Contig14_revised_geneMsil1676	22241	10.33333333	54.10333333	3
msil_20jun08_Contig14_revised_geneMsil1681	47722	15.66666667	44.99	3
msil_20jun08_Contig14_revised_geneMsil1693	32000	13.5	44.675	2
msil_20jun08_Contig14_revised_geneMsil1706	49863	20	46.14666667	3
msil_20jun08_Contig14_revised_geneMsil1712	59937	10.5	20.875	2
msil_20jun08_Contig14_revised_geneMsil1714	42610	12.33333333	40.49	3
msil_20jun08_Contig14_revised_geneMsil1716	42908	12.66666667	38.18	3

msil_20jun08_Contig14_revised_geneMsil1719	35013	9.5	45.075	2
msil_20jun08_Contig14_revised_geneMsil1745	16470	4	30.36	3
msil_20jun08_Contig14_revised_geneMsil1798	7536	3.666666667	56.52333333	3
msil_20jun08_Contig14_revised_geneMsil1806	14147	7	64.96	2
msil_20jun08_Contig14_revised_geneMsil1816	8581	2	22.785	2
msil_20jun08_Contig14_revised_geneMsil1827	36279	7.666666667	29.37	3
msil_20jun08_Contig14_revised_geneMsil1851	24880	3	13.9	2
msil_20jun08_Contig14_revised_geneMsil1861	12022	5	41.96333333	3
msil_20jun08_Contig14_revised_geneMsil1910	71137	20	43.155	2
msil_20jun08_Contig14_revised_geneMsil2091	35124	13.66666667	52.11	3
msil_20jun08_Contig14_revised_geneMsil2098	17176	5.5	57.01	2
msil_20jun08_Contig14_revised_geneMsil2110	46025	14.5	40.09	2
msil_20jun08_Contig14_revised_geneMsil2111	33360	7	20.88333333	3
msil_20jun08_Contig14_revised_geneMsil2114	27914	12	44.4	2
msil_20jun08_Contig14_revised_geneMsil2224	34262	7.333333333	28.67666667	3
msil_20jun08_Contig14_revised_geneMsil2246	37432	20.66666667	66.47333333	3
msil_20jun08_Contig14_revised_geneMsil2282	35524	11	42.535	2
msil_20jun08_Contig14_revised_geneMsil2301	55900	13.33333333	31.74666667	3
msil_20jun08_Contig14_revised_geneMsil2342	55659	19	52.43666667	3
msil_20jun08_Contig14_revised_geneMsil2345	12700	11	66.1	3
msil_20jun08_Contig14_revised_geneMsil2361	56550	15.66666667	45.24	3
msil_20jun08_Contig14_revised_geneMsil2385	30967	13	53.05	2
msil_20jun08_Contig14_revised_geneMsil2390	18570	10.66666667	57.90666667	3
msil_20jun08_Contig14_revised_geneMsil2442	13852	4.666666667	36.54333333	3
msil_20jun08_Contig14_revised_geneMsil2447	36024	10.33333333	43.17333333	3
msil_20jun08_Contig14_revised_geneMsil2453	16949	5.5	46.205	2

msil_20jun08_Contig14_revised_geneMsil2481	13273	5	44.76	2
msil_20jun08_Contig14_revised_geneMsil2501	33463	16.66666667	57.63	3
msil_20jun08_Contig14_revised_geneMsil2503	30071	11	46.26	2
msil_20jun08_Contig14_revised_geneMsil2505	44837	17	41.58666667	3
msil_20jun08_Contig14_revised_geneMsil2523	34049	8.666666667	30.43666667	3
msil_20jun08_Contig14_revised_geneMsil2544	16674	6.666666667	61.99333333	3
msil_20jun08_Contig14_revised_geneMsil2561	32370	8	40.655	2
msil_20jun08_Contig14_revised_geneMsil2592	18878	6.5	50	2
msil_20jun08_Contig14_revised_geneMsil2600	32884	5	23.63	2
msil_20jun08_Contig14_revised_geneMsil2601	38186	9.5	25.35	2
msil_20jun08_Contig14_revised_geneMsil2608	39954	14.33333333	43.75333333	3
msil_20jun08_Contig14_revised_geneMsil2623	59260	7	7.64	2
msil_20jun08_Contig14_revised_geneMsil2647	24390	13.66666667	64.69666667	3
msil_20jun08_Contig14_revised_geneMsil2655	79369	26	37.28666667	3
msil_20jun08_Contig14_revised_geneMsil2671	38305	13	40.41666667	3
msil_20jun08_Contig14_revised_geneMsil2674	21733	8	54.56	3
msil_20jun08_Contig14_revised_geneMsil2678	44881	5.5	24.04	2
msil_20jun08_Contig14_revised_geneMsil2753	23953	4.5	25.21	2
msil_20jun08_Contig14_revised_geneMsil2836	18231	4	28.06	3
msil_20jun08_Contig14_revised_geneMsil2847	37670	11.5	39.36	2
msil_20jun08_Contig14_revised_geneMsil2876	48780	13.5	34.17	2
msil_20jun08_Contig14_revised_geneMsil2912	50645	16	38.96	2
msil_20jun08_Contig14_revised_geneMsil2952	7316	4	62.38333333	3
msil_20jun08_Contig14_revised_geneMsil2953	78051	19	33.025	2
msil_20jun08_Contig14_revised_geneMsil2955	67811	29.33333333	52.41666667	3
msil_20jun08_Contig14_revised_geneMsil2991	37134	6.666666667	25.74	3

msil_20jun08_Contig14_revised_geneMsil2996	40209	19.66666667	69.31666667	3
msil_20jun08_Contig14_revised_geneMsil2997	25488	11.66666667	48.27	3
msil_20jun08_Contig14_revised_geneMsil3011	36643	7	28.90666667	3
msil_20jun08_Contig14_revised_geneMsil3127	16934	10.66666667	62.82666667	3
msil_20jun08_Contig14_revised_geneMsil3128	19963	11.33333333	68.84333333	3
msil_20jun08_Contig14_revised_geneMsil3129	16943	10.66666667	67.08333333	3
msil_20jun08_Contig14_revised_geneMsil3130	31470	16	61.03666667	3
msil_20jun08_Contig14_revised_geneMsil3131	88155	31	49.54	3
msil_20jun08_Contig14_revised_geneMsil3132	52371	20.66666667	57.4	3
msil_20jun08_Contig14_revised_geneMsil3135	17994	7	53.1	3
msil_20jun08_Contig14_revised_geneMsil3180	39523	11.33333333	31.76666667	3
msil_20jun08_Contig14_revised_geneMsil3181	63138	12	25.085	2
msil_20jun08_Contig14_revised_geneMsil3191	11219	3.5	42.855	2
msil_20jun08_Contig14_revised_geneMsil3199	14562	3.5	25	2
msil_20jun08_Contig14_revised_geneMsil3206	16043	7	45.03333333	3
msil_20jun08_Contig14_revised_geneMsil3211	14445	3.33333333	41.53333333	3
msil_20jun08_Contig14_revised_geneMsil3216	20999	7.5	62.565	2
msil_20jun08_Contig14_revised_geneMsil3220	55460	13.5	33.17	2
msil_20jun08_Contig14_revised_geneMsil3226	32689	8.66666667	32.55	3
msil_20jun08_Contig14_revised_geneMsil3234	65213	18.33333333	33.82	3
msil_20jun08_Contig14_revised_geneMsil3255	46063	9.5	26.34	2
msil_20jun08_Contig14_revised_geneMsil3278	34581	10	32.11	3
msil_20jun08_Contig14_revised_geneMsil3309	28875	7.5	47.205	2
msil_20jun08_Contig14_revised_geneMsil3530	45259	20.33333333	44.88	3
msil_20jun08_Contig14_revised_geneMsil3563	66498	17.5	28.595	2
msil_20jun08_Contig14_revised_geneMsil3584	30055	6	21.35	2

msil_20jun08_Contig14_revised_geneMsil3636	10342	3.333333333	51.70333333	3
msil_20jun08_Contig14_revised_geneMsil3698	39195	20.33333333	60.80666667	3
msil_20jun08_Contig14_revised_geneMsil3705	40379	12.66666667	33.15333333	3
msil_20jun08_Contig14_revised_geneMsil3712	16567	5	27.465	2
msil_20jun08_Contig14_revised_geneMsil3746	7075	4	73.81	2
msil_20jun08_Contig14_revised_geneMsil3763	59987	15.5	29.51	2
msil_20jun08_Contig14_revised_geneMsil3776	25471	6	39.27	2
msil_20jun08_Contig14_revised_geneMsil3828	20405	8.5	45.72	2
msil_20jun08_Contig14_revised_geneMsil3854	12064	5.666666667	54.76333333	3
msil_20jun08_Contig14_revised_geneMsil3855	48384	5.333333333	10.23	3
msil_20jun08_Contig14_revised_geneMsil3867	12947	4	34.925	2
msil_20jun08_Contig14_revised_geneMsil3868	153709	30	22.345	2
msil_20jun08_Contig14_revised_geneMsil3869	154444	27	23.98	2
msil_20jun08_Contig14_revised_geneMsil3874	51719	16	31.00666667	3
msil_20jun08_Contig14_revised_geneMsil3875	170267	49.33333333	37.14333333	3
msil_20jun08_Contig14_revised_geneMsil3881	53361	20.33333333	58.25	3

Supplementary Table 5

Protein identifications – label-free quantification – Acetate growth condition

Description	Molecular weight (Da)	Average number of peptides	Average sequence coverage (%)	Technical replicates
msil_20jun08_Contig14_revised_geneMsil0011	13843	9.333333333	60.35333333	3
msil_20jun08_Contig14_revised_geneMsil0061	25546	12.33333333	69.61333333	3
msil_20jun08_Contig14_revised_geneMsil0078	17055	6	38.985	2
msil_20jun08_Contig14_revised_geneMsil0081	35103	11.5	56.075	2
msil_20jun08_Contig14_revised_geneMsil0082	39534	18.66666667	60.23	3
msil_20jun08_Contig14_revised_geneMsil0096	21839	11.33333333	50.68	3
msil_20jun08_Contig14_revised_geneMsil0105	23664	5.666666667	33.65666667	3
msil_20jun08_Contig14_revised_geneMsil0106	54270	19.66666667	44.40333333	3
msil_20jun08_Contig14_revised_geneMsil0107	35134	15.33333333	60.36	3
msil_20jun08_Contig14_revised_geneMsil0131	49760	14.5	42.275	2
msil_20jun08_Contig14_revised_geneMsil0135	50049	10.5	30.18	2
msil_20jun08_Contig14_revised_geneMsil0136	18091	6.666666667	53.09333333	3
msil_20jun08_Contig14_revised_geneMsil0147	45138	8.333333333	28.94	3
msil_20jun08_Contig14_revised_geneMsil0154	14853	7	52.115	2
msil_20jun08_Contig14_revised_geneMsil0162	58645	19.33333333	45.41333333	3
msil_20jun08_Contig14_revised_geneMsil0178	34892	15	58.91333333	3
msil_20jun08_Contig14_revised_geneMsil0181	9466	3	55.55666667	3
msil_20jun08_Contig14_revised_geneMsil0188	53255	16	33.27	3
msil_20jun08_Contig14_revised_geneMsil0189	26072	9.5	44.59	2
msil_20jun08_Contig14_revised_geneMsil0193	23482	8.5	56.755	2

msil_20jun08_Contig14_revised_geneMsil0195	19814	10	56.65	3
msil_20jun08_Contig14_revised_geneMsil0202	37231	12.66666667	52.27666667	3
msil_20jun08_Contig14_revised_geneMsil0212	58102	15.33333333	38.13333333	3
msil_20jun08_Contig14_revised_geneMsil0239	32226	9	36.41333333	3
msil_20jun08_Contig14_revised_geneMsil0255	26326	8.66666667	42.42	3
msil_20jun08_Contig14_revised_geneMsil0268	77286	30.33333333	49.20666667	3
msil_20jun08_Contig14_revised_geneMsil0286	29882	10	50.60666667	3
msil_20jun08_Contig14_revised_geneMsil0289	18026	10.33333333	60.69666667	3
msil_20jun08_Contig14_revised_geneMsil0292	34508	9.5	32.52	2
msil_20jun08_Contig14_revised_geneMsil0314	17592	7.66666667	61.69666667	3
msil_20jun08_Contig14_revised_geneMsil0318	37265	13	39.155	2
msil_20jun08_Contig14_revised_geneMsil0325	10539	4	44.795	2
msil_20jun08_Contig14_revised_geneMsil0328	40649	11.5	51.555	2
msil_20jun08_Contig14_revised_geneMsil0340	20096	9	62.76666667	3
msil_20jun08_Contig14_revised_geneMsil0341	54742	25.33333333	42.24	3
msil_20jun08_Contig14_revised_geneMsil0342	31630	14.66666667	58.19333333	3
msil_20jun08_Contig14_revised_geneMsil0343	51344	25.66666667	67.63333333	3
msil_20jun08_Contig14_revised_geneMsil0344	14406	5	56.13666667	3
msil_20jun08_Contig14_revised_geneMsil0351	25878	8.66666667	51.73	3
msil_20jun08_Contig14_revised_geneMsil0352	15502	6.33333333	48.02	3
msil_20jun08_Contig14_revised_geneMsil0367	49316	11.5	32.065	2
msil_20jun08_Contig14_revised_geneMsil0380	28036	12.33333333	49.30666667	3
msil_20jun08_Contig14_revised_geneMsil0381	36470	15.66666667	42.71666667	3
msil_20jun08_Contig14_revised_geneMsil0420	39448	12	44.16666667	3
msil_20jun08_Contig14_revised_geneMsil0426	38837	13.33333333	40.44333333	3
msil_20jun08_Contig14_revised_geneMsil0427	52067	19.66666667	48.61666667	3

msil_20jun08_Contig14_revised_geneMsil0428	12019	4	49.10666667	3
msil_20jun08_Contig14_revised_geneMsil0450	20681	4	29.23333333	3
msil_20jun08_Contig14_revised_geneMsil0471	68529	42	71.68	3
msil_20jun08_Contig14_revised_geneMsil0472	31060	17.33333333	58.59666667	3
msil_20jun08_Contig14_revised_geneMsil0485	22673	6.66666667	35.58	3
msil_20jun08_Contig14_revised_geneMsil0493	12257	5	53.50666667	3
msil_20jun08_Contig14_revised_geneMsil0495	65056	21.33333333	41.69333333	3
msil_20jun08_Contig14_revised_geneMsil0499	35950	11.5	39.65	2
msil_20jun08_Contig14_revised_geneMsil0515	44953	17.66666667	45.97666667	3
msil_20jun08_Contig14_revised_geneMsil0519	38148	8.66666667	34.30333333	3
msil_20jun08_Contig14_revised_geneMsil0520	48780	17.33333333	49.78333333	3
msil_20jun08_Contig14_revised_geneMsil0521	45795	23	44.29666667	3
msil_20jun08_Contig14_revised_geneMsil0522	50947	19.33333333	50.76333333	3
msil_20jun08_Contig14_revised_geneMsil0534	30949	7	30.42	2
msil_20jun08_Contig14_revised_geneMsil0559	20538	13	67.70333333	3
msil_20jun08_Contig14_revised_geneMsil0561	17054	4.5	34.455	2
msil_20jun08_Contig14_revised_geneMsil0563	20320	5.66666667	28.22	3
msil_20jun08_Contig14_revised_geneMsil0573	15501	5	30.17333333	3
msil_20jun08_Contig14_revised_geneMsil0574	26639	6.66666667	33.33333333	3
msil_20jun08_Contig14_revised_geneMsil0576	10354	4.33333333	54.34666667	3
msil_20jun08_Contig14_revised_geneMsil0579	22138	7.33333333	46.44	3
msil_20jun08_Contig14_revised_geneMsil0582	43134	24.33333333	74.15666667	3
msil_20jun08_Contig14_revised_geneMsil0583	76289	35.33333333	55.33	3
msil_20jun08_Contig14_revised_geneMsil0584	17850	6.66666667	42.09333333	3
msil_20jun08_Contig14_revised_geneMsil0586	51588	17	32.75	2
msil_20jun08_Contig14_revised_geneMsil0612	34378	12.66666667	47.51666667	3

msil_20jun08_Contig14_revised_geneMsil0619	19849	9	64.81666667	3
msil_20jun08_Contig14_revised_geneMsil0620	28630	18.33333333	89.19666667	3
msil_20jun08_Contig14_revised_geneMsil0643	47785	21.33333333	55.48333333	3
msil_20jun08_Contig14_revised_geneMsil0651	12125	5	33.48	2
msil_20jun08_Contig14_revised_geneMsil0660	17046	5.5	49.695	2
msil_20jun08_Contig14_revised_geneMsil0672	118596	23.66666667	32.16	3
msil_20jun08_Contig14_revised_geneMsil0673	16443	7.33333333	43.64	3
msil_20jun08_Contig14_revised_geneMsil0675	64133	25	48.51	3
msil_20jun08_Contig14_revised_geneMsil0701	49297	10.33333333	28.25666667	3
msil_20jun08_Contig14_revised_geneMsil0737	53465	21.33333333	53.74	3
msil_20jun08_Contig14_revised_geneMsil0738	52264	20	53.21333333	3
msil_20jun08_Contig14_revised_geneMsil0747	91754	27	32.80666667	3
msil_20jun08_Contig14_revised_geneMsil0749	12177	6.5	61.945	2
msil_20jun08_Contig14_revised_geneMsil0758	8371	4	59.72	2
msil_20jun08_Contig14_revised_geneMsil0764	31189	11.66666667	53.58666667	3
msil_20jun08_Contig14_revised_geneMsil0774	33332	5	22.57	3
msil_20jun08_Contig14_revised_geneMsil0780	70259	18.66666667	30.69	3
msil_20jun08_Contig14_revised_geneMsil0786	35857	18.5	67.12	2
msil_20jun08_Contig14_revised_geneMsil0794	10224	4.66666667	47.37	3
msil_20jun08_Contig14_revised_geneMsil0795	57608	41.33333333	70.32333333	3
msil_20jun08_Contig14_revised_geneMsil0805	53987	17	40.715	2
msil_20jun08_Contig14_revised_geneMsil0806	26970	7	38.22	3
msil_20jun08_Contig14_revised_geneMsil0811	27406	11	55.555	2
msil_20jun08_Contig14_revised_geneMsil0823	14700	11.66666667	55.31333333	3
msil_20jun08_Contig14_revised_geneMsil0832	41778	14.5	40	2
msil_20jun08_Contig14_revised_geneMsil0839	36345	10	33.145	2

msil_20jun08_Contig14_revised_geneMsil0848	31544	11.66666667	47.87666667	3
msil_20jun08_Contig14_revised_geneMsil0852	13555	6.5	64.615	2
msil_20jun08_Contig14_revised_geneMsil0870	19884	9	64.03333333	3
msil_20jun08_Contig14_revised_geneMsil0887	9398	5	57.36666667	3
msil_20jun08_Contig14_revised_geneMsil0895	25268	9.333333333	58.50333333	3
msil_20jun08_Contig14_revised_geneMsil0898	16594	5.666666667	37.44333333	3
msil_20jun08_Contig14_revised_geneMsil0899	9233	2.333333333	24.16666667	3
msil_20jun08_Contig14_revised_geneMsil0903	36627	11	32.57	2
msil_20jun08_Contig14_revised_geneMsil0931	51409	26.33333333	61.32333333	3
msil_20jun08_Contig14_revised_geneMsil0968	15036	6.333333333	46.90333333	3
msil_20jun08_Contig14_revised_geneMsil0970	18853	10.66666667	36.89333333	3
msil_20jun08_Contig14_revised_geneMsil0971	20604	10.33333333	55.79666667	3
msil_20jun08_Contig14_revised_geneMsil0999	14246	5.5	53.875	2
msil_20jun08_Contig14_revised_geneMsil1003	30140	13.66666667	62.17666667	3
msil_20jun08_Contig14_revised_geneMsil1006	67355	12.5	28.705	2
msil_20jun08_Contig14_revised_geneMsil1007	46510	16	50.465	2
msil_20jun08_Contig14_revised_geneMsil1012	31539	11.33333333	51.54666667	3
msil_20jun08_Contig14_revised_geneMsil1025	20017	5.666666667	51.39666667	3
msil_20jun08_Contig14_revised_geneMsil1043	19114	6.333333333	48.31	3
msil_20jun08_Contig14_revised_geneMsil1068	30027	9.333333333	44.56333333	3
msil_20jun08_Contig14_revised_geneMsil1086	33521	13.33333333	45.26666667	3
msil_20jun08_Contig14_revised_geneMsil1117	63869	23.66666667	48.20333333	3
msil_20jun08_Contig14_revised_geneMsil1125	54423	19.5	41.735	2
msil_20jun08_Contig14_revised_geneMsil1133	27825	10	32.76	2
msil_20jun08_Contig14_revised_geneMsil1138	27183	7.5	41.94	2
msil_20jun08_Contig14_revised_geneMsil1140	26523	9.5	45.04	2

msil_20jun08_Contig14_revised_geneMsil1149	28664	7.666666667	33.99333333	3
msil_20jun08_Contig14_revised_geneMsil1154	52933	21.33333333	54.76	3
msil_20jun08_Contig14_revised_geneMsil1158	58989	22	50.92	3
msil_20jun08_Contig14_revised_geneMsil1160	31761	11	40.49333333	3
msil_20jun08_Contig14_revised_geneMsil1165	35463	10.33333333	36.91666667	3
msil_20jun08_Contig14_revised_geneMsil1171	17691	7.333333333	54.85	3
msil_20jun08_Contig14_revised_geneMsil1186	12265	9.333333333	69.94	3
msil_20jun08_Contig14_revised_geneMsil1191	37974	7.666666667	26.70333333	3
msil_20jun08_Contig14_revised_geneMsil1193	70227	22.33333333	48.87666667	3
msil_20jun08_Contig14_revised_geneMsil1194	39096	14.66666667	61.49333333	3
msil_20jun08_Contig14_revised_geneMsil1203	30940	8.666666667	43.65	3
msil_20jun08_Contig14_revised_geneMsil1205	30006	11.66666667	32.87	3
msil_20jun08_Contig14_revised_geneMsil1215	47989	10.33333333	23.91333333	3
msil_20jun08_Contig14_revised_geneMsil1224	10019	5.333333333	61.04666667	3
msil_20jun08_Contig14_revised_geneMsil1226	70912	30.66666667	50.28333333	3
msil_20jun08_Contig14_revised_geneMsil1236	19181	3.333333333	28.28666667	3
msil_20jun08_Contig14_revised_geneMsil1259	34728	7.5	32.09	2
msil_20jun08_Contig14_revised_geneMsil1262	59789	18	37.58	3
msil_20jun08_Contig14_revised_geneMsil1263	44933	16.5	55.245	2
msil_20jun08_Contig14_revised_geneMsil1264	15306	7.333333333	51.34	3
msil_20jun08_Contig14_revised_geneMsil1314	12494	3.666666667	53.58333333	3
msil_20jun08_Contig14_revised_geneMsil1325	79102	33.33333333	56.12666667	3
msil_20jun08_Contig14_revised_geneMsil1328	52734	21	46.03666667	3
msil_20jun08_Contig14_revised_geneMsil1347	36232	8	34.89333333	3
msil_20jun08_Contig14_revised_geneMsil1353	44433	10	32.76	3
msil_20jun08_Contig14_revised_geneMsil1365	8505	3.666666667	51.38666667	3

msil_20jun08_Contig14_revised_geneMsil1368	44591	11	37.745	2
msil_20jun08_Contig14_revised_geneMsil1375	58235	28	61.19333333	3
msil_20jun08_Contig14_revised_geneMsil1381	42704	11	31.46	2
msil_20jun08_Contig14_revised_geneMsil1382	46216	19.33333333	56.97	3
msil_20jun08_Contig14_revised_geneMsil1388	22303	9	28.47	2
msil_20jun08_Contig14_revised_geneMsil1395	75669	29.33333333	53.1	3
msil_20jun08_Contig14_revised_geneMsil1411	11268	7.666666667	61.37333333	3
msil_20jun08_Contig14_revised_geneMsil1429	45127	12.66666667	33.49	3
msil_20jun08_Contig14_revised_geneMsil1493	31367	7	46.27333333	3
msil_20jun08_Contig14_revised_geneMsil1559	42669	16.33333333	62.98	3
msil_20jun08_Contig14_revised_geneMsil1574	28579	10.66666667	35.92333333	3
msil_20jun08_Contig14_revised_geneMsil1575	29163	10	34.77666667	3
msil_20jun08_Contig14_revised_geneMsil1587	65135	16	40.1	3
msil_20jun08_Contig14_revised_geneMsil1603	24476	7.333333333	31.87666667	3
msil_20jun08_Contig14_revised_geneMsil1611	20289	6.333333333	29.63	3
msil_20jun08_Contig14_revised_geneMsil1616	45805	17	39.5	2
msil_20jun08_Contig14_revised_geneMsil1624	42542	7.333333333	28.46333333	3
msil_20jun08_Contig14_revised_geneMsil1627	76934	17.33333333	30.16	3
msil_20jun08_Contig14_revised_geneMsil1628	8288	2	16.88666667	3
msil_20jun08_Contig14_revised_geneMsil1630	39298	13.33333333	47.85333333	3
msil_20jun08_Contig14_revised_geneMsil1633	23930	6	34.36333333	3
msil_20jun08_Contig14_revised_geneMsil1649	40230	17	41.46	2
msil_20jun08_Contig14_revised_geneMsil1676	22241	16	77.55666667	3
msil_20jun08_Contig14_revised_geneMsil1677	22932	10	44.97666667	3
msil_20jun08_Contig14_revised_geneMsil1680	51617	16	40.59	3
msil_20jun08_Contig14_revised_geneMsil1681	47722	21	57.57666667	3

msil_20jun08_Contig14_revised_geneMsil1688	100172	18.66666667	25.49666667	3
msil_20jun08_Contig14_revised_geneMsil1691	24368	8.5	58.695	2
msil_20jun08_Contig14_revised_geneMsil1693	32000	15.33333333	57.63666667	3
msil_20jun08_Contig14_revised_geneMsil1694	37702	16.33333333	54.15333333	3
msil_20jun08_Contig14_revised_geneMsil1706	49863	23.33333333	60.98	3
msil_20jun08_Contig14_revised_geneMsil1707	42671	12.5	40.76	2
msil_20jun08_Contig14_revised_geneMsil1713	33845	13.33333333	56.66	3
msil_20jun08_Contig14_revised_geneMsil1714	42610	19.66666667	59.59666667	3
msil_20jun08_Contig14_revised_geneMsil1716	42908	16.66666667	50.88	3
msil_20jun08_Contig14_revised_geneMsil1717	31773	10.5	39.775	2
msil_20jun08_Contig14_revised_geneMsil1719	35013	14	56.615	2
msil_20jun08_Contig14_revised_geneMsil1743	38477	13	43.56333333	3
msil_20jun08_Contig14_revised_geneMsil1744	13791	4.5	52.615	2
msil_20jun08_Contig14_revised_geneMsil1745	16470	6.66666667	44.16	3
msil_20jun08_Contig14_revised_geneMsil1759	41928	16.33333333	50.94	3
msil_20jun08_Contig14_revised_geneMsil1760	55723	20.33333333	50.72	3
msil_20jun08_Contig14_revised_geneMsil1777	14945	5	25	2
msil_20jun08_Contig14_revised_geneMsil1807	23932	7.66666667	35.37666667	3
msil_20jun08_Contig14_revised_geneMsil1808	30842	16.33333333	52.43	3
msil_20jun08_Contig14_revised_geneMsil1816	8581	2	44.935	2
msil_20jun08_Contig14_revised_geneMsil1821	35733	11	38.30333333	3
msil_20jun08_Contig14_revised_geneMsil1827	36279	5.5	17.68	2
msil_20jun08_Contig14_revised_geneMsil1843	14867	4	53.4	2
msil_20jun08_Contig14_revised_geneMsil1858	26586	8.33333333	36.18666667	3
msil_20jun08_Contig14_revised_geneMsil1860	52564	4	14.95333333	3
msil_20jun08_Contig14_revised_geneMsil1861	12022	8.33333333	61.01333333	3

msil_20jun08_Contig14_revised_geneMsil1865	32182	12	52.83	3
msil_20jun08_Contig14_revised_geneMsil1869	31672	9	34.375	2
msil_20jun08_Contig14_revised_geneMsil1872	17457	7.333333333	36.89666667	3
msil_20jun08_Contig14_revised_geneMsil1881	32802	11.33333333	50.84	3
msil_20jun08_Contig14_revised_geneMsil1898	27222	8.666666667	47.78666667	3
msil_20jun08_Contig14_revised_geneMsil1900	25195	9	37.445	2
msil_20jun08_Contig14_revised_geneMsil1905	89867	26.5	38.98	2
msil_20jun08_Contig14_revised_geneMsil1907	95512	29.33333333	37.34666667	3
msil_20jun08_Contig14_revised_geneMsil1910	71137	34	58.35666667	3
msil_20jun08_Contig14_revised_geneMsil2074	45579	11	35.86	2
msil_20jun08_Contig14_revised_geneMsil2079	37620	8.333333333	29.16333333	3
msil_20jun08_Contig14_revised_geneMsil2091	35124	17.66666667	64.66	3
msil_20jun08_Contig14_revised_geneMsil2110	46025	22	52.30666667	3
msil_20jun08_Contig14_revised_geneMsil2111	33360	9	34.21666667	3
msil_20jun08_Contig14_revised_geneMsil2114	27914	11	48.66666667	3
msil_20jun08_Contig14_revised_geneMsil2141	17478	6.333333333	52.29666667	3
msil_20jun08_Contig14_revised_geneMsil2206	19288	8	64	2
msil_20jun08_Contig14_revised_geneMsil2212	11167	6.5	76.53	2
msil_20jun08_Contig14_revised_geneMsil2221	55725	17.5	41.005	2
msil_20jun08_Contig14_revised_geneMsil2224	34262	13	46.87666667	3
msil_20jun08_Contig14_revised_geneMsil2232	101369	26	27.83	2
msil_20jun08_Contig14_revised_geneMsil2245	21640	11.5	65.735	2
msil_20jun08_Contig14_revised_geneMsil2246	37432	8.666666667	35.14666667	3
msil_20jun08_Contig14_revised_geneMsil2253	20419	5	43.205	2
msil_20jun08_Contig14_revised_geneMsil2282	35524	15.66666667	57.61	3
msil_20jun08_Contig14_revised_geneMsil2283	43212	15	55.69333333	3

msil_20jun08_Contig14_revised_geneMsil2289	18866	4.666666667	37.75333333	3
msil_20jun08_Contig14_revised_geneMsil2295	95767	26.66666667	36.38333333	3
msil_20jun08_Contig14_revised_geneMsil2297	59527	15.66666667	34.83666667	3
msil_20jun08_Contig14_revised_geneMsil2301	55900	11.33333333	30.28333333	3
msil_20jun08_Contig14_revised_geneMsil2329	47962	19	41.08	3
msil_20jun08_Contig14_revised_geneMsil2342	55659	26.33333333	65.01666667	3
msil_20jun08_Contig14_revised_geneMsil2344	28678	6	31.91333333	3
msil_20jun08_Contig14_revised_geneMsil2345	12700	13.66666667	86.16	3
msil_20jun08_Contig14_revised_geneMsil2360	19922	8.666666667	48.99666667	3
msil_20jun08_Contig14_revised_geneMsil2361	56550	23.66666667	56.08333333	3
msil_20jun08_Contig14_revised_geneMsil2385	30967	11.66666667	65.65	3
msil_20jun08_Contig14_revised_geneMsil2387	33960	12.66666667	58.38666667	3
msil_20jun08_Contig14_revised_geneMsil2390	18570	9	58.66666667	3
msil_20jun08_Contig14_revised_geneMsil2400	39210	17	58.80666667	3
msil_20jun08_Contig14_revised_geneMsil2403	27359	8.666666667	38.68333333	3
msil_20jun08_Contig14_revised_geneMsil2427	66206	19.5	44.985	2
msil_20jun08_Contig14_revised_geneMsil2428	61868	12	32.875	2
msil_20jun08_Contig14_revised_geneMsil2436	16019	8.333333333	71.33333333	3
msil_20jun08_Contig14_revised_geneMsil2442	13852	5.666666667	52.1	3
msil_20jun08_Contig14_revised_geneMsil2447	36024	12.33333333	53.21666667	3
msil_20jun08_Contig14_revised_geneMsil2485	36213	11.66666667	48.84333333	3
msil_20jun08_Contig14_revised_geneMsil2501	33463	21.66666667	80.58333333	3
msil_20jun08_Contig14_revised_geneMsil2503	30071	14.66666667	57.59666667	3
msil_20jun08_Contig14_revised_geneMsil2504	113111	36.66666667	43.68	3
msil_20jun08_Contig14_revised_geneMsil2505	44837	19.66666667	45.87	3
msil_20jun08_Contig14_revised_geneMsil2523	34049	19	56.73	3

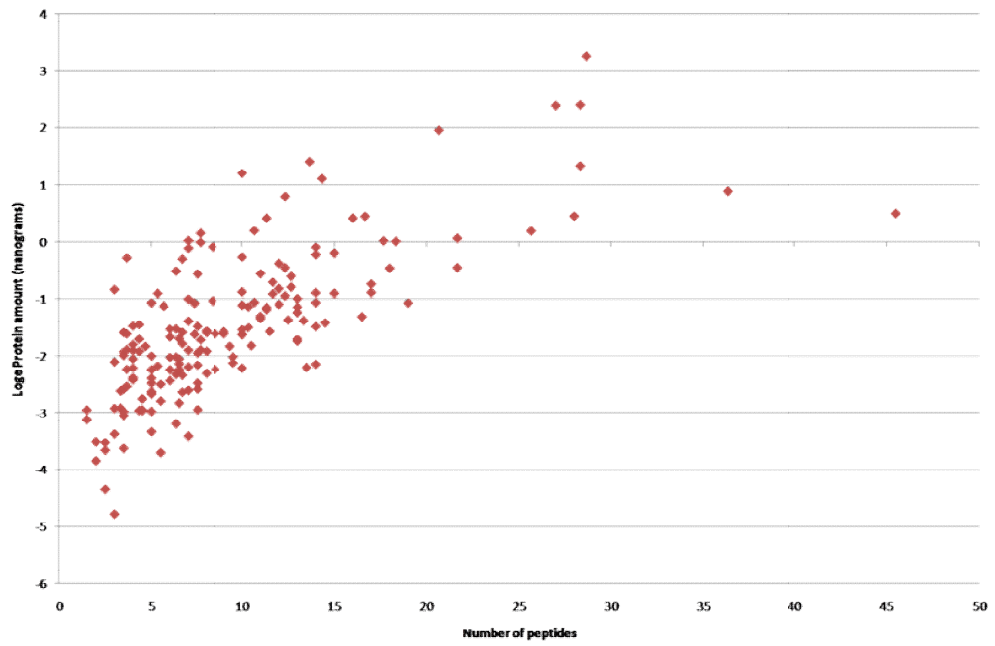
msil_20jun08_Contig14_revised_geneMsil2525	105844	27	32.295	2
msil_20jun08_Contig14_revised_geneMsil2530	23299	7	45.375	2
msil_20jun08_Contig14_revised_geneMsil2532	47274	14	31.28	2
msil_20jun08_Contig14_revised_geneMsil2544	16674	4.333333333	54.14	3
msil_20jun08_Contig14_revised_geneMsil2576	69080	20	35.165	2
msil_20jun08_Contig14_revised_geneMsil2598	31231	16	63.27666667	3
msil_20jun08_Contig14_revised_geneMsil2608	39954	18.66666667	63.61	3
msil_20jun08_Contig14_revised_geneMsil2645	15175	7	51.51333333	3
msil_20jun08_Contig14_revised_geneMsil2647	24390	13	61.36333333	3
msil_20jun08_Contig14_revised_geneMsil2653	36393	12	48.25	2
msil_20jun08_Contig14_revised_geneMsil2655	79369	40.33333333	65.39333333	3
msil_20jun08_Contig14_revised_geneMsil2678	44881	5.5	21.535	2
msil_20jun08_Contig14_revised_geneMsil2707	17807	7.5	49.68	2
msil_20jun08_Contig14_revised_geneMsil2796	13182	5.666666667	47.78	3
msil_20jun08_Contig14_revised_geneMsil2800	20504	4	34.22333333	3
msil_20jun08_Contig14_revised_geneMsil2810	54003	20.33333333	46.16	3
msil_20jun08_Contig14_revised_geneMsil2811	12104	4	26.89	3
msil_20jun08_Contig14_revised_geneMsil2815	8480	4.333333333	45.52666667	3
msil_20jun08_Contig14_revised_geneMsil2816	61931	14.66666667	35.64	3
msil_20jun08_Contig14_revised_geneMsil2864	47166	11.5	32.99	2
msil_20jun08_Contig14_revised_geneMsil2894	38212	11.66666667	45.28666667	3
msil_20jun08_Contig14_revised_geneMsil2903	15538	5	42.28	3
msil_20jun08_Contig14_revised_geneMsil2912	50645	22.66666667	53.75	3
msil_20jun08_Contig14_revised_geneMsil2913	24076	10.33333333	59.22666667	3
msil_20jun08_Contig14_revised_geneMsil2914	46270	15.66666667	42.83666667	3
msil_20jun08_Contig14_revised_geneMsil2915	89017	26.66666667	36.23333333	3

msil_20jun08_Contig14_revised_geneMsil2952	7316	6	87.62	3
msil_20jun08_Contig14_revised_geneMsil2955	67811	29	52.52333333	3
msil_20jun08_Contig14_revised_geneMsil2969	25250	10	49.21666667	3
msil_20jun08_Contig14_revised_geneMsil2977	49364	15	44.25333333	3
msil_20jun08_Contig14_revised_geneMsil2991	37134	15	54.24333333	3
msil_20jun08_Contig14_revised_geneMsil2996	40209	30.33333333	79.74333333	3
msil_20jun08_Contig14_revised_geneMsil2997	25488	14.33333333	61.27333333	3
msil_20jun08_Contig14_revised_geneMsil3000	54595	14.5	35.645	2
msil_20jun08_Contig14_revised_geneMsil3002	35388	8	46.96333333	3
msil_20jun08_Contig14_revised_geneMsil3011	36643	11.33333333	43.85333333	3
msil_20jun08_Contig14_revised_geneMsil3070	45240	11	43.42	2
msil_20jun08_Contig14_revised_geneMsil3149	68915	11	21.395	2
msil_20jun08_Contig14_revised_geneMsil3157	59816	23.33333333	54.73333333	3
msil_20jun08_Contig14_revised_geneMsil3159	27522	6	37.255	2
msil_20jun08_Contig14_revised_geneMsil3172	88382	24	38.39333333	3
msil_20jun08_Contig14_revised_geneMsil3206	16043	7	60.70666667	3
msil_20jun08_Contig14_revised_geneMsil3210	37801	13.66666667	56.62666667	3
msil_20jun08_Contig14_revised_geneMsil3211	14445	5.33333333	55.02666667	3
msil_20jun08_Contig14_revised_geneMsil3216	20999	8.66666667	71.02333333	3
msil_20jun08_Contig14_revised_geneMsil3218	11717	7	59.05	2
msil_20jun08_Contig14_revised_geneMsil3226	32689	7	29.46	2
msil_20jun08_Contig14_revised_geneMsil3238	45932	13	44.85666667	3
msil_20jun08_Contig14_revised_geneMsil3248	122929	28.33333333	23.29666667	3
msil_20jun08_Contig14_revised_geneMsil3255	46063	17.66666667	45.46	3
msil_20jun08_Contig14_revised_geneMsil3270	18324	5.5	33.625	2
msil_20jun08_Contig14_revised_geneMsil3285	25563	11	43.97333333	3

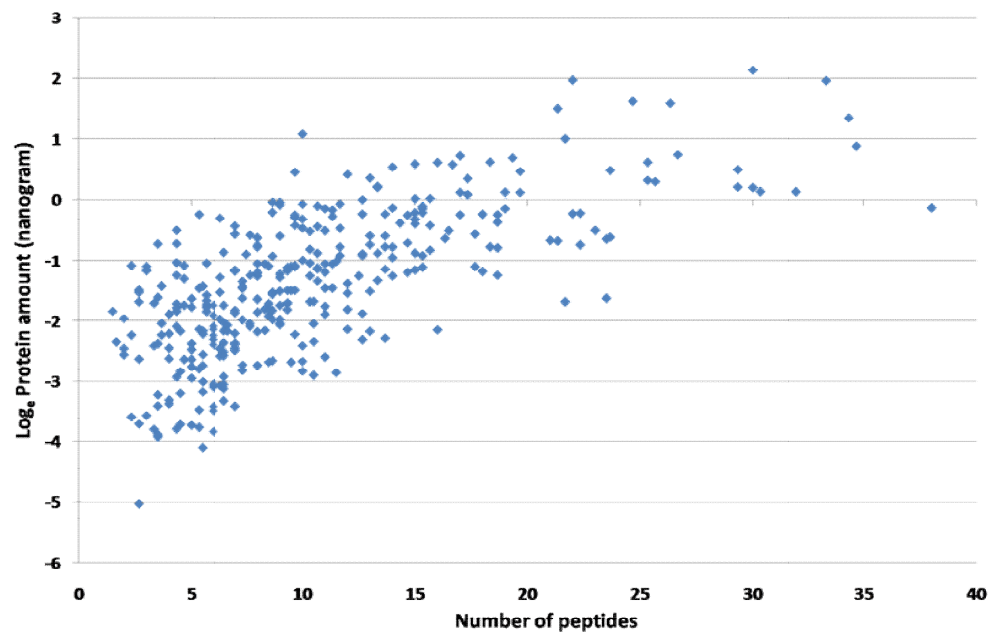
msil_20jun08_Contig14_revised_geneMsil3287	76131	28.66666667	49.83333333	3
msil_20jun08_Contig14_revised_geneMsil3292	26279	8	29.00666667	3
msil_20jun08_Contig14_revised_geneMsil3306	39546	13.5	37.37	2
msil_20jun08_Contig14_revised_geneMsil3309	28875	14	72.385	2
msil_20jun08_Contig14_revised_geneMsil3314	33266	7	33.43666667	3
msil_20jun08_Contig14_revised_geneMsil3336	36160	14	51.53333333	3
msil_20jun08_Contig14_revised_geneMsil3475	6308	3.333333333	82.46	3
msil_20jun08_Contig14_revised_geneMsil3483	48455	15.66666667	45.04666667	3
msil_20jun08_Contig14_revised_geneMsil3502	13955	5	52.97333333	3
msil_20jun08_Contig14_revised_geneMsil3509	98993	31.66666667	44.68666667	3
msil_20jun08_Contig14_revised_geneMsil3510	40214	19	62.25333333	3
msil_20jun08_Contig14_revised_geneMsil3530	45259	27	58.39333333	3
msil_20jun08_Contig14_revised_geneMsil3538	73080	24	33.18	2
msil_20jun08_Contig14_revised_geneMsil3544	33697	11	46.08	3
msil_20jun08_Contig14_revised_geneMsil3545	136112	41.33333333	39.09	3
msil_20jun08_Contig14_revised_geneMsil3550	37460	12	41.885	2
msil_20jun08_Contig14_revised_geneMsil3552	19790	12.33333333	70.07	3
msil_20jun08_Contig14_revised_geneMsil3553	16543	7.333333333	58.05666667	3
msil_20jun08_Contig14_revised_geneMsil3563	66498	20.66666667	41.67	3
msil_20jun08_Contig14_revised_geneMsil3660	28431	4.5	25.885	2
msil_20jun08_Contig14_revised_geneMsil3674	28144	7.5	41.175	2
msil_20jun08_Contig14_revised_geneMsil3676	17015	6.333333333	76.75	3
msil_20jun08_Contig14_revised_geneMsil3687	76633	23.66666667	42.47666667	3
msil_20jun08_Contig14_revised_geneMsil3696	35129	12.5	33.89	2
msil_20jun08_Contig14_revised_geneMsil3697	31854	14	56.38333333	3
msil_20jun08_Contig14_revised_geneMsil3698	39195	21	72.16	3

msil_20jun08_Contig14_revised_geneMsil3702	53285	17	43.39	3
msil_20jun08_Contig14_revised_geneMsil3705	40379	13.33333333	48.11666667	3
msil_20jun08_Contig14_revised_geneMsil3746	7075	3.666666667	58.2	3
msil_20jun08_Contig14_revised_geneMsil3780	31701	10	45.65666667	3
msil_20jun08_Contig14_revised_geneMsil3781	26419	7	38.55666667	3
msil_20jun08_Contig14_revised_geneMsil3816	45238	7.5	27.845	2
msil_20jun08_Contig14_revised_geneMsil3820	19350	6.333333333	39.16	3
msil_20jun08_Contig14_revised_geneMsil3821	7381	2.5	56	2
msil_20jun08_Contig14_revised_geneMsil3828	20405	10.66666667	59.71333333	3
msil_20jun08_Contig14_revised_geneMsil3831	26234	10	42.26	2
msil_20jun08_Contig14_revised_geneMsil3832	20774	5.5	45.43	2
msil_20jun08_Contig14_revised_geneMsil3845	13993	6	58.27	2
msil_20jun08_Contig14_revised_geneMsil3851	15220	4.666666667	44.05333333	3
msil_20jun08_Contig14_revised_geneMsil3863	19697	7.666666667	48.48666667	3
msil_20jun08_Contig14_revised_geneMsil3867	12947	3	26.45666667	3
msil_20jun08_Contig14_revised_geneMsil3868	153709	49.66666667	43.41	3
msil_20jun08_Contig14_revised_geneMsil3869	154444	51	46.36333333	3
msil_20jun08_Contig14_revised_geneMsil3874	51719	17	30.22333333	3
msil_20jun08_Contig14_revised_geneMsil3881	53361	26	73.63666667	3
msil_20jun08_Contig14_revised_geneMsil3907	14659	3.333333333	23.15	3
msil_20jun08_Contig14_revised_geneMsil3913	18654	5.333333333	59.15666667	3

(a)



(b)



Supplementary figure 1 – Correlation of protein concentration and number of peptides identified in M. silvestris when grown under (a) methane, (b) acetate.

Appendix B

Dissemination of work

Journal papers

Patel, V. J., Thalassinou, K., Slade, S. E., Connolly, J. B., Crombie, A., Murrell, J. C. and Scrivens, J. H. (2009) A comparative, qualitative and quantitative proteomic study of the methanotrophic bacterium *Methylocella silvestris*. *Journal of Proteome Research* **8**(7): 3752-3759

Scarff, C. A., **Patel, V. J.**, Thalassinou, K., Scrivens, J. H. (2009) Probing hemoglobin structure by means of travelling wave ion mobility mass spectrometry. *Journal of the American Society for Mass Spectrometry* **20**: 625-631.

Williams, J. P., **Patel, V. J.**, Holland, R., Scrivens, J. H. (2006) The use of recently described ionisation techniques for the rapid analysis of some common drugs and samples of biological origin. *Rapid Communications in Mass Spectrometry* **20**: 1447-1456

Williams, J. P., Nibbering, N. M. M., Green, B. N., **Patel, V. J.**, Scrivens, J. H. (2006) Collision-induced fragmentation pathways including odd-electron ion formation from desorption electrospray ionisation generated protonated and deprotonated drugs derived from tandem accurate mass spectrometry. *Journal of Mass Spectrometry* **41**: 1277-1286

Williams, J. P., Lock, R., **Patel, V. J.**, Scrivens, J. H. (2006) Polarity switching accurate mass measurement of pharmaceutical samples using desorption electrospray ionization and a dual ion source interfaced to an orthogonal acceleration time-of-flight mass spectrometer. *Analytical Chemistry* **78**: 7440-7445

Conference papers

Patel, V. J., Crombie, A., Thalassinou, K., Connolly, J. B., Murrell, J. C., Slade, S. E. and Scrivens, J. H. Identifying and quantifying novel biochemical pathways in newly discovered prokaryotes. *56th ASMS Conference on Mass Spectrometry & Allied Topics Denver, Colorado* (2008)

Connolly, J. B., Tomczyk, N., Slade, S. E., **Patel, V. J.**, Boden, R., Thalassinou, K., Shaefer, H. and Scrivens, J. H. Strategies for stringent cataloguing of *Methylophaga thiooxidans* using an alternative scanning LCMS approach. *56th ASMS Conference on Mass Spectrometry & Allied Topics Denver, Colorado* (2008)

Patel, V. J., Scarff, C. A., Thalassinou, K., Williams, J. P. and Scrivens, J. H. Structure-property relationships of hemoglobin tetramers studied by travelling wave-based ion mobility-mass spectrometry. *BMSS 29th Annual Meeting Heriot-Watt University, Edinburgh Scotland* (2007)

Patel, V. J., Williams, J. P., Giles, K., Bateman, R. H., Green, B. N., Scrivens, J. H. Studying non-covalent metalloprotein complexes by travelling wave-based ion mobility mass spectrometry, *Proceedings of the 54th ASMS Conference on Mass Spectrometry and Allied Topics, Seattle, Washington and 17th International Mass Spectrometry Conference, Prague, Czech Republic (2006)*

Oral Presentations

Patel, V. J., Thalassinou, K., Slade, S. E., Scrivens, J. H. Experimental approaches to profiling and comparative proteomics, *13th Indian Society for Mass Spectrometry Workshop, Anushakti Nagar, Mumbai (2008)* – Recipient of Best Student Presentation Award

Patel, V. J., Proteomic approaches to the characterisation of a newly discovered prokaryote, *Annual Postgraduate Symposium, Department of Biological Sciences, University of Warwick (2008)* – Recipient of Best Final Year Presentation Award

Invited Oral Presentations

Patel, V. J., Thalassinou, K., Slade, S. E., Crombie, A., Murrell, J. C., Scrivens, J. H. An evaluation of profiling and differential proteomics approaches for the characterisation of biochemical pathways in a newly-discovered prokaryote. *Waters Nordic Users' Meetings, Helsinki, Stockholm, Oslo, Copenhagen and Waters French Users' Meeting, Paris (2008)*

The use of recently described ionisation techniques for the rapid analysis of some common drugs and samples of biological origin

Jonathan P. Williams*, Vibhuti J. Patel, Richard Holland and James H. Scrivens

Department of Biological Sciences, University of Warwick, Gibbet Hill Rd, Coventry CV4 7AL, UK

Received 5 February 2006; Revised 8 March 2006; Accepted 13 March 2006

Three ionisation techniques that require no sample preparation or extraction prior to mass analysis have been used for the rapid analysis of pharmaceutical tablets and ointments. These methods were (i) the novel direct analysis in real time (DART), (ii) desorption electrospray ionisation (DESI), and (iii) desorption atmospheric pressure chemical ionisation (DAPCI). The performance of the three techniques was investigated for a number of common drugs. Significant differences between these approaches were observed. For compounds of moderate to low polarity DAPCI produced more effective ionisation. Accurate DESI and DAPCI tandem mass spectra were obtained and these greatly enhance the selectivity and information content of the experiment. The detection from human skin of the active ingredients from ointments is reported together with the detection of ibuprofen metabolites in human urine. Copyright © 2006 John Wiley & Sons, Ltd.

Desorption electrospray ionisation (DESI) is a newly developed ionisation technique for the analysis and detection of samples present on a variety of surfaces.¹ Rapid high-throughput analysis can be undertaken for a variety of sample types. The technique can be used to detect analytes on surface areas of sub-mm² dimensions. Applications utilising the DESI technique include the detection of rhodamine dyes and pharmaceutical samples separated on thin-layer chromatography (TLC) plates,² the rapid detection of pharmaceutical samples using ion mobility/time-of-flight mass spectrometry (ToFMS),³ high-throughput analysis of pharmaceutical samples,⁴ the detection of explosives,^{5,6} the rapid accurate mass tandem mass spectrometry (MS/MS) of pharmaceutical samples,⁷ intact biological tissue imaging,⁸ direct analysis of alkaloids from plant tissue,⁹ and the analysis of controlled substances.^{10,11} Several possible mechanisms of ionisation have been postulated, including chemical sputtering involving gas-phase ions generated by electrospray ionisation (ESI) or corona discharge and subsequent charge transfer between these primary ions and sample molecules on the surface. The occurrence of gas-phase ion-molecule reactions has also been suggested together with a droplet splashing or pick-up mechanism. This involves the impacting of multiply charged solvent droplets dissolving sample molecules from the surface leading to the formation of secondary charged droplets carrying sample molecules and resulting in ion formation mechanisms similar to that of ESI.^{1,4–6}

Desorption atmospheric pressure chemical ionisation (DAPCI) was first reported for the trace level detection of

TNT, PETN and RDX explosives,⁵ and has not received as much attention to date as DESI. Ionisation of the explosives was provided by the initial formation of toluene or methanol reagent ions produced by a corona discharge. These reagent ions formed are thought to ionise the analyte molecules by either electron or proton transfer in a chemical ionisation step.⁵ For those compounds that do not provide sufficient ion intensity by DESI, DAPCI offers an alternative option. DAPCI has been shown to provide increased sensitivity for compounds of moderate polarity.⁷ DAPCI generated higher signal intensities for the active ingredient hydrocortisone, a weakly polar corticosteroid, than DESI using the same solvent system normally used for conventional ESI experiments.⁷ The DAPCI technique previously reported⁷ used nitrogen sheath gas and a mixture of methanol/water (1:1) from which ions were produced by a corona discharge. Reagent ions formed in the corona discharge region react with desorbed analyte molecules forming, depending on the ionisation mode, for the most part, protonated or deprotonated molecules.⁷ A variant to the DAPCI technique for the rapid analysis of volatile and semi-volatile compounds, referred to as atmospheric pressure solids analysis probe (ASAP), was recently used in the analysis of a number of steroids and biological tissues.¹² Vapourisation of the sample is accomplished by placing it within the hot flowing nitrogen gas at atmospheric pressure.

Another rapid and newly developed ionisation method, direct analysis in real time (DART), was recently reported.¹³ Pharmaceutical samples, explosives and metabolites in urine were analysed on a single ToF instrument. Sample analysis

*Correspondence to: J. P. Williams, Department of Biological Sciences, University of Warwick, Gibbet Hill Road, Coventry CV4 7AL, UK.
E-mail: j.p.williams@warwick.ac.uk

by the DART technique is carried out at ambient temperature and ionisation is brought about by exposing the sample to a stream of excited gas, typically helium. An electrical discharge produces ions, electrons and excited state metastable neutral species. Several ionisation mechanisms have been postulated, including Penning ionisation, in which ionisation of the sample occurs by energy transfer from an excited atom or molecule of energy greater than the ionisation energy of the sample.¹³ It was reported that when helium is used as the gas, the mechanism involves the formation of ionised water clusters followed by proton transfer reactions.¹³

In this present study we have made use of these three newly developed ionisation techniques. The main focus of this study was to evaluate the potential of the ionisation techniques for the rapid analysis of active ingredients formulated into a variety of pharmaceutical tablets, gels and ointments. The rapid analysis of pharmaceutical drug formulations, the detection from human skin of the active ingredients from ointments or gels, and the detection of ibuprofen metabolites in human urine have been carried out. The increased selectivity and specificity of the DESI and DAPCI techniques used with a hybrid quadrupole ToF instrument were compared with that obtained by the use of DART with a single ToF instrument. Accurate mass MS and accurate mass MS/MS measurements, to within 2 mTh, has allowed elemental compositions for the product ions to be determined, thereby facilitating the identification of fragmentation pathways. We believe this to be the first report exploiting all three newly developed ionisation techniques combined with accurate mass DAPCI-MS/MS.

EXPERIMENTAL

The tablets and ointment formulations investigated are listed in Table 1, together with the molecular formulae and molecular weights of the active ingredients. Ibuprofen tablets and gel (Tesco, UK), Anadin Extra (Wyeth, UK) and Solpadeine Max (Tesco, UK) were purchased without prescription from a pharmacy. Proctosedyl ointment (Aventis)

and metoclopramide (APS) were obtained by prescription. Solvents were obtained from Sigma-Aldrich (Poole, UK).

Mass spectrometry

DART experiments were carried out on an AccuToF LC ToF mass spectrometer (Jeol, Peabody, MA, USA) and DESI/DAPCI experiments were carried out on a Q-ToF I (Waters, Manchester, UK). Experimental conditions for each are given below.

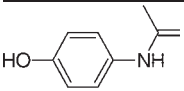
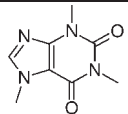
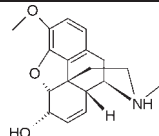
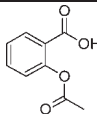
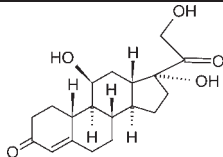
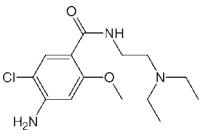
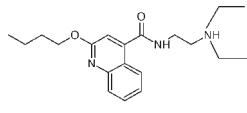
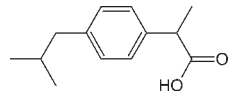
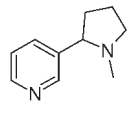
DART

A detailed description of the DART source can be found elsewhere.¹³ The AccuToF instrument was operated with helium flowing into the DART source and a voltage of 2 kV applied to the discharge needle in positive mode of ionisation. Orifice 1 of the interface was set to 27 eV. This voltage can be increased or decreased depending on the amount of fragmentation desired. The gas temperature was maintained at 80°C and the operating resolution of the instrument was approximately 6000 (FWHM). Mass spectra were acquired over the mass range of m/z 50–500 at an acquisition rate of 0.5 spectrum/s. For DART sample analysis the helium gas was directed towards the sample or allowed to interact with vapour-phase samples. Tablets were broken, to expose an uncoated sample surface, before being held with tweezers in the path of the flowing helium at atmospheric pressure. Samples in solution were analysed by placing filter paper (1 cm × 8 cm) in the solution prior to being held similarly. For ointments, approximately 100 mg was applied to the surface of a piece of matt-finished cardboard (1 cm × 2 cm) and held in the same position.

DESI

The Q-ToF I instrument was operated in positive and negative mode with a capillary voltage of 3.5 kV and –3.2 kV. The ion source block and nitrogen desolvation gas temperatures were set to 100°C and 400°C and the desolvation gas was set to a flow rate of 300 L/h. The cone voltage was set at 20 V for MS and MS/MS experiments and the collision energy used for MS/MS experiments was ramped between

Table 1. Structures, molecular formulae and molecular weights of the compounds investigated

				
$C_8H_9NO_2$ paracetamol MW:151	$C_8H_{10}N_4O_2$ caffeine MW:194	$C_{18}H_{22}NO_3$ codeine MW:299	$C_9H_8O_4$ aspirin MW:180	$C_{21}H_{30}O_5$ hydrocortisone MW:362
				
$C_{14}H_{22}ClN_3O_2$ metoclopramide MW:299	$C_{20}H_{29}N_3O_2$ cinchocaine hydrochloride MW:343	$C_{13}H_{18}O_2$ ibuprofen MW:206	$C_{10}H_{14}N_2$ nicotine MW:162	

10 and 25 eV during the acquisition. The ToF mass analyser was operated at a resolution of approximately 6000 (FWHM), with spectra acquired over the mass range of m/z 50–500 at an acquisition rate of 1 spectrum/s. For all MS/MS experiments argon was used as the collision gas. For DESI sample analysis, each tablet was broken, to expose an uncoated sample surface, before being held with tweezers, at an angle of approximately 45° to the solvent spray and a distance of 5 mm from the source sampling cone. Approximately 100 mg of the ointment was applied to the surface of a piece of matt-finished cardboard (1 cm × 2 cm) and held in the same position as the solid tablets. The surface of the tablet or card was then sprayed with a solution of acetonitrile/H₂O (1:1) in negative mode and a solution of acetonitrile/H₂O + 0.2% formic acid in positive mode at a flow rate of 10 μL/min, using a model 22 syringe pump from Harvard Apparatus (South Natick, MA, USA). No extensive modification of solvents, buffers and pH was carried out.

DAPCI

DAPCI experiments were performed on the Q-ToF I in both positive and negative modes of ionisation. The corona discharge pin voltage was set to 3.5 kV and –3.0 kV in positive and negative modes, respectively. The cone voltage was optimised between 10 and 25 V for each sample. The collision energy used for MS/MS experiments was ramped between 10 and 25 eV during the acquisition. The flow rate of the nitrogen desolvation gas was set to 150 L/h. The source and probe temperatures were set to 100°C and 400°C, respectively. A solvent mixture of methanol and water (1:1) flowing at 10 μL/min was infused into the heated nebuliser probe where it was converted into an aerosol which was rapidly heated in a stream of nitrogen gas, forming a vapour at the probe tip. The probe tip directly faced the tablet or ointment (which had been deposited onto the card) positioned between the corona discharge pin and the sampling cone. Reagent ions formed in the corona discharge region reacted with desorbed analyte molecules from the tablet or card forming, depending on the ionisation mode, for the most part, protonated or deprotonated molecules. The same experiments were also performed without any solvent flowing into the heated probe. This solventless DAPCI experiment is similar to the ASAP experiment previously described.¹²

Accurate mass measurement protocol for the Q-ToF I

Instrumental mass drift was corrected for by using a single internal reference lock mass in MS and MS/MS mode on the Q-ToF. Since the target compounds in this study are known, the precursor ion selected for MS/MS experiments provided the internal reference lock mass in MS/MS mode. Data acquisition and processing were carried out using the digital dead-time correction algorithm embedded in the operating software, MassLynx (V3.5), supplied by Waters UK. The high ion counts generated using the DAPCI and DESI techniques caused time-to-digital (TDC) dead-time saturation.⁷ The TDC correction software was utilised and displayed a peak centroid with the correct mass and signal intensity.

RESULTS AND DISCUSSION

An investigation has been carried out of a number of prescription and non-prescription pharmaceutical formulations. Active ingredients from a gel applied to human skin and drug metabolites in human urine were also investigated. DART, DESI and DAPCI (solvent/solventless) techniques were employed. The DESI and DAPCI techniques can be used as a rapid screen for the analysis of solids and liquids and can be set up quickly on most instruments by the use of a home-built source,^{1,10} or where the samples are handled manually.^{7,11} The relative signal intensities of the active ingredients are different when samples are held manually. This is due to the non-reproducibility of positioning of the sample in the solvent spray. The use of these techniques for the rapid analysis of samples provides similar information to that provided by the DART source. The active ingredients formulated in the solid tablets were detected by the DART technique and the analysis time in MS mode of operation was, for the majority of samples, as rapid as in the DESI and DAPCI techniques, requiring, in most cases, less than 5 s.

The analysis of solid tablets and the identification of ranitidine metabolites in human urine using the DART source have previously been reported.¹³ Since DAPCI has received little attention, a representative sample of results will focus on this together with others obtained using DESI. A comparison of the three techniques is made for the detection of a range of drugs. Tandem mass spectrometry (MS/MS) was performed on either the protonated or deprotonated molecule. The ion selected for collision-induced dissociation (CID) using the DAPCI and DESI techniques is annotated with a filled black circle and included in each figure for clarity. Some of the fragmentation pathways are complex and a detailed description of these has not been included. The proposed fragmentation schemes for some of the molecules are not assumed to be sequential. MS/MS of the protonated molecule $[M+H]^+$ of nicotine, for example, forms a product ion at m/z 106. This ion could possibly be formed by more than one route; m/z 163 → 106 (loss of C₃H₇N) or m/z 163 → 132 → 106 (sequential loss of CH₃NH₂ and C₂H₂).

Analysis of solid tablets in positive ion DAPCI, DESI and DART mode

A solid Anadin Extra tablet, which contained 45 mg of caffeine, 200 mg of paracetamol and 300 mg of aspirin, was analysed by DAPCI, DESI and DART. The mass spectra obtained for Anadin Extra using the DAPCI, DESI and the DART sources are shown in Figs. 1(A)–1(C). The protonated molecules for each of the active ingredients, paracetamol at m/z 152, aspirin at m/z 181 and caffeine at m/z 195, were observed in the DAPCI and DESI mass spectra but the protonated aspirin molecule was absent in the DART spectrum. The base peak in all the spectra is m/z 195, protonated caffeine. Accurate mass measurement confirms that the ion at m/z 163 is formed by loss of H₂O from protonated aspirin. This ion was observed in all three spectra. The ion at m/z 198 in the DAPCI and DESI spectra is ammoniated aspirin, $[M+NH_4]^+$, m/z 203 in the DESI spectrum is sodiated aspirin, $[M+Na]^+$. Figure 1(D) shows

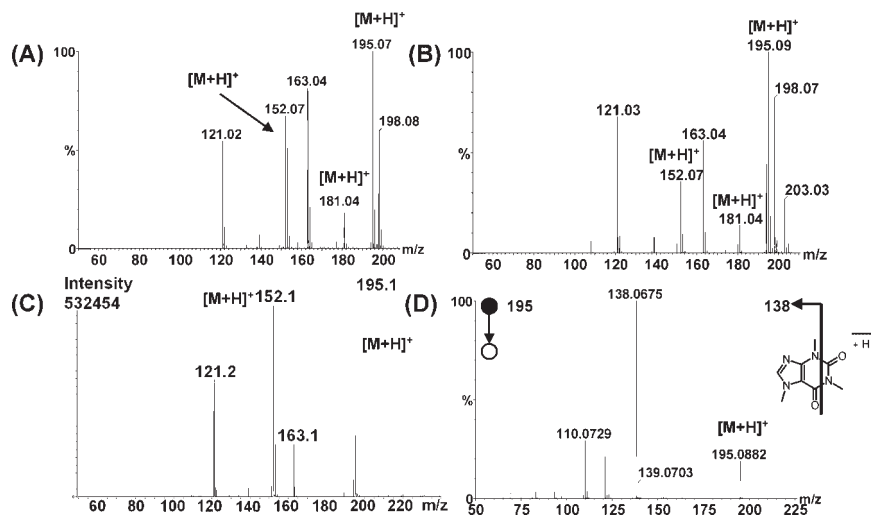


Figure 1. (A) Positive ion DAPCI-MS spectrum of an Anadin Extra tablet. (B) Positive ion DESI-MS spectrum of an Anadin Extra tablet. (C) Positive ion DART MS spectrum of an Anadin Extra tablet. (D) Positive ion DAPCI accurate mass MS/MS spectrum obtained from the protonated molecule of caffeine, $[M+H]^+$ of m/z 195.

the accurate mass MS/MS spectrum obtained for protonated caffeine. The elemental composition assignments for the product ions obtained for caffeine are given in Table 2. The base peak in the MS/MS spectrum is m/z 138. The MS/MS spectrum obtained during this study is very similar to that obtained previously.¹⁴ Elemental composition assignment shows that the protonated molecule primarily fragments by the loss of methyl isocyanate (57 Da) forming an even-electron ion at m/z 138 with the formula of $C_6H_8N_3O$. Two other product ions were generated with sufficient ion counts for the generation of elemental formulae: m/z 121 is an odd-electron radical cation with probable formula of $C_5H_3N_3O$ and the ion at m/z 110 has the elemental formula of $C_5H_8N_3$ formed through loss of CO from m/z 138.

A Solpadeine Max tablet containing 500 mg of paracetamol and 12.8 mg of codeine phosphate was analysed. The mass spectra obtained using the DAPCI, DESI and DART sources are shown in Figs. 2(A)–2(C). The m/z 152 base peak in the DAPCI and DART spectra is from protonated paracetamol. The base peak in the DESI spectrum is the dimer $[2M+H]^+$ at m/z 303 from paracetamol. The formation of dimers has been observed in all three techniques and is related to the local concentration of the active ingredient under study. The ratio of monomer to dimer species can vary with the exposed part of the tablet under investigation. The protonated molecule from codeine was more abundant than the protonated dimer of paracetamol using DART. The reverse was observed for

DAPCI and DESI. Other minor ions observed in the spectra are probably due to additives in the tablet formulation. The MS/MS spectrum of protonated codeine is shown in Fig. 2(D). The product ion spectrum generated from protonated codeine can be seen to be very complex but it is also specific and reproducible. The elemental composition assignments for some of the product ions obtained for codeine are given in Table 3.

Analysis of the active ingredients of a gel formulation by desorption from human skin by DESI

A thin layer of Ibuprofen gel containing 5% w/w of the active ingredient was applied to the surface of a human finger. The gel was gently massaged until absorbed by the skin. Using the DESI technique, with a source and desolvation temperature of 100°C, one could readily detect the drug at the point of application. Figure 3(A) shows the negative DESI mass spectrum, generated in 2 s, obtained 20 min after applying the gel. The base peak in the spectrum is deprotonated ibuprofen. Figure 3(B) shows the accurate mass MS/MS spectrum obtained for the deprotonated ibuprofen molecule 20 min after applying the gel. The deprotonated ibuprofen, $[M-H]^-$ at m/z 205.1229, was used as the internal lock mass. Accurate mass measurement of the product ion at m/z 161.1321 confirms that CO_2 is lost from the deprotonated molecule. This ion has an empirical formula of $C_{12}H_{17}$, a mass difference of -0.9 mTh from the theoretical monoisotopic mass. [Caution: there is a risk of electric shock from the capillary. Although the capillary carries only limited current, it is advisable to keep the finger well away from the capillary tip in the experiment].

Comparison of DART/DESI/DAPCI for the detection of two active ingredients in Proctosedyl ointment

It has been previously demonstrated that the detection of the two active ingredients formulated into proctosedyl ointment, cinchocaine hydrochloride (5 mg/g) and hydrocortisone

Table 2. DAPCI-MS/MS accurate mass of m/z 195 $[M+H]^+$ of caffeine

Measured mass	Formulae	Error (mTh)
138.0675	$C_8H_{10}O_2$	-0.6
	$C_6H_8N_3O$	0.8
121.0293	$C_7H_5O_2$	0.4
	$C_5H_3N_3O$	1.7
110.0729	$C_6H_{10}O$	-0.3
	$C_5H_8N_3$	1.1

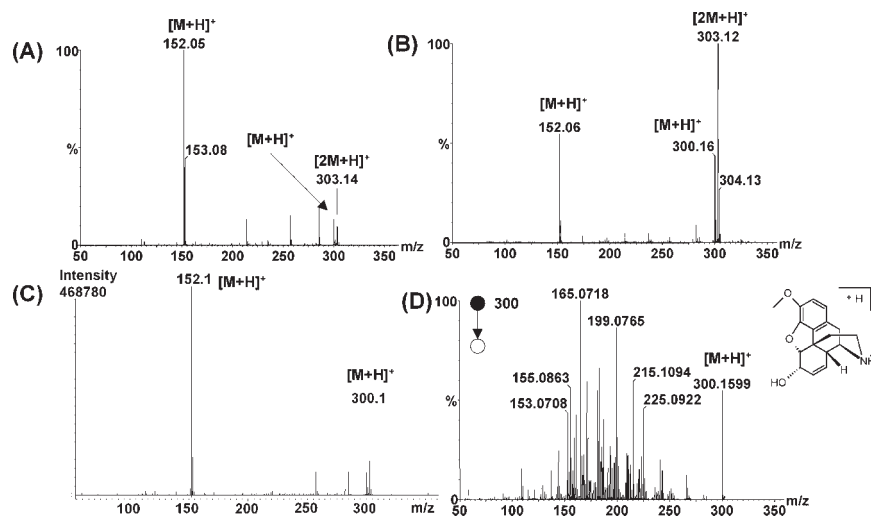


Figure 2. (A) Positive ion DAPCI-MS spectrum of a Solpadeine Max tablet. (B) Positive ion DESI-MS spectrum of a Solpadeine Max tablet. (C) Positive ion DART MS spectrum of a Solpadeine Max tablet. (D) Positive ion DAPCI accurate mass MS/MS spectrum obtained from the protonated molecule of codeine, $[M+H]^+$ of m/z 300.

(5 mg/g), was better with DAPCI than with DESI experiments. In positive ion mode, hydrocortisone, a weakly polar corticosteroid, was more effectively ionised by DAPCI.⁷ During this investigation, the ointment was applied to the card in the usual way and investigated in both negative and positive ion DAPCI modes of operation. The mass spectrum obtained for the ointment in negative ion mode is shown in Fig. 4(A). High ion counts were generated, highlighting the sensitive response of the ointment to the DAPCI process in negative mode. The base peak in the spectrum is the deprotonated molecule of cinchocaine at m/z 342. Other ions observed in the spectrum arise from additives in the ointment formulation. No ion was observed for deprotonated hydrocortisone. The mass spectrum obtained for the ointment in positive ion mode is shown in Fig. 4(B). High ion counts were generated, highlighting the sensitive response of the ointment to the DAPCI process in positive ion mode. The base peak in the spectrum is protonated cinchocaine at m/z 344. The ion at m/z 363 corresponds to protonated hydrocortisone.

The product ion spectra of the deprotonated and protonated molecules of cinchocaine can be compared in Figs. 4(C) and 4(D). Figure 4(C) shows the negative ion accurate mass MS/MS spectrum obtained for deprotonated cinchocaine. The elemental composition assignments for the product ions are given in Table 4. The base peak in the MS/MS spectrum is formed by the loss of $C_7H_{14}N_2O$ from the deprotonated molecule to the ion at m/z 200. This ion further

fragments to m/z 144 by loss of 56 Da. Accurate mass measurement confirms this to be a loss of C_4H_8 . A proposed fragmentation scheme is shown in Fig. 4(E).

Figure 4(D) shows the positive ion accurate mass MS/MS spectrum obtained for protonated cinchocaine. The probable elemental composition assignments for the product ions are given in Table 5. The base peak in the MS/MS spectrum is formed by the loss of $C_4H_{11}N$ from the protonated molecule to the ion at m/z 271. This ion further fragments to m/z 215 by loss of 56 Da. Accurate mass measurement confirms this to be a loss of C_4H_8 . A proposed fragmentation scheme is shown in Fig. 4(F).

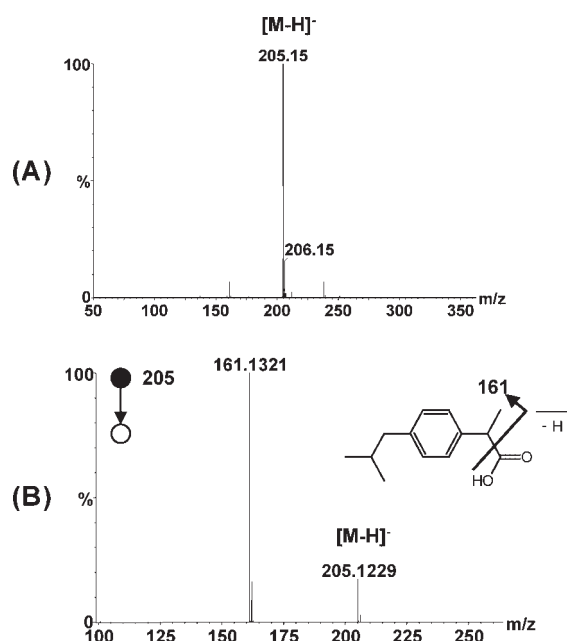


Figure 3. (A) Negative ion DESI-MS spectrum of Ibuprofen gel desorbed off skin. (B) Negative ion DAPCI accurate mass MS/MS spectrum obtained from the deprotonated molecule of ibuprofen, $[M-H]^-$ of m/z 205.

Table 3. DAPCI-MS/MS accurate mass of m/z 300 $[M+H]^+$ of codeine

Measured mass	Formulae	Error (mTh)
266.1196	$C_{17}H_{16}NO_2$	1.5
225.0922	$C_{15}H_{13}O_2$	0.6
215.1094	$C_{14}H_{15}O_2$	2.2
199.0765	$C_{13}H_{11}O_2$	0.6
183.0800	$C_{13}H_{11}O$	-0.9

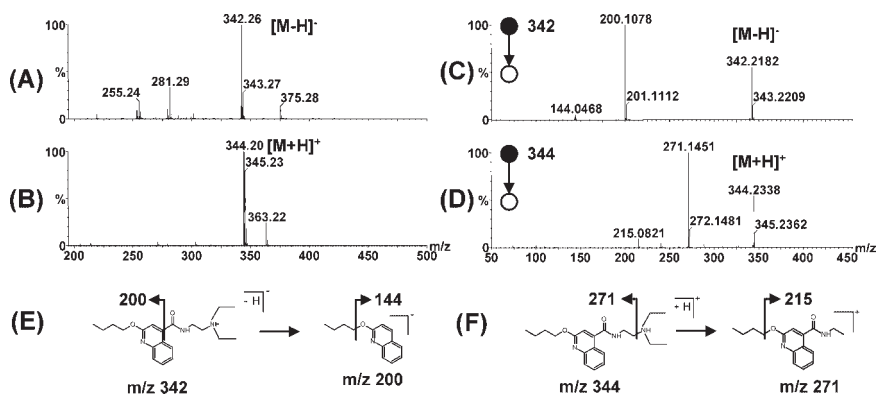


Figure 4. (A) Negative ion DAPCI-MS spectrum of Proctosedyl ointment. (B) Positive ion DAPCI-MS spectrum of Proctosedyl ointment. (C) Negative ion DAPCI accurate mass MS/MS spectrum of m/z 342, the deprotonated molecule $[M-H]^-$ of cinchocaine hydrochloride. (D) Positive ion DAPCI accurate mass MS/MS spectrum of m/z 344, the protonated molecule $[M+H]^+$ of cinchocaine hydrochloride. (E) Proposed negative ion fragmentation pathway of $[M-H]^-$ of cinchocaine hydrochloride. (F) Proposed positive ion fragmentation pathway of $[M+H]^+$ of cinchocaine hydrochloride.

The detection of active ingredients in ointments and creams using this approach is non-labour-intensive since the analysis requires no prior analyte extraction. A comparison of the DESI and DAPCI (solvent and solventless) techniques was made for this ointment since it contained both polar and weakly polar ingredients. Figure 5 shows the comparison. The mass spectrum of each is shown over the m/z region of 300–400. The top spectrum was obtained using DESI, the middle spectrum obtained using DAPCI without the use of solvent, and the bottom spectrum was obtained using DAPCI with solvent flowing into the heated probe. Protonated molecules were detected for both cinchocaine (m/z 344) and hydrocortisone (m/z 363) by all three techniques. The top two spectra have been magnified ($\times 16$) over the m/z region for hydrocortisone. The $[M+H]^+$ ion from cinchocaine is the base peak in all three spectra showing that this compound ionises more efficiently than hydrocortisone. The hydrocortisone has a relative abundance of approximately 3%, 4% and 35% compared with cinchocaine in DESI, DAPCI without solvent and DAPCI with solvent modes, respectively. This demonstrates the more effective detection of all the active ingredients in this formulation when using DAPCI with solvent.

Figure 6(A) shows the positive ion mass spectra, over the region m/z 340–365, for the same ointment when using DART and DAPCI with solvent. DART shows results similar to the

Table 4. DAPCI-MS/MS accurate mass of m/z 342 $[M-H]^-$ of cinchocaine

Measured mass	Formulae	Error (mTh)
200.1078	$C_{13}H_{14}NO$	0.3
144.0468	C_9H_6NO	1.9

Table 5. DAPCI-MS/MS accurate mass of m/z 344 $[M+H]^+$ of cinchocaine

Measured mass	Formulae	Error (mTh)
271.1451	$C_{16}H_{19}N_2O_2$	0.4
215.0821	$C_{12}H_{11}N_2O_2$	0.0

DESI and DAPCI without solvent experiments. The active ingredient cinchocaine produces the base peak and the peak due to the hydrocortisone has a relative abundance of approximately 10% compared with cinchocaine.

Comparison of positive and negative DAPCI-MS/MS

Metoclopramide is a medicine that increases the movements or contractions of the stomach and intestines and it is used to treat the symptoms of a stomach problem called diabetic gastroparesis. A metoclopramide tablet which contained 10 mg of the active ingredient was investigated by DAPCI in positive and negative ion modes. The negative ion mass spectrum of metoclopramide is shown in Fig. 7(A). The deprotonated molecule for the active ingredient was

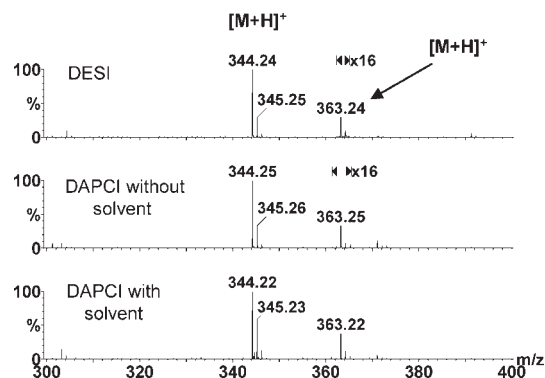


Figure 5. Top: positive ion DESI-MS spectrum of Proctosedyl ointment. The m/z range over m/z 363 (protonated molecule, $[M+H]^+$ of hydrocortisone) has been expanded showing a low abundance ion at m/z 363 which has been magnified ($\times 16$). Middle: positive ion DAPCI-MS spectrum of Proctosedyl ointment obtained without solvent. The m/z range over m/z 363 has been expanded showing a low abundance ion at m/z 363 which has been magnified ($\times 16$). Bottom: positive ion DAPCI-MS spectrum of Proctosedyl ointment obtained with solvent.

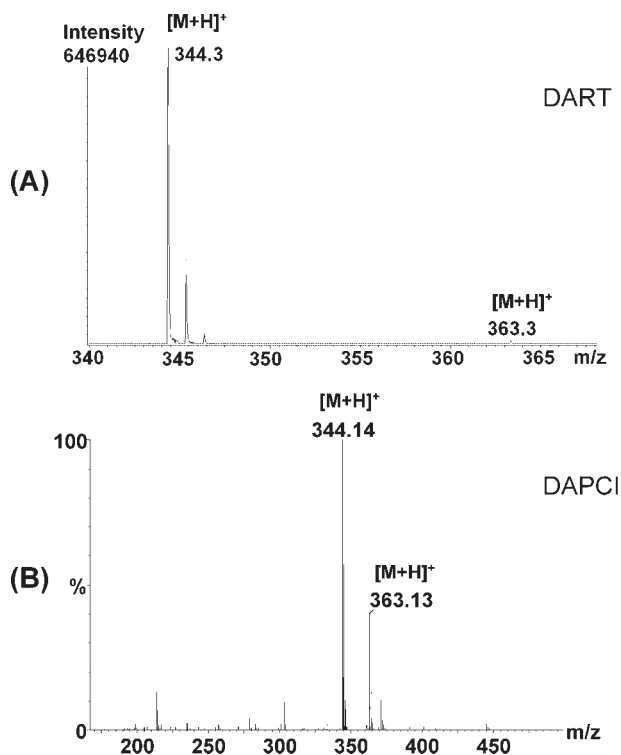


Figure 6. (A) Positive ion DART-MS spectrum of Proctosedyl ointment. (B) Positive ion DAPCI-MS spectrum of Proctosedyl ointment obtained with solvent.

observed in the spectrum at m/z 298 and 300. These two peaks have relative signal intensities of 3:1, consistent with the presence of one chlorine atom in the molecule. The positive ion mass spectrum is shown in Fig. 7(B). The protonated

molecule for the active ingredient was observed as the base peak in the spectrum at m/z 300. The positive ion mass spectrum obtained using DART for metoclopramide is shown in Fig. 7(E). Once again, the protonated molecule for the active ingredient at m/z 300 was the base peak.

A comparison of the product ion mass spectra of protonated and deprotonated metoclopramide is shown in Figs. 7(C) and 7(D). Figure 7(C) shows the accurate mass MS/MS spectrum obtained for deprotonated metoclopramide. The elemental composition assignments for the product ions are given in Table 6. The deprotonated molecule appears to fragment by the unusual loss of a methyl radical forming an odd-electron radical anion of m/z 283. This ion further fragments to m/z 211 by loss of 72 Da, the diethylamine moiety. Accurate mass measurement confirms this to be a loss of $C_4H_{10}N$. A proposed fragmentation scheme is shown in Fig. 7(F).

Figure 7(D) shows the accurate mass MS/MS spectrum obtained for protonated metoclopramide. The spectrum shows ions at m/z 227 and 184, in agreement with results obtained previously.¹⁵ The elemental composition assignments for the product ions are given in Table 7. The base peak in the MS/MS spectrum at m/z 227 shows that the protonated molecule primarily fragments by the loss of $C_4H_{11}N$, diethylamine. This ion further fragments to m/z 184 by loss of 43 Da. Accurate mass measurement confirms this to be a loss of C_2H_5N . A proposed fragmentation scheme is shown in Fig. 7(G).

Analysis of a naturally occurring plant alkaloid

DAPCI, DESI and DART analysis of tobacco was performed by removing the tobacco from a cigarette and holding the

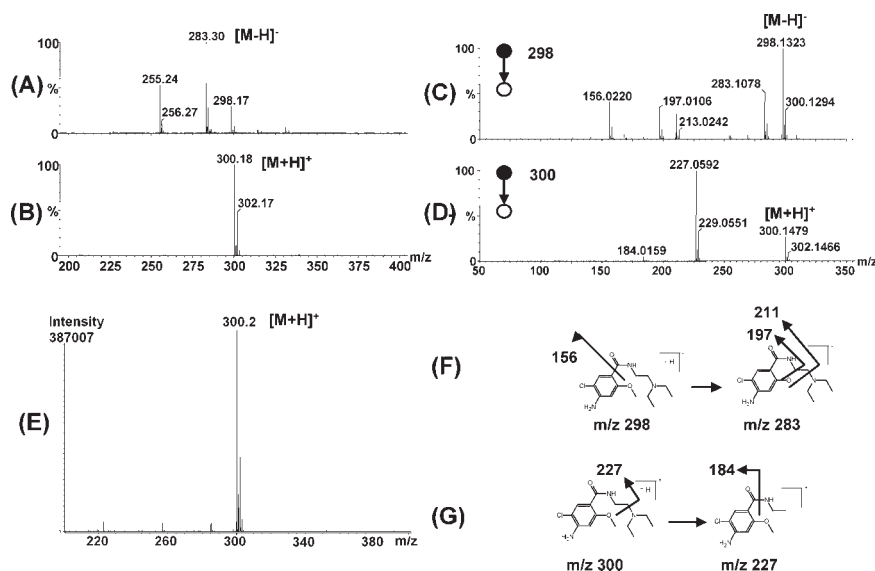


Figure 7. (A) Negative ion DAPCI-MS spectrum of metoclopramide. (B) Positive ion DAPCI-MS spectrum of metoclopramide. (C) Negative ion DAPCI accurate mass MS/MS spectrum of m/z 298, the deprotonated molecule $[M-H]^-$ of metoclopramide. (D) Positive ion DAPCI accurate mass MS/MS spectrum of m/z 300, the protonated molecule $[M+H]^+$ of metoclopramide. (E) Positive ion DART-MS spectrum of metoclopramide. (F) Proposed negative ion fragmentation pathway of $[M-H]^-$ of metoclopramide. (G) Proposed positive ion fragmentation pathway of $[M+H]^+$ of metoclopramide.

Table 6. DAPCI-MS/MS accurate mass of m/z 298 $[M-H]^-$ of metoclopramide

Measured mass	Formulae	Error (mTh)
283.1078	$C_{13}H_{18}ClN_3O_2$	-0.9
211.0258	$C_9H_8ClN_2O_2$	-1.6
197.0106	$C_8H_6ClN_2O_2$	-1.2
156.0220	C_7H_7ClNO	0.3

Table 7. DAPCI-MS/MS accurate mass of m/z 300 $[M+H]^+$ of metoclopramide

Measured mass	Formulae	Error (mTh)
227.0592	$C_{10}H_{12}ClN_2O_2$	0.4
184.0159	$C_8H_7ClNO_2$	-0.7

contents with tweezers by the method described in the Experimental section. Nicotine is an alkaloid and is known to be found naturally in the tobacco plant. The DAPCI, DESI and DART mass spectra obtained in positive ion mode are shown in Figs. 8(A)–8(C). A low cone voltage produced a single ion at m/z 163, corresponding to protonated nicotine, using DAPCI and DESI. Figure 8(C) shows the spectrum obtained using the DART source at a high cone voltage. A high cone voltage produced in-source fragmentation and the DART spectrum is very similar to the DAPCI-MS/MS spectrum of m/z 163 shown in Fig. 8(D). The elemental composition assignments for the product ions obtained for nicotine are given in Table 8. These show that the protonated molecule fragments by the loss of 31 Da to form an ion at m/z 132. Accurate mass measurement confirms this loss to be CH_3NH_2 from the methyl-substituted pyrrolidine ring. The base peak in the MS/MS spectrum is m/z 130 formed by loss

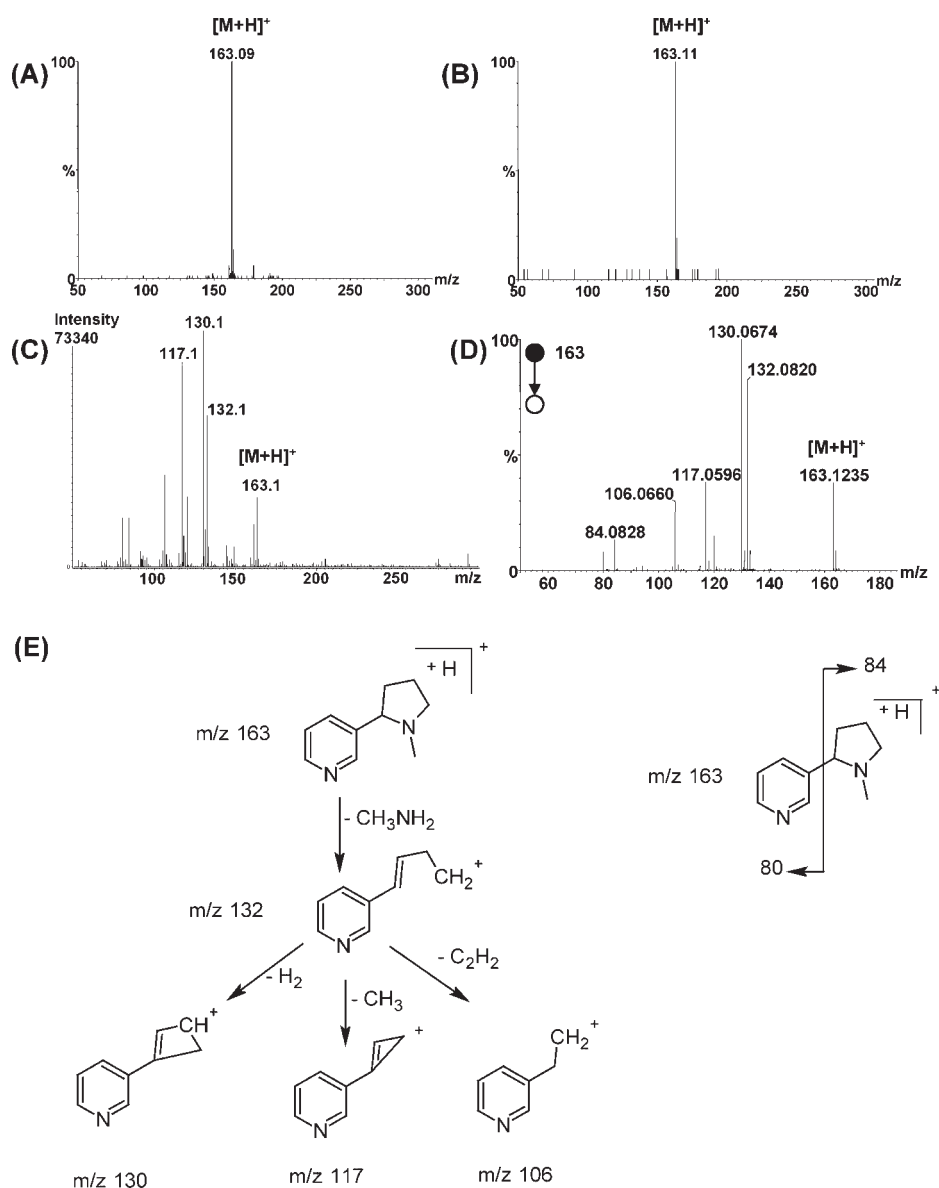


Figure 8. (A) Positive ion DAPCI-MS spectrum of tobacco. (B) Positive ion DESI-MS spectrum of tobacco. (C) Positive ion DART-MS spectrum of tobacco. (D) Positive ion DAPCI accurate mass MS/MS spectrum of m/z 163, the protonated molecule $[M+H]^+$ of the active ingredient nicotine found in tobacco. (E) Proposed positive ion fragmentation pathway of $[M+H]^+$ of nicotine.

Table 8. DAPCI-MS/MS accurate mass of m/z 163 $[M+H]^+$ of nicotine

Measured mass	Formulae	Error (mTh)
132.0820	C ₉ H ₁₀ N	0.7
130.0674	C ₉ H ₈ N	1.7
117.0596	C ₈ H ₇ N	1.8
106.0660	C ₇ H ₈ N	0.4
84.0828	C ₅ H ₁₀ N	1.5
80.0524	C ₅ H ₆ N	2.4

of H₂ from m/z 132. Other ions at m/z 117 and 106 are possibly formed by the loss of a methyl radical and a C₂H₂ molecule, respectively, from m/z 132. The methyl radical loss has been observed previously.¹⁶ Other easily recognised product ions include the ion of m/z 84 due to the methyl-substituted pyrrolidine ring moiety obtained from cleavage of the bond between the two rings and the ion of m/z 80 due to the pyridine ring moiety of the molecule, obtained in the same way. A proposed fragmentation scheme is shown in Fig. 8(E), and is similar to the scheme previously proposed.¹⁶ Accurate mass MS/MS confirms that the previous proposals are correct.

Identification of ibuprofen metabolites from human urine

The development of a rapid screen for the detection of drug or drug metabolites in urine is simplified by the use of these newly developed ionisation techniques. The study of drug metabolites in urine can be challenging due to the low levels of metabolites and high levels of endogenous materials such as salts. Ibuprofen is a non-steroidal anti-inflammatory drug and the known bio-transformations of ibuprofen include hydroxylation, carboxylation and glucuronidation.¹⁷ A control sample of human urine was obtained 60 min after administration of two ibuprofen tablets which each con-

tained 200 mg of the active ingredient. We used DAPCI, with a mixture of methanol/water flowing into the heated probe, for the detection of ibuprofen metabolites in human urine.

A urine sample was absorbed onto filter paper (1 cm × 8 cm) by placing the paper into the urine sample. Metabolites in the urine were then identified by holding the filter paper with tweezers as described in the Experimental section. Figure 9(A) shows the negative ion mass spectrum obtained without any pre-treatment of the urine sample. The spectrum contains ions attributable to the presence of endogenous components, of which many can be ascribed to components in the urine and some to background species. Ions detected other than the known bio-transformations of ibuprofen have been tentatively assigned to deprotonated pyruvic acid at m/z 87, lactic acid at m/z 89, methylmalonic acid at m/z 117, xanthine at m/z 151, and hippuric acid at m/z 178.

Figure 9(B) shows the accurate mass spectrum obtained using the deprotonated molecule $[M-H]^-$ at m/z 205.1229 from the parent drug ibuprofen, as the internal lock mass. Ions corresponding to the deprotonated molecules of two low-level metabolites were detected, the hydroxyl-ibuprofen at m/z 221 and carboxy-ibuprofen at m/z 235. No glucuronide metabolites were detected. The specificity of the technique coupled with accurate mass measurement has allowed the determination of an elemental composition for m/z 221. The probable elemental composition for the accurate mass of 221.1163 is C₁₃H₁₇O₃, consistent with the formula for a hydroxylated ibuprofen metabolite. The mass error calculated is -1.5 mTh from the theoretical monoisotopic mass for hydroxyl-ibuprofen metabolite. The elemental composition for the accurate mass of 235.0983 is C₁₃H₁₅O₄, consistent with the formula for a carboxylated ibuprofen metabolite. The mass error calculated is 1.2 mTh from the theoretical monoisotopic mass for the carboxy-ibuprofen metabolite. Figure 9(C) shows the MS/MS spectrum obtained for the

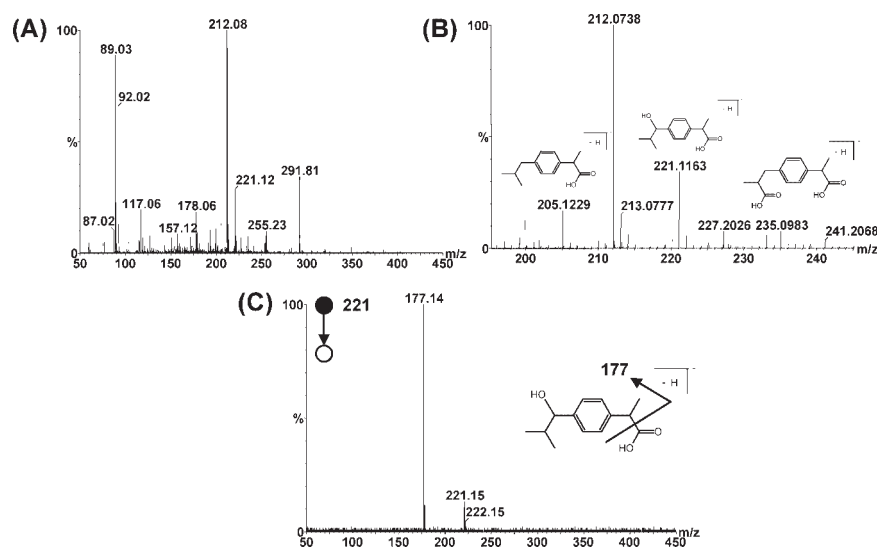


Figure 9. (A) Negative ion DAPCI-MS spectrum of urine sample, 75 min after oral dose of 400 mg of ibuprofen tablets. (B) Negative ion DAPCI accurate mass spectrum of a urine sample, 75 min after an oral dose of 400 mg of ibuprofen tablets. The mass spectrum has been expanded over the m/z range 195–245. (C) Negative ion DAPCI-MS/MS spectrum of m/z 221 showing loss of CO₂.

deprotonated hydroxy-ibuprofen metabolite at m/z 221. This primarily fragments with the formation of a single product ion at m/z 177 through the loss of CO_2 .

CONCLUSIONS

The use of DART, DESI and DAPCI (solvent and solventless) techniques has been demonstrated for a wide range of molecules of pharmaceutical interest. In all cases the sample to be studied has been desorbed from a solid substrate. The well-demonstrated ability of the DART technique to rapidly characterise gaseous samples was not utilised. Significant differences between these approaches were observed. For compounds of moderate to low polarity, DAPCI (solvent and solventless) produced more effective ionisation. DAPCI with solvent was observed to be more effective in the ionisation of these compounds than DART, DESI or solventless DAPCI. The previously described approach of ASAP¹² was found to be very similar to solventless DAPCI.

DESI and DAPCI techniques have provided a highly robust means of interrogating the active ingredients of a variety of pharmaceutical formulations. Sampling the formulation is rapid and ionisation occurs almost instantly. The orientation of the surface in front of the stream of excited gas seems to be more critical with the DART source. The current design of the DART source, although very effective for the analysis of gaseous samples, is less effective for the analysis of adsorbed molecules. This is due to the inability of visualise the gas stream and the relatively wide sampling geometry.

The surface to be analysed via DESI can be held in the solvent stream at an angle of 45° and at a distance of 5–20 mm prior to the sampling cone orifice or it can be held at an angle directly in front of the probe tip some distance from the sampling cone on the QToF 1 instrument. Since the spray emanating from the probe tip can be seen in DESI mode, the tip can be positioned exactly to spray the surface to be interrogated, which can be of the order of cm^2 to sub-mm^2 .

The use of a Q-ToF instrument, with elevated resolution and full-scan sensitivity, has improved the selectivity and specificity of the DAPCI technique originally shown by Cooks and co-workers⁵ by allowing the generation of accurate mass MS/MS information to within 2 mTh using a single point internal lock mass to correct mass scale drift. The high information content has allowed probable elemental compositions and fragmentation pathways to be determined.

The development of these new ionisation techniques offers very significant advantages for a number of important

scientific areas. The ability to rapidly analyse complex mixtures with little, or no, sample preparation is very important. The techniques DART, DESI and DAPCI (solvent and solventless) have complex, potentially inter-related mechanisms but have been shown to provide complementary information on a range of compounds of pharmaceutical interest. The direct analysis of urine and molecules adsorbed on skin is of particular interest. This work has focused on samples desorbed from solid substrates although the study of gaseous samples is also possible, particularly with the DART technique. The techniques produce little fragmentation and so the addition of accurate mass MS and MS/MS greatly increases the selectivity of the approach.

Acknowledgements

The authors would like to thank JEOL for the kind access to the AccuToF incorporating the DART source, the Proteomics Facility at the University of Warwick, and Professor Keith Jennings for helpful discussions concerning the manuscript.

REFERENCES

1. Takats Z, Wiseman JM, Gologan B, Cooks RG. *Science* 2004; **306**: 471.
2. Van Berkel GJ, Ford MJ, Deibel MA. *Anal. Chem.* 2005; **77**: 1207.
3. Weston DJ, Bateman R, Wilson ID, Wood TR, Creaser CS. *Anal. Chem.* 2005; **77**: 7572.
4. Chen H, Talaty NN, Takats Z, Cooks RG. *Anal. Chem.* 2005; **77**: 6915.
5. Takats Z, Cotte-Rodriguez I, Talaty N, Chen H, Cooks RG. *Chem. Commun.* 2005; 1950.
6. Cotte-Rodriguez I, Takats Z, Talaty N, Chen H, Cooks RG. *Anal. Chem.* 2005; **77**: 6755.
7. Williams JP, Scrivens JH. *Rapid Commun. Mass Spectrom.* 2005; **19**: 3643.
8. Wiseman JM, Puolitaival SM, Takats Z, Cooks RG, Caprioli RM. *Angew. Chem. Int. Ed. Engl.* 2005; **44**: 7094.
9. Talaty N, Takats Z, Cooks RG. *Analyst* 2005; **130**: 1624.
10. Leuthold LA, Mandscheff JF, Fathi M, Giroud C, Augsburg M, Varesio E, Hopfgartner G. *Rapid Commun. Mass Spectrom.* 2006; **20**: 103.
11. Rodriguez-Cruz SE. *Rapid Commun. Mass Spectrom.* 2006; **20**: 53.
12. McEwen CN, McKay RG, Larsen BS. *Anal. Chem.* 2005; **77**: 7826.
13. Cody RB, Laramée JA, Durst HD. *Anal. Chem.* 2005; **77**: 2297.
14. Tuomi T, Johnsson T, Reijula K. *Clin. Chem.* 1999; **45**: 2164.
15. Abdel-Hamid ME, Sharma D. *J. Liquid Chromatogr. Rel. Tech.* 2004; **27**: 641.
16. Byrd GD, Davis RA, Ogden MW. *J. Chromatogr. Sci.* 2005; **43**: 133.
17. Kearney G, Khan T, Castro-Perez J, Pugh J. *Proc. 49th ASMS Conf. Mass Spectrometry and Allied Topics*, Chicago, Illinois, 2001.

Collision-induced fragmentation pathways including odd-electron ion formation from desorption electrospray ionisation generated protonated and deprotonated drugs derived from tandem accurate mass spectrometry

Jonathan P. Williams,^{1*} Nico M. M. Nibbering,² Brian N. Green,³ Vibhuti J. Patel¹ and James H. Scrivens¹

¹ Department of Biological Sciences, University of Warwick, Gibbet Hill Rd, Coventry, CV4 7AL, UK

² Laser Centre and Chemistry Department, Vrije Universiteit, De Boelelaan 1083, 1081 HV Amsterdam, The Netherlands

³ Waters MS Technologies Centre, Micromass UK Ltd., Atlas Park, Simonsway, Manchester, M22 5PP, UK

Received 8 May 2006; Accepted 28 June 2006

The rapid desorption electrospray ionisation (DESI) of some small molecules and their fragmentation using a triple-quadrupole and a hybrid quadrupole time-of-flight mass spectrometer (Q-ToF) have been investigated. Various scanning modes have been employed using the triple-quadrupole instrument to elucidate fragmentation pathways for the product ions observed in the collision-induced dissociation (CID) spectra. Together with accurate mass tandem mass spectrometry (MS/MS) measurements performed on the hybrid Q-ToF mass spectrometer, unequivocal product ion identification and fragmentation pathways were determined for deprotonated metoclopramide and protonated aspirin, caffeine and nicotine. Ion structures and fragmentation pathway mechanisms have been proposed and compared with previously published data. The necessity for elevated resolution for the differentiation of isobaric ions are discussed. Copyright © 2006 John Wiley & Sons, Ltd.

KEYWORDS: desorption electrospray ionisation; mass spectrometry; tandem mass spectrometry; drugs; accurate mass measurement

INTRODUCTION

Mass spectrometry sampling using desorption electrospray ionisation (DESI) is gaining popularity for the analysis of a variety of surface types under ambient conditions.¹ In DESI, charged solvent droplets are sprayed towards the surface under investigation. These charged solvent droplets collide with the analytes on the surface which are desorbed, and ionisation is thought to occur by a droplet pick-up mechanism.² These secondary ions produced from the surface are then sampled by the mass spectrometer. For the analysis of solid pharmaceutical drug formulations, mass spectrometry sampling using DESI is extremely rapid since it requires little sample preparation. The DESI technique together with desorption atmospheric pressure chemical ionisation³ (DAPCI) and direct analysis in real time⁴ (DART) has been utilised for the analysis of some common drugs and samples of biological origin using mass spectrometry and tandem mass spectrometry accurate mass measurements.^{5,6} DESI of pharmaceutical

samples has previously been studied in an ion trap mass spectrometer⁷ and an ion-mobility mass spectrometer⁸ and analysed directly from thin layer chromatography (TLC) plates.⁹

Triple-quadrupole mass spectrometers are extremely useful as a research tool since they provide an efficient means of compound identification because of their unique scanning capabilities. The combination of the various scanning techniques employed in a triple-quadrupole instrument together with elevated-resolution time-of-flight (ToF) mass spectrometers, which provide accurate mass measurements allowing generation of elemental formulae, represents a powerful approach for unequivocal product ion identification and fragmentation pathways. The DESI technique is rapid since little or no sample preparation is required. The combination of both these mass spectrometry techniques provides high information content.

Experiments have been undertaken using various triple-quadrupole scanning techniques and/or elevated-resolution ToF accurate mass, with the focus on fully characterising fragmentation pathways for some low-molecular-weight species. Three compounds, caffeine, nicotine and metoclopramide, were of particular interest since they generated odd-electron

*Correspondence to: Jonathan P. Williams, Department of Biological Sciences, University of Warwick, Gibbet Hill Road, Coventry, CV4 7AL, UK. E-mail: j.p.williams@warwick.ac.uk

product ions from even-electron precursor ions via low-energy, collision-induced dissociation (CID). Protonated or deprotonated molecules do not usually lose a radical to form an odd-electron cation or anion since violation of the 'even-electron rule' occurs. However, fragmentations of many small even-electron ions that have not obeyed the rule have previously been reported from electron-impact (EI) and chemical-ionisation (CI) experiments.^{10–12} Using DESI, we demonstrate that under low-energy CID conditions, radical elimination can occur from even-electron ions especially if the generated product ions have a high stability: for example, they are aromatic. The necessity for elevated resolution for the differentiation of isobaric ions and accurate mass measurement for identification and confirmation has been demonstrated for an in-source CID-generated product ion of m/z 110 formed from protonated caffeine and paracetamol. The combination of a triple-quadrupole instrument, which can operate in full scan, product ion, precursor ion and neutral scan modes, together with a ToF instrument which provided elevated resolution and full scan sensitivity, will be discussed. The rapid analysis using DESI of a variety of compounds is shown, in which these experiments enhance the information content.

EXPERIMENTAL

Mass spectrometry

Experiments were performed in a triple-quadrupole (Q_1qQ_2 , where q is a hexapole) mass spectrometer (Quattro Ultima, Waters MS Technologies, Manchester, UK) and a hybrid quadrupole time-of-flight mass spectrometer (Q-ToF 1, Waters MS Technologies, Manchester, UK). Experimental details using the Q-ToF 1 have been previously reported.^{5,6}

The Q_1qQ_2 instrument was equipped with the standard Z-spray electrospray ion source and operated at a source and desolvation temperature of 110 °C and 250 °C, respectively. The desolvation gas was set to a flow rate of 200 l/h. The instrument was operated in the positive and negative modes with capillary voltages of 3.0 kV and –2.6 kV, respectively. Tandem mass spectrometry (MS/MS or MS^2) was carried out using argon as collision gas at a pressure of 2.5×10^{-3} mbar within the radio frequency (r.f.r.f.)-only hexapole collision cell. For DESI analysis, each tablet was broken, to expose an uncoated sample surface, before being held with tweezers, at an angle of approximately 45° to the solvent spray. The surface of the tablet was sprayed with a solution of acetonitrile/ H_2O + 0.2% formic acid at a flow rate of 10 μ l/min using a Harvard Apparatus (South Natick, MA, USA) Model 22 syringe pump. Mass spectra were acquired in the MCA mode at an acquisition rate of 1 spectrum per 8 s. Data acquisition and processing were carried out using MassLynx (V3.5).

Compounds investigated

The structures of the compounds investigated are listed in Fig. 1, together with the molecular formulae and molecular weights of their active ingredients. Diclofenac (50 mg) and naproxen (500 mg, Roche) were obtained by prescription. Ibuprofen (200 mg, Tesco, UK) and Anadin Extra (300 mg of aspirin, 200 mg of paracetamol and 45 mg of caffeine, Wyeth UK) were purchased from a local pharmacy. Nicotine was obtained directly from tobacco contained in a cigarette.

RESULTS AND DISCUSSION

The triple-quadrupole utilised in these experiments does not provide sufficient accurate mass measurements for the

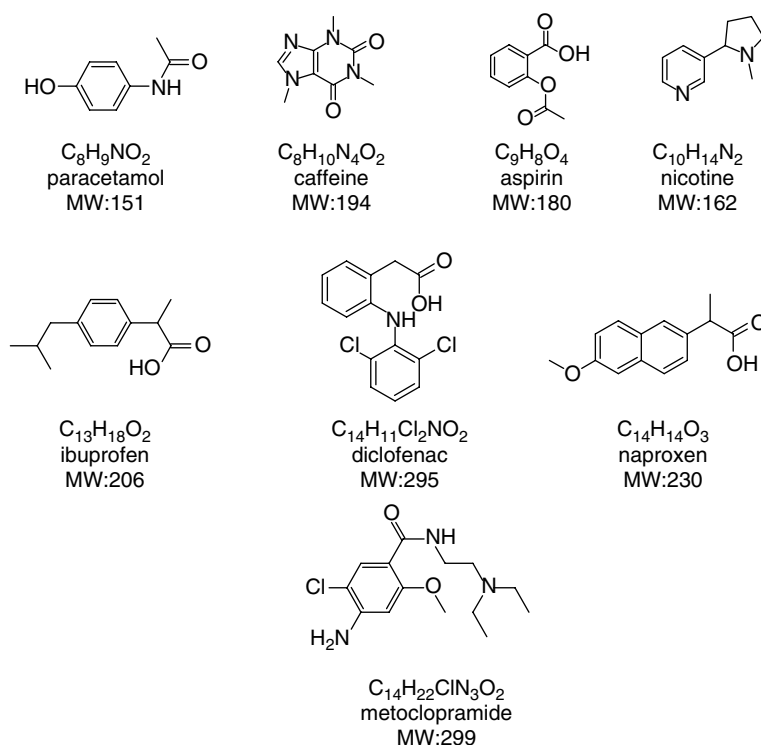


Figure 1. Structures, molecular formulae and molecular weights of the compounds studied.

unequivocal identification of the product ions resulting from CID. Accurate mass MS/MS performed in a Q-ToF-type instrument has been used to provide complementary structural information of the fragmentation pathways for many of the product ions observed in the CID spectra allowing fragmentation pathways to be determined. Ion structures and fragmentation pathway mechanisms have been proposed for deprotonated metoclopramide and protonated aspirin, caffeine and nicotine.

Negative ion DESI: product ion and neutral-loss scan modes

The rapid sampling of the DESI technique may be useful for metabolite identification during the drug discovery process. The detection of metabolites in urine has previously been reported by DESI.^{13,14} Signal suppression was observed for the analysis of metabolites in urine; however, the results demonstrate the applicability of the DESI approach as a rapid screen for metabolite identification. We have previously detected ibuprofen metabolites in human urine using DAPCI.⁶

Precursor and neutral-loss scan modes are efficient in detecting targeted or untargeted compounds, for example, metabolites of a drug. Sulfate, glutathione and glucuronide conjugates when fragmented by CID produce a characteristic loss of 80 Da (SO_3), 129 Da (loss of $\text{C}_5\text{H}_7\text{NO}_3$, the pyroglutamic acid moiety) and 176 Da ($\text{C}_6\text{H}_8\text{O}_6$), respectively, by positive ion mode analysis.^{15–17} MS/MS analysis in negative ion mode of sulphate conjugates leads to the formation of characteristic product ions of m/z 80 (SO_3^-) and m/z 97 (HSO_4^-). Glucuronide conjugates form a product ion of m/z 175 ($\text{C}_6\text{H}_7\text{O}_6^-$), the free glucuronide moiety, when

fragmented by CID. Q_2 is set to transmit only these low mass ions, and the precursor ions that fragment to these specific product ions are detected by scanning Q_1 for precursor ions.

Three non-steroidal anti-inflammatory drugs (NSAIDs) were chosen as model pharmaceutical tablets to evaluate the suitability of various scanning modes of a triple-quadrupole instrument using DESI for the detection of targeted group-specific losses. The three tablets, diclofenac, ibuprofen and naproxen, were sampled separately during the same acquisition for 8 s. Each of the compounds is known to deprotonate, owing to the acidic functional group in the molecules.

The mass spectrum obtained is shown in Fig. 2(A) over the m/z range 200–300. Intense deprotonated molecules together with some in-source CID fragment ions were observed in the mass spectrum. The peaks in the mass spectrum can be ascribed to $[\text{M}-\text{H}]^-$ from ibuprofen of m/z 205, the base peak $[\text{M}-\text{H}]^-$ from naproxen of m/z 229 and $[\text{M}-\text{H}]^-$ from diclofenac of m/z 294. The peak at m/z 250 corresponds to loss of CO_2 from the deprotonated molecule of diclofenac. The product ion mass spectrum generated for each of the deprotonated NSAID molecules is shown in Fig. 2(B). Deprotonated ibuprofen and diclofenac produced a single product ion under the low collision energy used during the experiment of m/z 161 and m/z 250, respectively. Naproxen generated two product ions of m/z 185 and m/z 170. It can be seen that all deprotonated molecules fragment with a characteristic loss of CO_2 (44 Da). Figure 2(C) shows the neutral-loss mass spectrum obtained by sampling the three NSAIDs during the same acquisition. As expected, the three deprotonated molecules of m/z 205, 229 and 294 were detected. Q_1 and Q_2 were scanned together with a

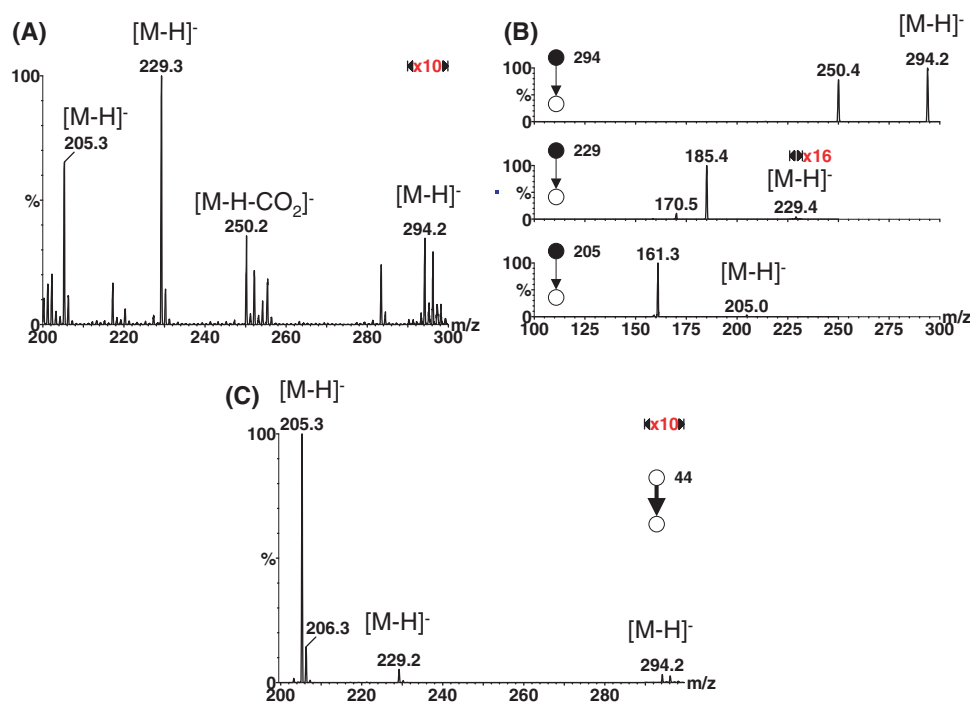


Figure 2. (A) Negative ion DESI-MS spectrum of ibuprofen, naproxen and diclofenac, (B) negative ion DESI-MS/MS spectrum of $[\text{M}-\text{H}]^-$ of m/z 205 from ibuprofen, $[\text{M}-\text{H}]^-$ of m/z 229 from naproxen and $[\text{M}-\text{H}]^-$ of m/z 294 from diclofenac, and (C) negative ion DESI-MS/MS neutral-loss spectrum in which both quadrupoles were scanned with a pre-defined mass difference of 44 Da, corresponding to CO_2 loss.

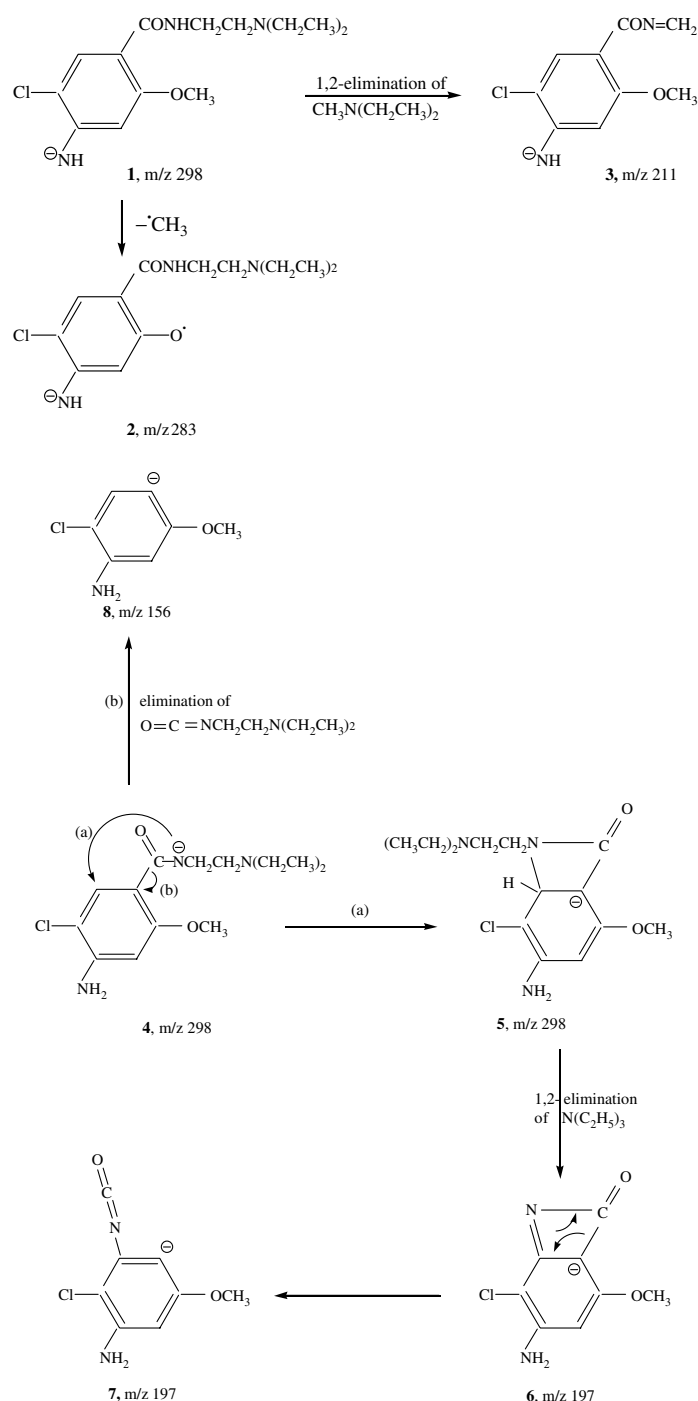
pre-defined mass difference of 44 Da. We believe this to be the first report combining rapid neutral-loss scanning methods and DESI for targeted group specific losses.

Accurate mass MS/MS of deprotonated metoclopramide, $[M-H]^-$ of m/z 298, led to the formation of four intense product ions as shown in Fig. 3. The ion of m/z 283 is of particular interest since it is a delocalised aromatic odd-electron anion formed through the loss of a methyl radical from deprotonated metoclopramide. The proposed mechanism for the generation of this product ion, together with the product ions of m/z 211, 197 and 156, is shown in Scheme 1. The formation of the ion with m/z 283 occurs via the route $1 \rightarrow 2$ and the ion with m/z 211 via route

$1 \rightarrow 3$, both starting from the deprotonated amino $[M-H]^-$ species. Possible mechanisms for the route of formation have been suggested for the ions **8** with m/z 156 and **6** or **7** with m/z 197. These ions are suggested to be generated from the metoclopramide molecules, which have been deprotonated at the amide nitrogen. Ion **7** is more likely than ion **6** because it does not have the strained four-membered ring fused with the aromatic ring as ion **6** has.

Positive ion DESI: product ion and precursor ion scan modes

Q-ToF accurate mass MS/MS experiments were carried out and combined with precursor ion tandem quadrupole



Scheme 1. The proposed fragmentation pathways of $[M-H]^-$ of metoclopramide by means of DESI-MS/MS.

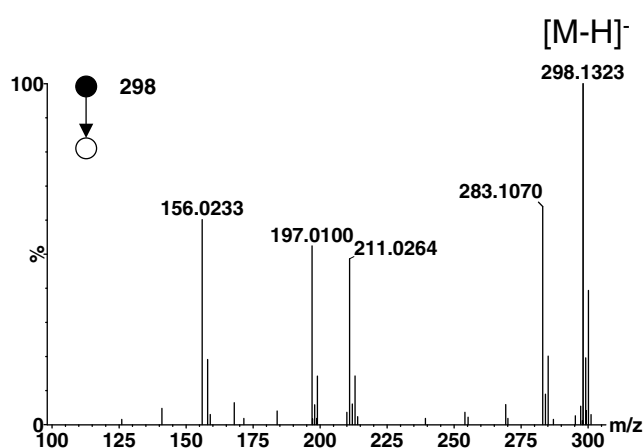


Figure 3. Negative ion DESI-MS/MS spectrum of $[M-H]^-$ of m/z 298 from metoclopramide.

experiments to determine unequivocal fragmentation pathways for protonated aspirin, caffeine, paracetamol and nicotine.

A solid Anadin Extra tablet, which contained 45 mg of the active ingredient caffeine, 200 mg of paracetamol and 300 mg of aspirin, was analysed using DESI. The mass spectrum obtained for Anadin Extra is in Fig. 4(A) and represents two 8 s scans obtained in the tandem quadrupole instrument. The protonated molecules for each of the active ingredients, paracetamol of m/z 152 and caffeine of m/z 195, were observed in the mass spectrum but the protonated aspirin molecule was absent. The sodiated molecule, $[M+Na]^+$ for the active ingredient aspirin was observed at m/z 203. The base peak in the mass spectrum is at m/z 195, the protonated molecule of caffeine. Other ions detected at m/z 163, 121 and m/z 110 have been previously characterised by accurate mass MS/MS.⁶ Accurate mass confirmed the ion of m/z 163 to be

$[MH-H_2O]^+$ from aspirin. The ions at m/z 110 and m/z 121 are products of in-source CID. We have previously shown that the ion at m/z 110 is a fragment produced by loss of CO from a product ion of m/z 138 from protonated caffeine and by loss of ketene from protonated paracetamol with the elemental formulae of $C_5H_8N_3$ and C_6H_8NO , respectively.^{5,6} The quadrupole instrument does not have the resolving power to distinguish between these two fragment ions since a resolution of ~ 9800 full width at half-maximum (FWHM) is required.

A precursor ion scan of m/z 110 is shown in Fig. 4(B). The optimum intensity of the ion of m/z 110 was first reached by increasing the in-source declustering voltage. The collision energy was optimised manually so that the yield of the m/z 110 ion generated within the collision cell from all precursors was maximised. The data confirms that the ion of m/z 110 is formed by loss of ketene (42 Da) from the $[M+H]^+$ of paracetamol at m/z 152. The data also confirms that the ion of m/z 110 is formed by loss of CO from m/z 138, which is subsequently formed by loss of methyl isocyanate (57 Da) from $[M+H]^+$ of caffeine at m/z 195.

A precursor ion scan of m/z 121 is shown in Fig. 4(C). The data confirms that the ion of m/z 121 is formed from m/z 139 and m/z 163 from aspirin. Accurate mass measurement of m/z 139.0411 suggested a formula of $C_7H_7O_3$, loss of ketene from the protonated aspirin molecule $[MH-CH_2=CO]^+$. Accurate mass measurement of m/z 149.0243 suggested a formula of $C_8H_5O_3$. This ion probably arises from the dissociation of an unknown phthalate: a ubiquitous contaminant often found in ESI mass spectra, which further dissociates via loss of CO to form m/z 121. Accurate mass measurement of m/z 163.0391 suggested a formula of $C_9H_7O_3$, loss of H_2O from the protonated molecule $[MH-H_2O]^+$. The MS/MS spectrum of protonated aspirin (m/z 181, data not shown) shows $C_2H_4O_2$ elimination

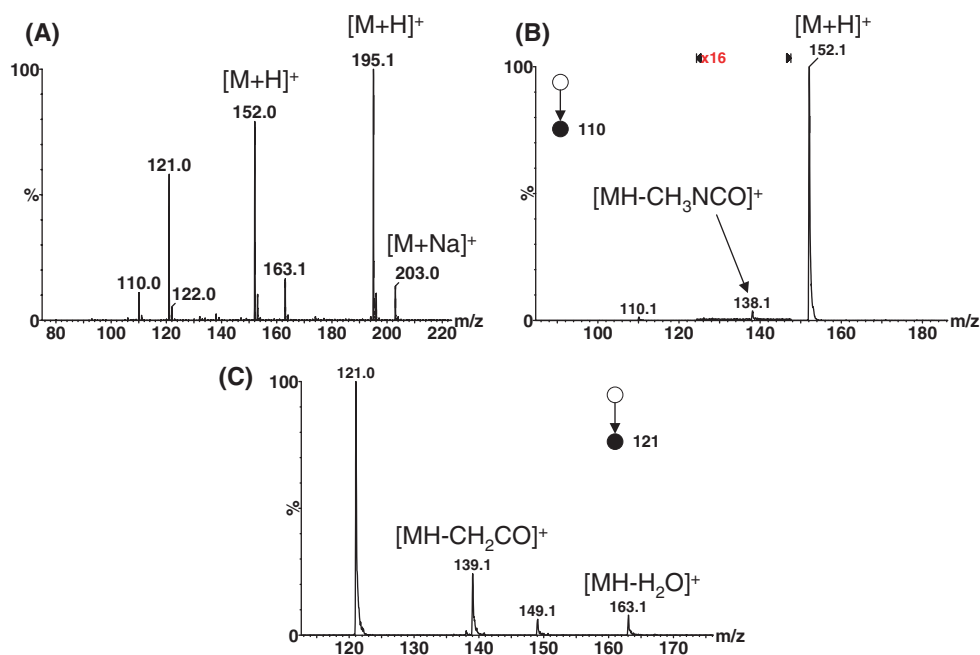


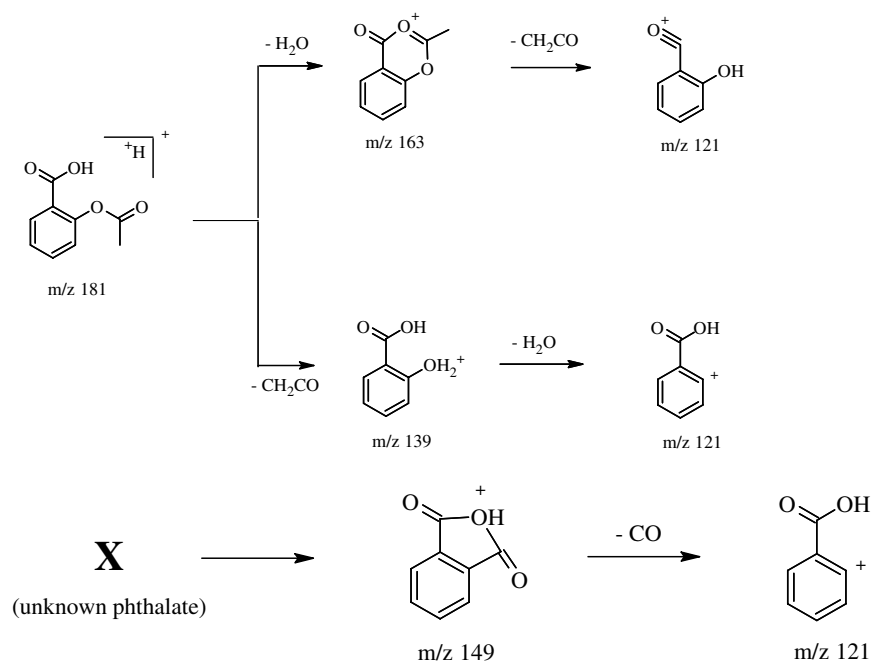
Figure 4. (A) Positive ion DESI-MS spectrum of Anadin Extra, (B) precursor ion spectrum of m/z 110 from the mass spectrum of Anadin Extra, and (C) precursor ion spectrum of m/z 121 from the mass spectrum of Anadin Extra.

$[\text{MH}-\text{HOCOCH}_3]^+$ from the side chain substituent. These fragmentation pathways are summarised in Scheme 2, and were generated from the data provided by the precursor ion spectrum together with accurate mass MS/MS. The ortho-hydroxybenzoyl cation will be energetically preferred because of the delocalisation of the positive charge over the aromatic ring and the hydroxyl group.

DESI accurate mass MS/MS of $[\text{M} + \text{H}]^+$ of m/z 195 from caffeine in Anadin Extra generated three product ions of m/z 138, 123 and 110 with sufficient intensity for the exact mass and hence generation of an elemental formula, shown in Fig. 5(A). The ion of m/z 138 is formed through loss of methyl isocyanate from protonated caffeine and m/z 110 is formed through loss of carbon monoxide from m/z 138. An odd-electron product ion of m/z 123 with elemental formula $\text{C}_5\text{H}_5\text{N}_3\text{O}$ was detected but its formation was unknown. Radical-induced bond cleavages of protonated molecules (even-electron ions) generated by DESI are unusual since molecules are generally eliminated from even-electron species. The aim was to determine which precursor ion generated the product ion of m/z 123 from

MS/MS experiments on protonated caffeine. A precursor ion mass spectrum of m/z 123 is shown in Fig. 5(B). The proposed mechanism for generation of this product ion together with the product ions of m/z 138 and m/z 110 is shown in Scheme 3. The data together with accurate mass MS/MS confirms that the ion of m/z 123 is formed from the loss of a methyl radical probably from the ring nitrogen atom of ion 2 of m/z 138 in Scheme 3. Methyl radical loss by homolytic cleavage of the N–C bond generates a radical that can be delocalised over the other nitrogen atoms and the carbon atom of the protonated carbonyl group of the resulting radical cation resembling an aromatic system. The routes of product ion formation for m/z 138 and m/z 110 from caffeine are $1 \rightarrow 2 \rightarrow 3 \rightarrow 4$. The route of formation of the odd-electron ion of m/z 123 is $1 \rightarrow 2 \rightarrow 5$ as shown in Scheme 3.

The DESI MS/MS spectrum of nicotine is shown in Fig. 6(A). Fragmentation pathways for protonated nicotine were previously described in part using the precursor ion scan mode in a tandem quadrupole instrument.¹⁸ During this study, accurate mass MS/MS data together with



Scheme 2. The proposed fragmentation pathways of $[\text{M} + \text{H}]^+$ of aspirin by means of DESI-MS/MS.

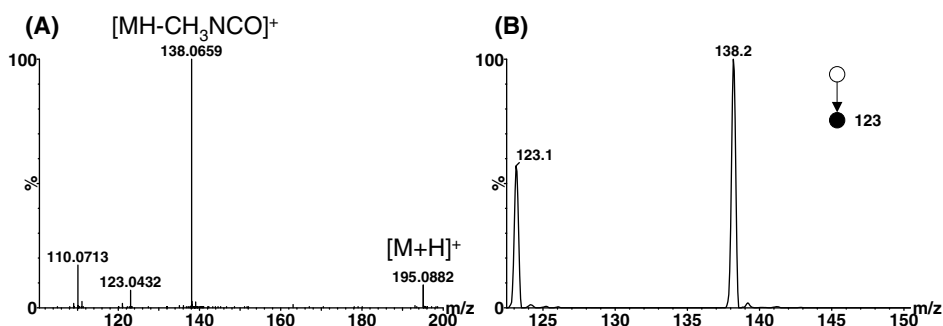


Figure 5. (A) Positive ion DESI-MS/MS spectrum of $[\text{M} + \text{H}]^+$ of m/z 195 from caffeine in Anadin Extra and (B) precursor ion spectrum of m/z 123 from the product ion spectrum of protonated caffeine.

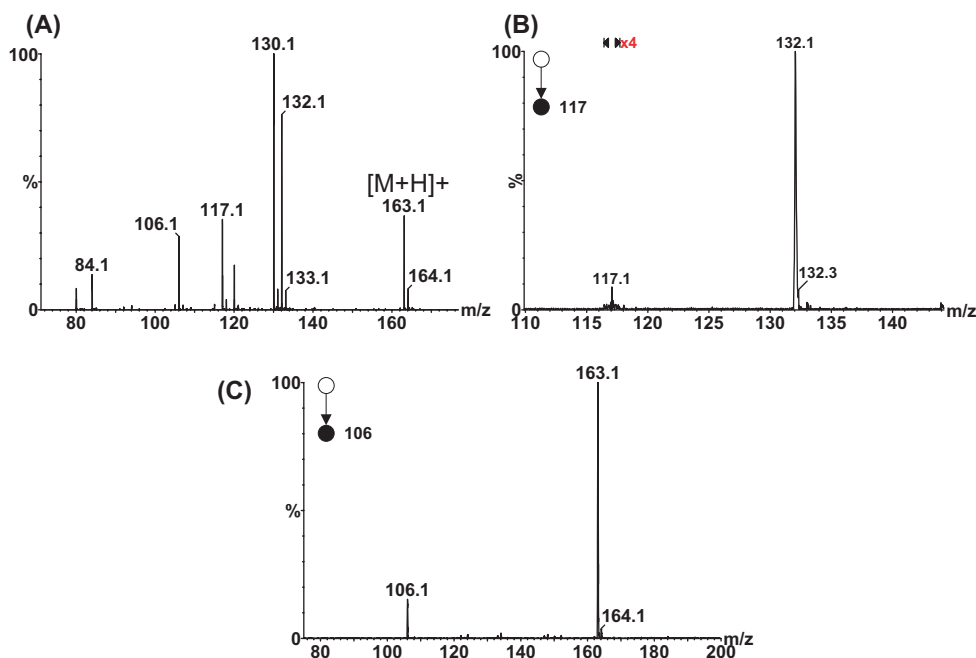
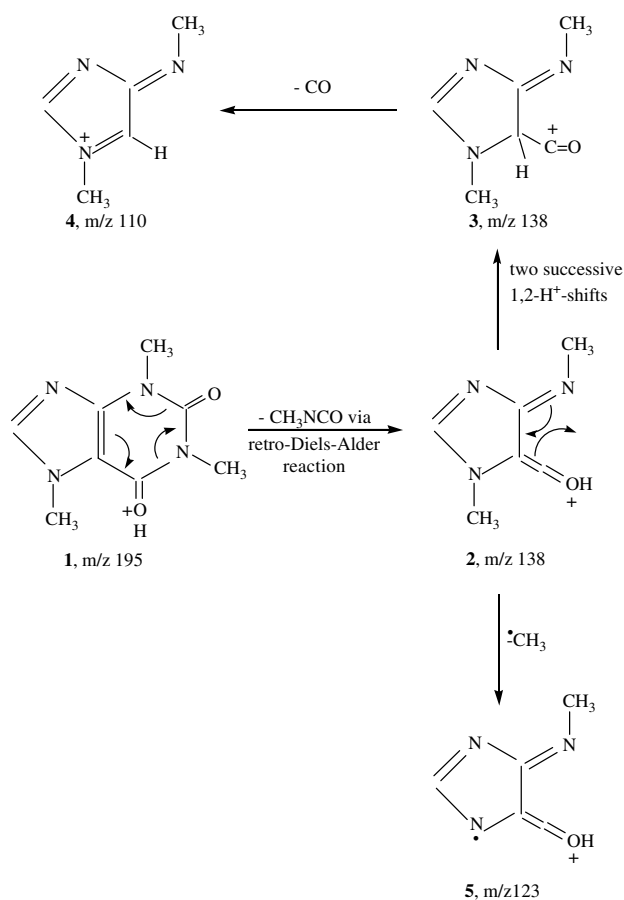


Figure 6. (A) Positive ion DESI-MS/MS spectrum of $[M + H]^+$ of m/z 163 from nicotine, (B) precursor ion spectrum of m/z 110 from the product ion mass spectrum of $[M + H]^+$ of m/z 163 from nicotine and (C) precursor ion spectrum of m/z 106 from the product ion mass spectrum of $[M + H]^+$ of m/z 163 from nicotine.



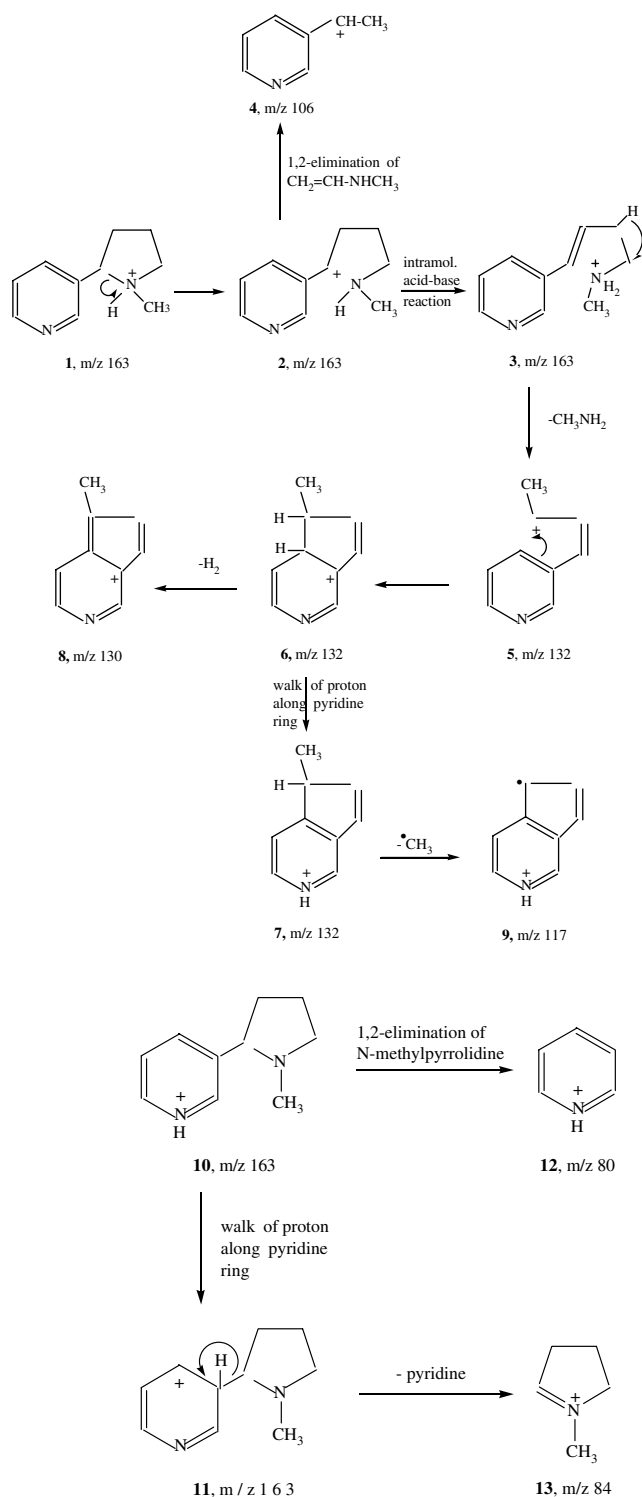
Scheme 3. The proposed fragmentation pathways of $[M + H]^+$ of caffeine by means of DESI-MS/MS.

detailed precursor ion scans of selected ions generated fragmentation pathways for protonated nicotine. Protonated

nicotine fragments under CID with the formation of an ion of m/z 132 through loss of 31 Da $[MH-CH_3NH_2]^+$ from the methyl-substituted pyrrolidine ring. The base peak in the MS/MS spectrum is m/z 130 $[MH-CH_3NH_2-H_2]^+$ formed by loss of H_2 from m/z 132. Accurate mass MS/MS of $[M + H]^+$ of nicotine generated an unusual odd-electron product ion of m/z 117. The aim was to determine which precursor ion generated this product ion from MS/MS experiments together with the precursor ions that fragment to form m/z 106. Ions of m/z 80 and m/z 84 are obtained from cleavage of the bond between the two rings.

A precursor ion scan of m/z 117 is shown in Fig. 6(B). The optimum intensity of the ion of m/z 117 was reached by increasing the in-source declustering voltage. The collision energy was optimised manually so that the yield of the m/z 117 ion generated within the collision cell from all precursors was maximised. The data, together with accurate mass MS/MS, confirms that the ion of m/z 117 is formed by the loss of a methyl radical from m/z 132 in nicotine $[MH-CH_3NH_2-\cdot CH_3]^+$. A precursor ion scan of m/z 106 is shown in Fig. 6(C). The data together with accurate mass MS/MS confirms that the ion of m/z 106 is formed solely from the loss of C_3H_7N from m/z 163 in nicotine $[MH-C_3H_7N]^+$, as previously proposed.¹⁸ These fragmentation pathways are summarised in Scheme 4, and were generated from the data provided by the precursor ion spectra together with accurate mass MS/MS. The ion structures proposed are significantly different from the previously proposed structures.^{6,18}

The routes of product ion formation for nicotine are: routes 1 \rightarrow 2 \rightarrow 4 for formation of m/z 106, 1 \rightarrow 2 \rightarrow 3 \rightarrow 5 for formation of m/z 132, 5 \rightarrow 6 \rightarrow 8 for formation of m/z 130 if 5 has enough energy and 5 \rightarrow 6 \rightarrow 7 \rightarrow 9 for formation of the odd-electron ion with m/z 117 if 5 has enough energy (in principle the carbenium ion centre in ion 5 could form



Scheme 4. The proposed fragmentation pathways of $[M + H]^+$ of nicotine by means of DESI-MS/MS.

also a C–C bond with carbon atom 2 of the pyridine ring, but this has been omitted in Scheme 4). Ion 9 with m/z 117 is energetically an attractive aromatic radical cation, and that is the reason why it can be generated from the even-electron ion 7 with m/z 132. The mechanistic routes of formation of the ions 12 with m/z 80 and 13 with m/z 84 are also included. Previous ion structures contained carbenium ion structures that are not in a potential energy minimum and with strained rings^{6,18} and are therefore unlikely to be formed.

It has been shown that utilising the DESI technique coupled with a triple-quadrupole instrument offers an efficient and quick method for gaining structural molecular information. Individual ions can be mass-selected and allowed to fragment through collisions with neutral gas atoms contained in a collision cell. The product ions generated can be detected; precursor ions that form fragments of a specific m/z value can be detected together with group-specific neutral-loss detection. Triple-quadrupole instrumentation provides nominal mass information, which is not sufficient for the provision of elemental formulae. It is possible to establish the elemental composition of an organic ion if its mass can be mass-measured with sufficient accuracy. For low-molecular-weight compounds the number of combinations is small or the composition may even be unique if the mass can be measured to within 5 ppm.

For instrumentation where MS/MS is not an option, for example, single-quadrupole and single ToF, fragment ions can be produced prior to entry into the mass spectrometer by increasing the voltage applied to the sampling cone. In-source CID has been utilised to the analysis of a Paracetamol Plus tablet which contains both paracetamol and caffeine, in a tandem quadrupole and a hybrid quadrupole-ToF instrument equipped with a 4 GHz time-to-digital converter (TDC) (Q-ToF Ultima, Waters MS Technologies) as an example of in-source CID. We have previously shown that the ion of m/z 110 is a fragment produced by CID of protonated caffeine and paracetamol with the elemental formulae of $C_5H_8N_3$ and C_6H_8NO , respectively. The quadrupole and single ToF instrument used in a previous study⁶ did not have the resolving power to distinguish between these two product ions and therefore had difficulty in its characterisation.

Figure 7(A) shows the in-source CID mass spectrum obtained in the quadrupole instrument for Paracetamol Plus. Figure 7(B) and (C) shows a comparison of the peak widths obtained for the product ion of m/z 110 obtained in the quadrupole and ToF instruments, respectively. The enhanced mass resolution of the hybrid Q-ToF (~10000 FWHM, in V-mode) aids exact mass measurement and thus facilitates assignment of an elemental formula, as shown in Fig. 7(D), for both product ions of m/z 110. Here, protonated paracetamol $[M + H]^+$ of m/z 152.0711 was used as the lock mass to correct for mass scale drift. Although the peaks are partially resolved, the ToF instrument has allowed unambiguous identification for these low mass ions separated by only 11.2 mDa. The ion of m/z 110.0624 is generated from protonated paracetamol and shows a mass error 1.8 mDa from the theoretical monoisotopic mass of 110.0606 Da. The ion of m/z 110.0701 is generated by loss of CO from m/z 138 from protonated caffeine and shows a mass error –1.7 mDa from the theoretical monoisotopic mass of 110.0718 Da.

CONCLUSIONS

The rapid DESI of some small molecules and their fragmentation using a triple-quadrupole and a hybrid Q-ToF has allowed for fragmentation pathways to be determined. Deprotonated metoclopramide and protonated caffeine and

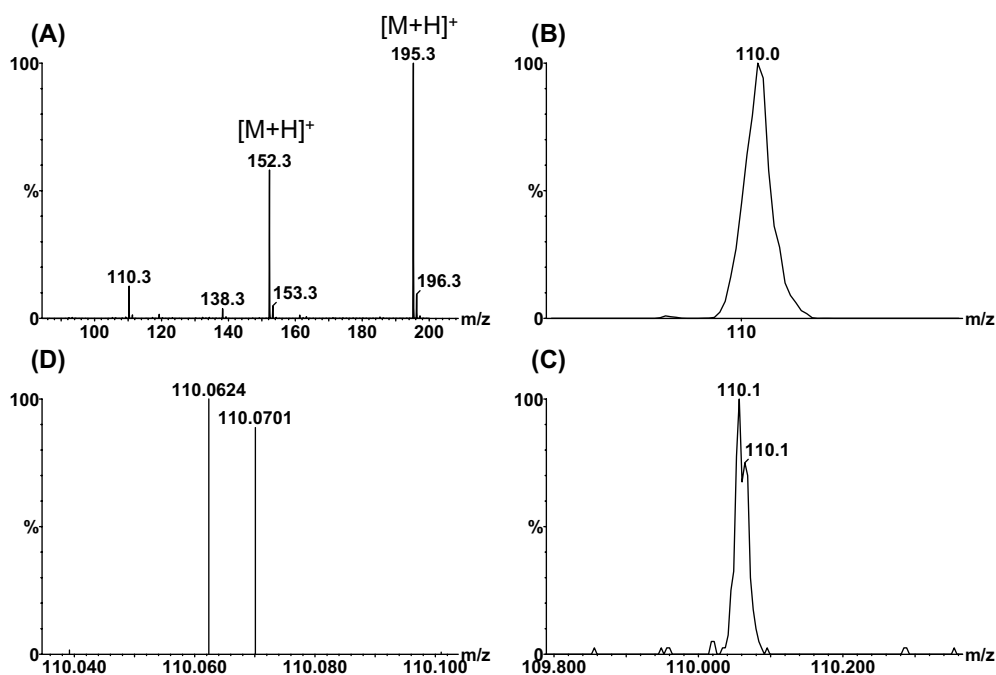


Figure 7. (A) Positive ion in-source CID DESI-MS spectrum of Anadin Extra, (B) expanded region of mass spectrum of fragment ion of m/z 110 from a quadrupole-type instrument, (C) expanded region of mass spectrum of fragment ion of m/z 110 from a ToF-type instrument and (D) accurate mass in-source CID of both fragment ions of m/z 110 from protonated paracetamol and caffeine from a ToF-type instrument.

nicotine have been shown to form odd-electron product ions upon low-energy CID by elimination of a methyl radical forming either a stable aromatic radical anion or cation. The formation of odd-electron product ion is unusual since even-electron ions generally produce even-electron product ions upon low-energy CID by elimination of a neutral.

Ion structures and ion formation mechanisms have been proposed for these low-molecular-weight compounds. Ion structures proposed for the product ions formed from CID of protonated nicotine are significantly different from previous proposals.^{6,18} The combination of the DESI technique with various scanning techniques employed in a triple-quadrupole instrument together with elevated resolution ToF mass spectrometry has been shown to represent a rapid and powerful analytical approach for unequivocal product ion identification and fragmentation pathways. The improved information content derived from ToF mass spectrometry has been demonstrated for the identification of isobaric ions to within 2 mDa. This could not have been achieved on instruments of low resolution.

The recently developed DESI approach has been shown to provide a method of obtaining rapid mass spectral data. The information content of the DESI experiment is limited, however, in most cases, to protonated and deprotonated molecules. The use of precursor, product and neutral-loss mass spectrometry experiments and ToF MS/MS accurate mass measurement enhances the information content of the DESI experiment and provides significantly greater selectivity. The experiments can be carried out on a time scale compatible with the DESI experiment.

REFERENCES

1. Takats Z, Wiseman JM, Gologan B, Cooks RG. Mass spectrometry sampling under ambient conditions with desorption electrospray ionization. *Science* 2004; **306**: 471.
2. Cooks RG, Ouyang Z, Takats Z, Wiseman JM. Detection technologies, ambient mass spectrometry. *Science* 2006; **311**: 1566.
3. Takats Z, Cotte-Rodriguez I, Talaty N, Chen H, Cooks RG. Direct, trace level detection of explosives on ambient surfaces by desorption electrospray ionization mass spectrometry. *Chem. Commun. (Camb.)* 2005; **15**: 1950.
4. Cody RB, Laramée JA, Durst HD. Versatile new ion source for the analysis of materials in open air under ambient conditions. *Anal. Chem.* 2005; **77**: 2297.
5. Williams JP, Scrivens JH. Rapid accurate mass desorption electrospray ionisation tandem mass spectrometry of pharmaceutical samples. *Rapid Commun. Mass Spectrom.* 2005; **19**: 3643.
6. Williams JP, Patel VJ, Holland R, Scrivens JH. The use of recently described ionisation techniques for the rapid analysis of some common drugs and samples of biological origin. *Rapid Commun. Mass Spectrom.* 2006; **20**: 1447.
7. Chen H, Talaty NN, Takats Z, Cooks RG. Desorption electrospray ionization mass spectrometry for high-throughput analysis of pharmaceutical samples in the ambient environment. *Anal. Chem.* 2005; **77**: 6915.
8. Weston DJ, Bateman R, Wilson ID, Wood TR, Creaser CS. Direct analysis of pharmaceutical drug formulations using ion mobility spectrometry/quadrupole-time-of-flight mass spectrometry combined with desorption electrospray ionization. *Anal. Chem.* 2005; **77**: 7572.
9. Van Berkel GJ, Ford MJ, Deibel MA. Thin-layer chromatography and mass spectrometry coupled using desorption electrospray ionization. *Anal. Chem.* 2005; **77**: 1207.
10. Karni M, Mandelbaum A. The even-electron rule. *Org. Mass Spectrom.* 1980; **15**: 53.
11. Sigsby ML, Day RJ, Cooks RG. Fragmentation of even electron ions: protonated ketones and ethers. *Org. Mass Spectrom.* 1979; **14**: 273.

12. Sigsby ML, Day RJ, Cooks RG. Fragmentation of even electron ions: protonated amines and esters. *Org. Mass Spectrom.* 1979; **14**: 556.
13. Takats Z, Wiseman JM, Cooks RG. Ambient mass spectrometry using desorption electrospray ionization (DESI): instrumentation, mechanisms and applications in forensics, chemistry, and biology. *J. Mass Spectrom.* 2005; **40**: 1261.
14. Kauppila TJ, Wiseman JM, Ketola RA, Kotiaho T, Cooks RG, Kostianen R. Desorption electrospray ionization mass spectrometry for the analysis of pharmaceuticals and metabolites. *Rapid Commun. Mass Spectrom.* 2006; **20**: 387.
15. Castro-Perez J, Plumb R, Liang L, Yang E. A high-throughput liquid chromatography/tandem mass spectrometry method for screening glutathione conjugates using exact mass neutral loss acquisition. *Rapid Commun. Mass Spectrom.* 2005; **19**: 798.
16. Castro-Perez J, Major H, Hoyes J, Preece S. In *Proceedings of the American Society for Mass Spectrometry*, Chicago, 2001.
17. Kostianen R, Kotiaho T, Kuuranne T, Auriola S. Liquid chromatography/atmospheric pressure ionization-mass spectrometry in drug metabolism studies. *J. Mass Spectrom.* 2003; **38**: 357.
18. Byrd GD, Davis RA, Ogden MW. A rapid LC-MS-MS method for the determination of nicotine and cotinine in serum and saliva samples from smokers: validation and comparison with a radioimmunoassay method. *J. Chromatogr. Sci.* 2005; **43**: 133.

Polarity Switching Accurate Mass Measurement of Pharmaceutical Samples Using Desorption Electrospray Ionization and a Dual Ion Source Interfaced to an Orthogonal Acceleration Time-of-Flight Mass Spectrometer

Jonathan P. Williams,*† Richard Lock,‡ Vibhuti J. Patel,† and James H. Scrivens†

Department of Biological Sciences, University of Warwick, Gibbet Hill Road, Coventry, CV4 7AL, UK, and Waters MS Technologies Centre, Micromass UK Ltd., Atlas Park, Simonsway, Manchester, M22 5PP, UK

A novel approach for the rapid, accurate mass analysis of pharmaceutical solid, liquid, and cream formulations using desorption electrospray ionization (DESI) is described. The method is based on polarity switching and real-time accurate mass measurement in an orthogonal acceleration time-of-flight mass spectrometer fitted with a dual-inlet electrospray ion source. Infusion of a reference compound into one inlet provides a single-point “lock mass” for accurate mass measurement. The other inlet sprays solvent at the sample being investigated using DESI. Minimal sample preparation was required. Results demonstrate the ability to acquire simultaneously positive and negative accurate mass DESI data within the same acquisition, thus negating the need for repeat analysis in each ion mode. In this paper, drugs that preferentially ionize in a particular mode depending on their physicochemical properties are presented. Mass accuracy to within 2 mTh was obtained for all drugs sampled.

The rapid ambient mass spectrometry sampling of pharmaceutical formulations using desorption electrospray ionization¹ (DESI) is analytically attractive since little or no sample preparation is required. In DESI, charged solvent droplets are sprayed toward the pharmaceutical being analyzed. These charged solvent droplets collide with the surface releasing desorbed secondary gas-phase analyte ions, which are sampled by the mass spectrometer (MS).² There have been a number of recent reports exploiting the DESI technique for the analysis of active ingredients in drug formulations^{3–8} together with high-throughput analysis

of selected tablets.⁹ The rapid identification of active ingredients in pharmaceutical formulations by accurate mass MS and MS/MS can be obtained in seconds using the DESI technique.^{4,5} DESI may therefore be a convenient method for high-throughput assays within pharmaceutical method development, without requiring separation by liquid chromatography (liquid chromatography) prior to mass spectrometric detection, as is often used. Although some sensitivity would be expected to be sacrificed using DESI-MS compared to electrospray ionization (ESI) LC–MS, analysis times will be shorter. Shorter analysis time would therefore lead to throughput enhancement.

For the rapid analysis of pharmaceutical drug formulations, it would be useful to extend the sampling approach using DESI to simultaneously detect both basic and acidic active ingredients that easily protonate and deprotonate in a particular formulation during the same acquisition. Many pharmaceuticals contain more than one active ingredient and fall into this category. Data for such tablets that contain a range of active ingredients will be presented. Pharmaceutical ingredients have diverse chemical properties, which require different polarity MS detection modes. The present study extends the previous approach⁹ by using DESI and polarity switching real-time accurate mass measurement. This negates the need for time-consuming repeat analysis in each ion mode and can be of benefit to compound screening applications where the ionization mode for analysis is not known. The time-of-flight (TOF) analyzer used in this present study made use of dynamic range enhancement (DRE) technology,¹⁰ facilitating routine accurate mass measurement of pharmaceutical formulations. Polarity switching accurate mass measurement using DESI will be presented for pharmaceutical solid, liquid, and cream formulations.

* To whom correspondence should be addressed. Fax: +44 2476 523701. Tel: +44 2476 528379. E-mail: j.p.williams@warwick.ac.uk.

† University of Warwick.

‡ Micromass UK Ltd.

- (1) Takats, Z.; Wiseman, J. M.; Gologan, B.; Cooks, R. G. *Science* **2004**, *306*, 471.
- (2) Cooks, R. G.; Ouyang, Z.; Takats, Z.; Wiseman, J. M. *Science* **2006**, *311*, 1966.
- (3) Rodriguez-Cruz, S. E. *Rapid Commun. Mass Spectrom.* **2006**, *20*, 53.
- (4) Williams, J. P.; Scrivens, J. H. *Rapid Commun. Mass Spectrom.* **2005**, *19*, 3643.
- (5) Williams, J. P.; Patel, V. J.; Holland, R.; Scrivens, J. H. *Rapid Commun. Mass Spectrom.* **2006**, *20*, 1447.

- (6) Weston, D. J.; Bateman, R.; Wilson, I. D.; Wood, T. R.; Creaser, C. S. *Anal. Chem.* **2005**, *77*, 7572.
- (7) Leuthold, L. A.; Mandscheff, J. F.; Fathi, M.; Giroud, C.; Augsburg, M.; Varesio, E.; Hopfgartner, G. *Rapid Commun. Mass Spectrom.* **2006**, *20*, 103.
- (8) Van Berkel, G. J.; Ford, M. J.; Deibel, M. A. *Anal. Chem.* **2005**, *77*, 1207.
- (9) Chen, H.; Talaty, N. N.; Takats, Z.; Cooks, R. G. *Anal. Chem.* **2005**, *77*, 6915.
- (10) Green, M. R.; Jackson, M.; Hoyes, J.; Bateman, R. H.; Newton, A.; Platt, S. J. *Proceedings of the American Society for Mass Spectrometry*, Chicago, IL, 2001.

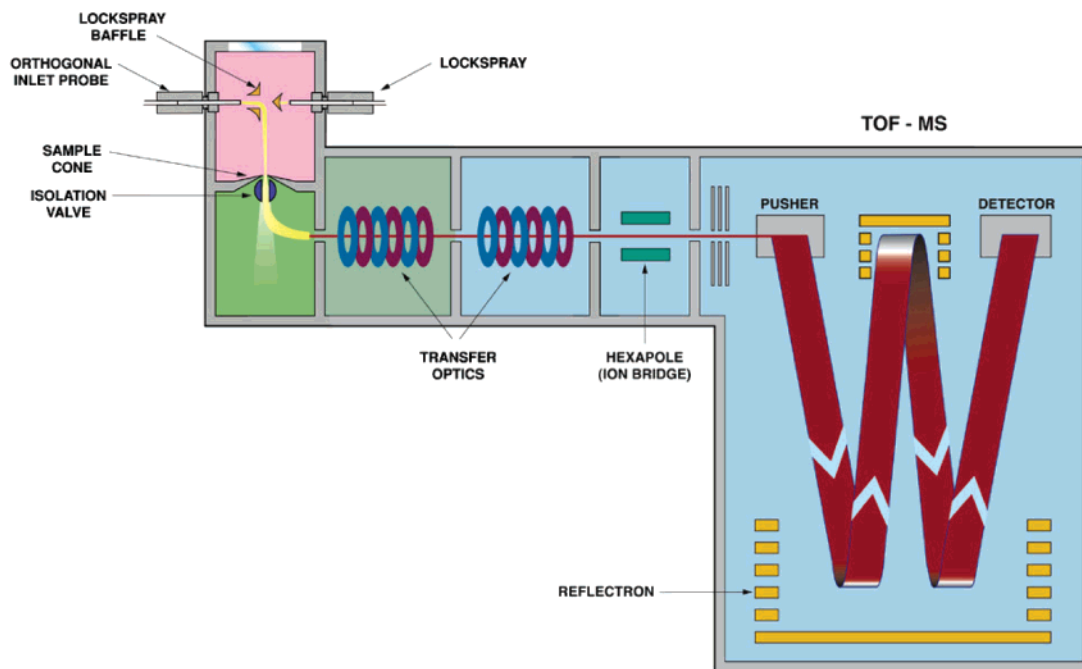


Figure 1. Schematic diagram of the LCT Premier.

Accurate mass and accurate mass in-source collision-induced data to within 2 mTh were obtained in <5 s for the drug formulations investigated. To the best of our knowledge, this is the first report combining real-time accurate mass measurement, polarity switching, and DESI.

EXPERIMENTAL SECTION

The tablets, liquid, and cream formulations investigated were as follows: erythromycin tablets (500 mg, Abbot) and chloramphenicol eye ointment (1% w/w/ 4 g, IVAX) obtained by prescription. Anadin Extra (300 mg of aspirin, 200 mg of paracetamol, and 45 mg of caffeine, Wyeth, UK) and Night Nurse liquid medicine (1000 mg/20 mL paracetamol, 20 mg/20 mL promethazine hydrochloride, and 15 mg/20 mL dextromethorphan hydrobromide, GSK) were purchased without prescription from a pharmacy.

Experiments were carried out using a Waters MS Technologies (Manchester, UK) LCT Premier orthogonal acceleration time-of-flight mass spectrometer (oa-TOFMS) fitted with a two-way electrospray ionization source^{11,12} (LockSpray). The instrument was operated in positive and negative modes with a capillary voltage of 3.0 and 2.5 kV, respectively. The ion source block and nitrogen desolvation gas temperature were set to 120 and 350 °C, respectively, and the desolvation gas was set to a flow rate of 300 L/h. The TOF mass analyzer was tuned in W-optic mode for an operating resolution of 10 000 (fwhm) (Figure 1). Mass spectra were acquired at an acquisition rate of 1 spectrum/150 ms with an interscan delay of 50 ms in centroid mode.

For DESI sample analysis, each tablet was broken, to expose an uncoated sample surface, before being held with tweezers, at an angle of ~45° to the solvent spray emanating from inlet one.

The source was open to the laboratory atmosphere to allow manual introduction of the samples. Approximately 100 mg of the ointment or 50 μ L of the liquid medicine was applied to the surface of a piece of matt-finished cardboard (1 cm \times 2 cm) and held in the same position as the solid tablets. The surface of the tablet or card was then sprayed with a solution of acetonitrile/H₂O + 0.1% formic acid at a flow rate of 10 μ L/min., using a Waters Acquity binary Solvent Manager pump. Accurate mass measurement was provided by infusing leucine enkephalin ($[M + H]^+ = 556.2771$ and $[M - H]^- = 554.2615$) into inlet two of the dual source and used as a single-point "lock mass" against which any subsequently acquired mass spectra were mass measured in real time. The lock mass was infused at a flow rate of 5 μ L/min at a concentration of 200 pg/ μ L (acetonitrile/H₂O + 0.1% formic acid) using a single-piston Waters Reagent Manager pump. No significant modification of solvents, buffers, and pH was carried out.

RESULTS AND DISCUSSION

Accurate Mass Protocol. We have previously shown the usefulness of mass spectrometry and tandem mass spectrometry accurate mass for the generation of empirical formulas of small molecules using DESI coupled with a hybrid quadrupole-time-of-flight mass spectrometer.^{4,5} Instrumental mass drift was corrected for by a single internal reference lock mass in MS and MS/MS mode. For MS experiments, a reference standard is sampled during the acquisition after analysis of the surface under investigation. For MS/MS experiments, the precursor ion selected for collision-induced dissociation (CID) provided the internal reference lock mass.

During this study, we have made use of a dual ESI source interfaced to a single TOF mass analyzer. The interface provides a simpler method of obtaining accurate mass measurements through infusion of a reference compound into a separate electrospray inlet rather than sampling a reference standard during the acquisition after analysis of the surface under investiga-

(11) Eckers, C.; Wolff, J. C.; Haskins, N. J.; Sage, A. B.; Giles, K.; Bateman, R. *Anal. Chem.* **2000**, *72*, 3683.

(12) Wolff, J. C.; Eckers, C.; Sage, A. B.; Giles, K.; Bateman, R. *Anal. Chem.* **2001**, *73*, 2605.

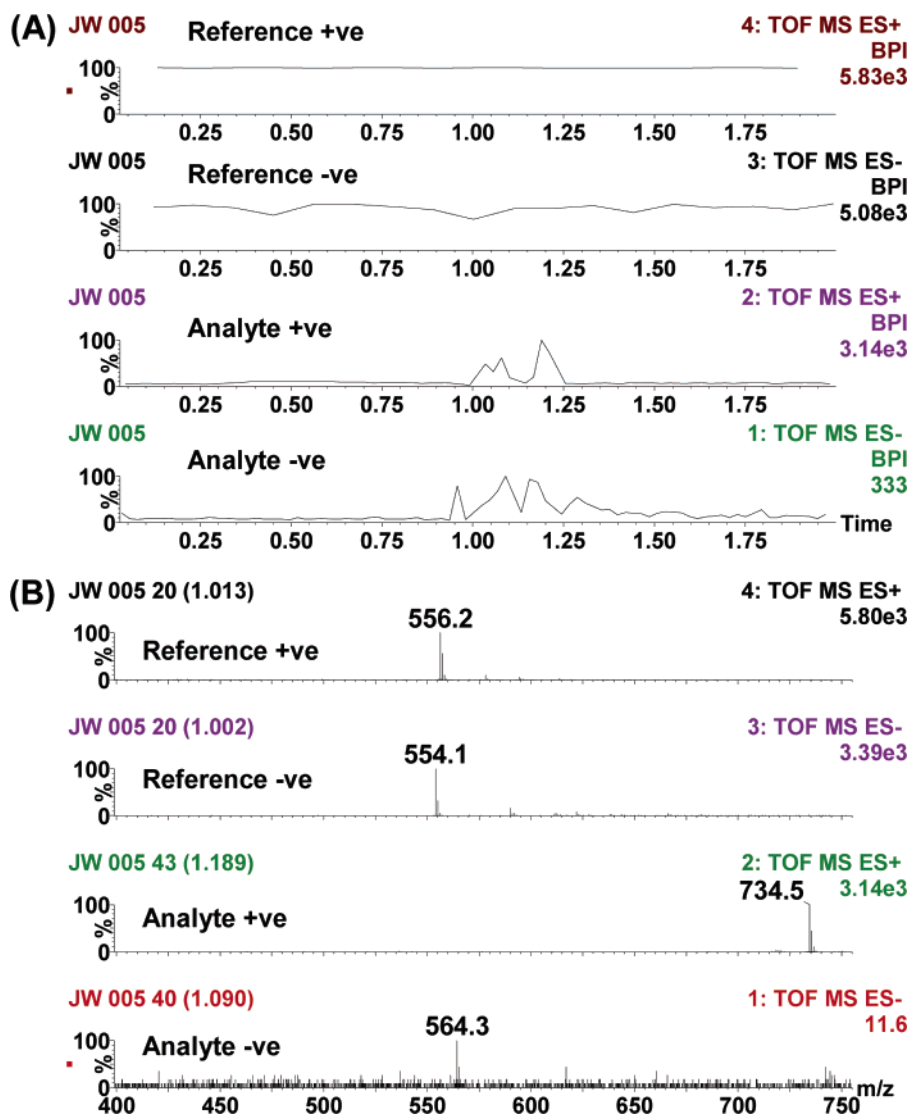


Figure 2. (A) Ion chromatograms obtained from a single acquisition of erythromycin and leucine enkephalin reference using DESI and polarity switching. (B) Mass spectra obtained from time of DESI analysis of erythromycin and leucine enkephalin at ~ 1 min.

tion used with a single probe.⁴ Using the dual probe interface, the spray from each capillary tip is sampled independently through use of a rotating baffle, which is externally driven by a stepper motor. The sprays are admitted to the sampling cone region of the LCT through an aperture in the baffle. The two positions of the sampling rotor are indexed, which allows data from each of the two sprays to be stored into separate functions within a single data file during acquisition.

Sampling of pharmaceutical drug formulations using DESI combined with a TOF mass spectrometer incorporating a time-to-digital (TDC) based detection system generates very high ion counts since the active ingredient can typically range from 1 to 1000 mg. Mass measurement accuracy can be compromised at high ion counts due to TDC "dead-time" saturation. TDC dead time is the time after an ion is recorded within which the TDC is not able to record another ion arriving at the detector. Dead-time saturation occurs when more than one ion arrives at the detector within the TDC dead-time period. Under multiple ion arrival conditions, some of the ions are not recorded resulting in a nonlinear intensity response and a shift in measured m/z value

toward lower m/z . A software-based correction method, based on statistical calculation, is employed to correct for dead-time saturation effects over a limited dynamic range. Previously it was necessary to manually select an area of the total ion chromatogram (TIC) where this correction software could be successfully employed to display data with accurate mass measurement and intensity values.^{4,5}

During this study, we have successfully used a single TOF instrument incorporating DRE¹⁰ technology together with polarity switching, for routine accurate mass measurement of DESI-generated ions. A full description of the mode of operation of DRE technology is given elsewhere^{13,14} and so will only be described briefly here. In principle, the DRE technique employs modified transfer optics capable of reducing the transmission of the ion beam into the orthogonal sampling (pushout) region by means

(13) Howes, K.; Wildgoose, J.; Green, M.; McCullagh, M. *Proceedings of the American Society for Mass Spectrometry*, Montreal, Canada, 2003. Available at <http://www.waters.com/WATERSDIVISION/pdfs/720000648EN.pdf>.

(14) McCullagh, M. *Application Note*, Waters UK, 2005. Available at <http://www.waters.com/WATERSDIVISION/pdfs/720001098EN.pdf>.

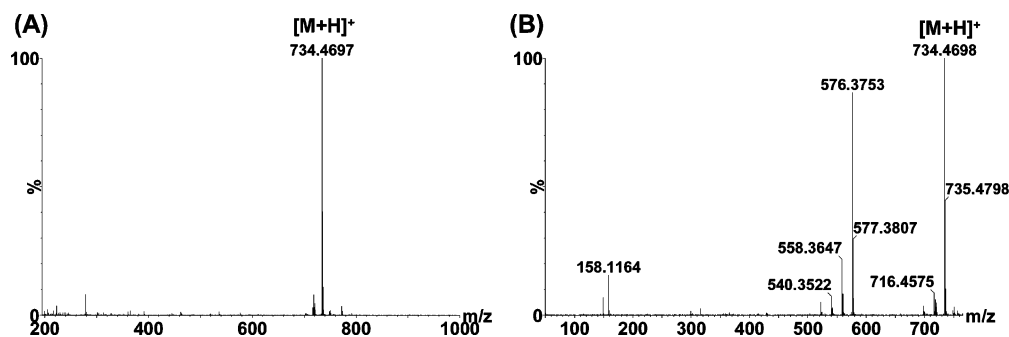


Figure 3. (A) Accurate mass spectrum obtained for protonated erythromycin. (B) Accurate mass in-source CID mass spectrum of protonated erythromycin.

of an applied voltage. These optics are arranged to defocus the ion beam in the Z direction (normal to the plane of the time-of-flight mass spectrometer), resulting in minimal effect on resolution and mass measurement. During acquisition, the transmission of the instrument is switched repeatedly between two calibrated values resulting in ion currents of low and high intensity. Rapidly switching between the two values allows saturated high-intensity data to be replaced by low-intensity data. The ion transmission ratio between switching is multiplied by the low-intensity ion signal to give an unsaturated high-intensity ion signal allowing mass measurements to be made in “real time” in regions of the TIC where previously detector saturation would have occurred.¹⁴ When TDC saturation effects limit measurements for the data recorded using full transmission, data from the low transmission cycle can be used.

During any one acquisition utilizing polarity switching, Lock-Spray and DESI data from each of the two sprays are sampled independently and stored as separate functions within the same data file. During an acquisition, four separate functions are acquired, a reference and an analyte function in positive and negative ion modes. The data system acquires data points for each of these functions in quick succession by rapidly switching from negative to positive and from analyte sprayer to reference sprayer in a repetitive cycle, resulting in a data file with four simultaneously acquired chromatograms. To demonstrate the ability to acquire simultaneously positive and negative accurate mass DESI data within the same acquisition, thus negating the need for repeat analysis in each ion mode, samples were chosen that preferentially ionize in a particular mode depending on their chemical structure. This type of analysis using DESI provides a rapid screen combining detection and characterization. It provides significant potential for the rapid identification of unknown materials.

Figure 2A shows the four base peak ion chromatograms obtained from sampling both an erythromycin tablet using DESI and leucine enkephalin (reference lock mass) using conventional ESI. (The polarity function is shown for each chromatogram in this example only and annotated in the top right corner of each chromatogram.) In this example, the erythromycin tablet was exposed to the solvent spray twice for 5 s. The point of ionization can be seen in function two in positive mode at ~ 1 min. The relative signal intensities of the active ingredient are slightly different due to the nonreproducibility of manually holding the tablet in the solvent spray. Figure 2B shows the mass spectrum obtained from each of the four functions over the m/z range of 400–750. The polarity function is shown for each spectrum in this

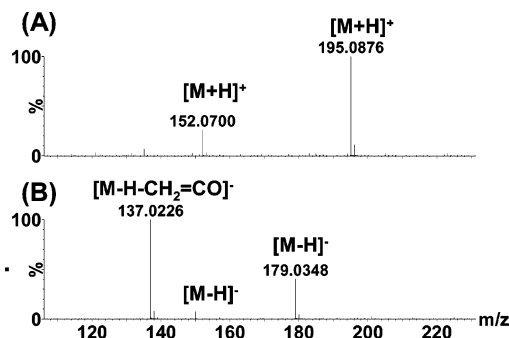


Figure 4. (A) Accurate mass spectrum obtained for Anadin Extra in positive ion mode. (B) Accurate mass spectrum obtained for Anadin Extra in negative ion mode

example only and annotated in the top right corner of each spectrum. It can be clearly seen that leucine enkephalin is suitably ionized in both modes of operation in contrast to erythromycin, which is only ionized in the positive mode of operation forming $[M + H]^+$ of m/z 734. Although very high ion counts were generated, no carryover of protonated erythromycin was observed in the reference mass spectrum.

Figure 3A shows the accurate mass spectrum obtained in real time. The data obtained show a mass error of 0.7 mDa from the theoretical monoisotopic mass of 734.4690 Da. This mass accuracy could not have been achieved by selecting the top of the TIC peak obtained on the instrument used previously⁴ since the ion counts generated by the DESI technique were far too high to allow for successful mass measurement. High ion counts are expected since the erythromycin tablet contained 500 mg of the active ingredient. Figure 3B shows a typical accurate mass in-source CID mass spectrum obtained from sampling an erythromycin tablet using DESI.

In-source collision-induced product ions generated from protonated erythromycin show ions characteristic of the cladinose and desosamine sugars present in the molecule. Elemental formula assignments for the product ions formed are as follows: m/z 734 ($C_{37}H_{68}NO_{13}$) corresponding to $[M + H]^+$, m/z 716 ($C_{37}H_{66}NO_{12}$) corresponding to $[MH - H_2O]^+$, m/z 576 ($C_{29}H_{54}NO_{10}$) corresponding to loss of the cladinose sugar moiety $[MH - \text{cladinose}]^+$ and not loss of the desosamine sugar moiety as previously reported,¹⁵ m/z 558 ($C_{29}H_{52}NO_9$) corresponding to either $[MH - H_2O - \text{cladinose}]^+$ or H_2O loss from m/z 576, m/z

(15) Yang, S.; Carlson, K. H. *J. Chromatogr., A* **2004**, *1038*, 141.

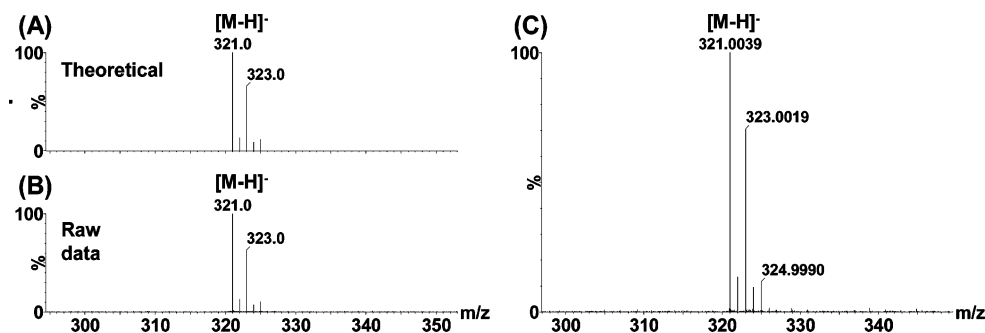


Figure 5. (A) Theoretical isotope distribution for deprotonated chloramphenicol. (B) Mass spectrum obtained for deprotonated chloramphenicol showing comparison with the theoretical isotope distribution. (C) Accurate mass spectrum obtained for deprotonated chloramphenicol.

Table 1. Summary of Results from Five Accurate Mass Measurements of the Active Ingredients in an Anadin Extra Tablet

drug	theoretical [M + H] ⁺	calculated [M + H] ⁺	deviation (mTh)	theoretical [M - H] ⁻	calculated [M - H] ⁻	deviation (mTh)
paracetamol	152.0712	152.0700 152.0698 152.0696	-1.2 -1.4 -1.6	150.0555	150.0544 150.0547 150.0543	-1.1 -0.8 -1.2
mean SD ^c		152.0698	-1.4 0.2		150.0545	-1.0 0.2
caffeine	195.0882	195.0876 195.0871 195.0883	-0.6 -1.1 +0.1	193.0726		
mean SD		195.0877	-0.5 0.6		X ^b	X
aspirin	181.0501			179.0344	179.0348 179.0336 179.0336	+0.4 -0.8 -0.8
mean SD		X	X		179.0340	-0.4 0.7

^a SD, standard deviation. ^b X, not detected.

540 (C₂₉H₅₀NO₈) corresponding to either [MH - (H₂O)₂ - cladinose]⁺ or H₂O loss from *m/z* 558, and *m/z* 522 (C₂₉H₄₈NO₇) corresponding possibly to loss of H₂O from *m/z* 540. The product ion of *m/z* 158 (C₈H₁₆NO₂) corresponds to the desosamine sugar moiety.

Anadin Extra tablets contain three active ingredients, aspirin, paracetamol, and caffeine. A tablet was analyzed with the intention of simultaneously detecting all three compounds using polarity switching within the same acquisition. Figure 4A shows the accurate mass spectrum obtained in positive mode. Two intense ions dominate the mass spectrum, [M + H]⁺ from paracetamol of *m/z* 152 and the base peak [M + H]⁺ from caffeine of *m/z* 195. No protonated molecule was detected for aspirin. A low-abundant ion of *m/z* 163 was detected, which corresponds to [MH - H₂O]⁺ from aspirin. Figure 4B shows the accurate mass spectrum obtained in negative mode. Three ions were detected. Two intense ions in the spectrum correspond to [M - H]⁻ from aspirin of *m/z* 179 and the base peak of *m/z* 137 to [M - H - CH₂=CO]⁻ formed by loss of ketene from deprotonated aspirin. An ion of low abundance with an accurate mass measurement of *m/z* 150.0544 corresponds to deprotonated paracetamol. No deprotonated molecule was detected for caffeine. This shows

the power of polarity switching since all three protonated or deprotonated molecules could be detected in one acquisition, thus negating the need for repeat analysis in each ion mode. The mean and standard deviation of three repeat accurate mass measurements of the active ingredients formulated into Anadin Extra acquired simultaneously in positive and negative modes are shown in Table 1.

Figure 5B shows the partial negative ion mass spectrum obtained by spotting eye ointment containing the active ingredient chloramphenicol on to a piece of matt-finished cardboard. Very high ion counts were generated for the deprotonated molecule [M - H]⁻ of *m/z* 321. An abundant dimer was also detected of *m/z* 643 (not shown). Figure 5A shows the theoretical isotope pattern for deprotonated chloramphenicol as a comparison to the raw data, indicating excellent agreement with that obtained experimentally. The deprotonated molecule for the active ingredient was observed in the mass spectrum at *m/z* 321. Three peaks separated by two *m/z* units were observed that have relative intensities of 9:6:1, consistent with the presence of two chlorine atoms in the molecule. Figure 5C shows the accurate mass spectrum for deprotonated chloramphenicol. Table 2 shows the empirical formulas generated for chloramphenicol taking into

Table 2. Summary of Accurate Mass Measurements for Isotope Peaks of Deprotonated Chloramphenicol

measured mass	theoretical mass	formulas	error (mTh)
321.0039	321.0045	C ₁₁ H ₁₁ O ₅ N ₂ ³⁵ Cl ₂	-0.6
323.0019	323.0015	C ₁₁ H ₁₁ O ₅ N ₂ ³⁵ Cl ³⁷ Cl	0.4
324.9990	324.9985	C ₁₁ H ₁₁ O ₅ N ₂ ³⁷ Cl ₂	0.5

Table 3. Summary of Results from Four Positive Ion Mode Accurate Mass Measurements of the Active Ingredients in Night Nurse Liquid Medicine

drug [M + H] ⁺	measured mass	theoretical mass	error (mTh)
paracetamol	152.0709	152.0711	-0.2
dextromethorphan	272.2012	272.2014	-0.2
promethazine	285.1414	285.1425	-1.1

account the relative contribution of the ³⁵Cl and ³⁷Cl isotopes. No protonated molecule was detected for chloramphenicol in the positive ion mode.

Finally, Night Nurse liquid medicine was analyzed, which contained three active ingredients, paracetamol, promethazine, and dextromethorphan, and was spotted on to cardboard and held with tweezers in the usual way. All three protonated active ingredients were detected in positive ionization mode of *m/z* 152 for paracetamol, *m/z* 272 for dextromethorphan, and *m/z* 285 for promethazine. Only deprotonated paracetamol of *m/z* 150 was detected in negative ionization mode (data not shown). The accurate mass measurements, obtained in positive ionization mode only, are shown in Table 3.

CONCLUSION

The simultaneous analysis of pharmaceutical tablets that contain both basic and acidic polar active ingredients has been demonstrated using polarity switching, a dual-inlet source, and DESI. Analysis of pharmaceutical liquid and cream formulations that contain active ingredients that have different ionization preferences has also been shown. Polarity switching has negated the need for repeat analysis in each ionization mode for formulations that contain a range of compounds that require both positive and negative modes of mass spectrometric detection.

DRE has been successfully employed in these examples by selecting the top of the peak of the ion chromatogram for the generation of an accurate mass measurement through transmission switching experiments. The data generated show mass accuracy to within 2 mTh sufficient for the generation of an empirical formula.

Rapid sample analysis and high-throughput characteristics for the detection of multiple active ingredients in pharmaceutical formulations using polarity switching, DESI, and real-time accurate mass measurement make the technique useful for screening assays and may be of benefit to the pharmaceutical drug discovery process.

ACKNOWLEDGMENT

The authors thank Martin Green and Ashley Sage of Waters MS Technologies, UK for helpful discussions with the manuscript.

Received for review May 18, 2006. Accepted August 22, 2006.

AC0609125

Probing Hemoglobin Structure by Means of Traveling-Wave Ion Mobility Mass Spectrometry

Charlotte A. Scarff, Vibhuti J. Patel, Konstantinos Thalassinou, and James H. Scrivens

Department of Biological Sciences, University of Warwick, Coventry, United Kingdom

Hemoglobin (Hb) is a tetrameric noncovalent complex consisting of two α - and two β -globin chains each associated with a heme group. Its exact assembly pathway is a matter of debate. Disorders of hemoglobin are the most common inherited disorders and subsequently the molecule has been extensively studied. This work attempts to further elucidate the structural properties of the hemoglobin tetramer and its components. Gas-phase conformations of hemoglobin tetramers and their constituents were investigated by means of traveling-wave ion mobility mass spectrometry. Sickle (HbS) and normal (HbA) hemoglobin molecules were analyzed to determine whether conformational differences in their quaternary structure could be observed. Rotationally averaged collision cross sections were estimated for tetramer, dimer, apo-, and holo-monomers with reference to a protein standard with known cross sections. Estimates of cross section obtained for the tetramers were compared to values calculated from X-ray crystallographic structures. HbS was consistently estimated to have a larger cross section than that of HbA, comparable with values obtained from X-ray crystallographic structures. Nontetrameric species observed included apo- and holo- forms of α - and β -monomers and heterodimers; α - and β -monomers in both apo- and holo- forms were found to have similar cross sections, suggesting they maintain a similar fold in the gas phase in both the presence and the absence of heme. Heme-deficient dimer, observed in the spectrum when analyzing commercially prepared Hb, was not observed when analyzing fresh blood. This implies that holo- α -apo- β is not an essential intermediate within the Hb assembly pathway, as previously proposed. (J Am Soc Mass Spectrom 2009, 20, 625–631) © 2009 Published by Elsevier Inc. on behalf of American Society for Mass Spectrometry

Mass spectrometry (MS) has become an important tool for the study of various aspects of protein structure, including the assembly and disassembly of protein complexes, subunit interactions, and ligand interactions [1]. The study of noncovalent complexes has been facilitated by the use of collisional cooling or dampening of ions by elevating the pressure within the source region or by deceleration [2, 3].

The typical electrospray ionization (ESI) mass spectrum of a protein consists of an envelope of peaks attributed to a series of multiply charged gas-phase ions that can indicate the stability and compactness of its structure in the gas phase [4, 5]. Multiply charged ions are produced by proton attachment, predominantly to exposed basic sites on the protein [6], and those of lowest charge are thought to be most representative of native structure [7]. A highly compact protein would have fewer exposed basic sites than those of an unfolded conformation of the same protein and thus would accept fewer charges [4, 8].

Ion mobility spectrometry is a shape-selective technique, based on the time taken for an ion to traverse a mobility cell containing an inert gas under the influence of a weak electric field [9]. This time is related to the rotationally averaged collision cross section, mass, and charge of the ion. The coupling of ion mobility separation with mass spectrometry has provided a powerful method for the analysis of complex mixtures and for the determination of molecular structure [10].

Ion mobility mass spectrometry has emerged as a complementary technique to the well-established methods of X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy for three-dimensional protein structure analysis [11]. There is now substantial evidence to support the view that the gas-phase protein structure can reflect, under controlled conditions, the native solution-phase structure [1, 6, 12]. Multiple studies have shown good agreement between rotationally averaged cross-sectional measurements obtained from X-ray and NMR structures and those obtained by ion mobility experiments [7, 13–17].

In this study traveling-wave ion mobility mass spectrometry (TWIMS) was used to probe the gas-phase conformations of hemoglobin tetramers and their constituents.

Address reprint requests to Dr. James H. Scrivens, University of Warwick, Department of Biological Sciences, Coventry CV4 7AL, UK. E-mail: j.h.scrivens@warwick.ac.uk

A traveling-wave ion mobility separator (T-wave) [18] has been incorporated into a quadrupole time-of-flight (TOF) instrument, the Synapt HDMS system (Waters Corp., Milford, MA, USA) [9]. The T-wave allows for samples to be analyzed at biologically relevant concentrations, having speed and sensitivity advantages over the traditional drift-time ion mobility device. The T-wave device does not allow for absolute cross-sectional measurements to be obtained from drift-time information, although these may be estimated by using reference samples of known cross sections [7, 13, 19].

Hemoglobin (Hb) is a tetramer consisting of four globin chains, two α - and two β -, each associated with a heme group. It is the major oxygen-transport protein found in the red blood cells of all vertebrates. Disorders of hemoglobin are the most common of all inherited disorders and, consequently, the molecule has been extensively studied.

The most debilitating Hb variant is that which causes sickle-cell anemia. This disease occurs when a person inherits two particular mutated copies of the β -globin gene. The sickle-cell mutation results in the production of a β -chain with a single amino acid substitution (β^6 Glu \rightarrow Val) and changes the conformation of the assembled tetramer to allow molecular stacking. Polymerization of this sickle-cell hemoglobin molecule (HbS), in deoxygenated blood, causes the characteristic alteration in shape of red blood cells from biconcave discs to crescentic [20].

ESI-MS has been widely used to detect Hb variants in hemoglobin [21–25] and to investigate its structural assembly into a noncovalent complex [26–29] and its corresponding disassembly [30].

The exact assembly pathway for hemoglobin is still under debate. It is known that one α - and one β -monomer come together to form a heterodimer and that two of these dimers associate to form the tetramer. Since α - and β -monomers can exist in heme-free (apo, α^a and β^a) and heme-bound (holo, α^h and β^h) forms [28], it is unclear as to whether the heme groups are attached to both α - and β -monomers before dimer formation or whether association leads to heme recruitment.

Griffith and Kaltashov have suggested that the formation of a heme-deficient dimer intermediate ($\alpha^h\beta^a$) occurs, consisting of a natively folded holo- α -globin (α^h) and a partially unfolded apo- β -globin (β^a), before complete dimer formation, to ensure correct tetramer structure arrangement [26, 27]. The Konermann group, however, reported that the heme-deficient dimer seen consistently when using commercially available samples, in the form of lyophilized powder, was not observed when using freshly prepared samples [31]. They studied the acid-induced denaturation of bovine Hb and concluded that it followed a highly symmetric mechanism: $(\alpha^h\beta^h)_2 \rightarrow 2\alpha^h\beta^h \rightarrow 2\alpha^h_{\text{folded}} + 2\beta^h_{\text{folded}} \rightarrow 2\alpha^a_{\text{unfolded}} + 2\beta^a_{\text{unfolded}} + 4\text{ heme}$.

This work attempts to further elucidate the structural properties of the hemoglobin tetramer and its compo-

nents and to determine whether conformational differences between the HbA and HbS molecules can be observed by TWIMS.

Experimental

Samples and Sample Preparation

Samples of fresh whole blood were supplied by University Hospitals Coventry and Warwickshire NHS Trust. Sample preparation for mass spectral analysis was adapted from that detailed by Ofori-Acquah et al. [29]. Samples (20 μ L) were diluted 10-fold in 10 mM ammonium acetate (pH 6.8) and spun at 3000 g for 15 min in centrifugal filter units with a 10-kDa cutoff (Microcon YM-10, Millipore Corp., Billerica, MA, USA). Sample retained on the filter was diluted a further 20-fold with 10 mM ammonium acetate and desalted by agitating for two 10-min periods with close to 5 mg of ion-exchange mixed bed resin (AG 501-X8, Bio-Rad Laboratories, Hercules, CA, USA) that had been prepared for use by rinsing twice in liquid chromatography MS grade water. The resulting solutions were introduced into the ESI source of a Synapt HDMS System (Waters) by means of fused silica nanospray needles. All solvents and calibration and protein standards were obtained from Sigma–Aldrich (St. Louis, MO, USA).

Mass Spectral Analysis

Data were acquired by means of a Synapt HDMS system in ESI positive mode with a capillary voltage of 1.2 kV from 1000 to 4500 m/z . The TOF mass analyzer was calibrated using 2 mg/mL cesium iodide in 50% aqueous propan-2-ol.

Instrument acquisition parameters were adjusted to provide the optimal ion mobility separation. The cone voltage was 60 V and the collision energy in the trap region was 10 eV. Source temperature and gas flow were 110 $^{\circ}$ C and 35 mL/min, respectively. Nitrogen was used as the gas in the ion mobility cell and the indicated pressure within the cell was 0.68 mbar, equivalent to a flow rate of 38 mL/min. The backing pressure was increased in increments from 2 to 8 mbar to identify the ideal pressure conditions for transmission of the relevant ionic species. The traveling-wave velocity and wave height were altered in increments from 100 to 600 m/s and 8 to 20 V, respectively, and the conditions that provided the optimal mobility separation were used for all following experiments.

The synchronization of gated release of ions into the ion mobility separator with TOF acquisition allows arrival time distributions of ions to be obtained. For each gate pulse, 200 orthogonal acceleration pushes of the TOF analyzer are recorded to form one ion mobility experiment. The overall mobility recording time is $200 \times t_p$, where t_p is the pusher period [9]. The pusher period depends on the mass acquisition range; for these

experiments, a pusher period of 120 μ s was used, giving a mobility recording time of 24 ms.

Equine myoglobin at a concentration of 10 μ M in 50% aqueous acetonitrile containing 0.2% formic acid was used to provide data that were used to create a calibration curve for cross-sectional measurements.

Data obtained for each hemoglobin tetramer over the m/z range 3000–4500 were deconvoluted onto a true mass scale using maximum entropy modeling (MaxEnt for short) software to provide an estimate of molecular mass. Experiments were carried out in triplicate.

Calibration, Modeling, and Estimation of Cross Section

The equine myoglobin data were used to create a calibration curve for each set of experiments. Absolute cross sections for equine myoglobin were obtained from drift-time ion mobility mass spectrometry (DTIMS) studies (Prof. Michael T. Bowers, personal communication). The calibration was performed using a procedure developed in-house based on previously published work [13, 19, 32, 33]. In brief, normalized cross sections (corrected for charge and reduced mass) were plotted against corrected arrival times (corrected to exclude time spent outside the ion mobility cell) to create a calibration with a power-series fit. The calibration allows one to estimate the cross section of a molecule of interest provided that the mobilities (corrected arrival times) for that molecule lie within the mobilities observed for the calibrant used, irrespective of the size range of cross sections for the calibrant [34, 35]. The calibration was used to estimate rotationally averaged collision cross sections of hemoglobin monomer, dimer, and tetramer for the different charge states observed, based on their arrival time distributions, provided that their corrected arrival times fell along the calibration curve.

To compare the experimental cross sections for the normal and sickle hemoglobin tetramers with accepted values, cross sections were calculated using MOBCAL, a program to calculate mobilities [36, 37], from published X-ray structures held at the RCSB Protein Data Bank [38].

MOBCAL facilitates the use of three approximations to calculate cross sections. The projection approximation (PA) typically results in an underestimation of the cross section of a large ion. It calculates the cross section by averaging the projections produced by every orientation of a molecule and so does not take into account interactions with the buffer gas. The trajectory method (TM) takes into account all interactions, but is computationally intense. The exact hard-sphere scattering model (EHSS) carries out trajectory calculations, while ignoring long-range interactions, but nevertheless gives values within a few percent of the TM approximations [7, 39]. For this work, cross sections were calculated using

the PA and EHSS methods to reduce computational time.

Results and Discussion

Instrument Acquisition Parameters

Considerable optimization of instrument acquisition parameters is required for each individual application of ion mobility separation. This must be tailored to the sample of interest because optimal conditions are dependent on ionic species and mass-to-charge ratio [3]. Controlled optimization of instrument acquisition parameters indicated that a backing pressure of between 6.6 and 6.8 mbar was ideal for intact hemoglobin tetramer analysis. The optimal ion mobility separation of the tetramer was achieved at a traveling-wave velocity and wave height of 400 m/s and 18 V, respectively.

Calibration

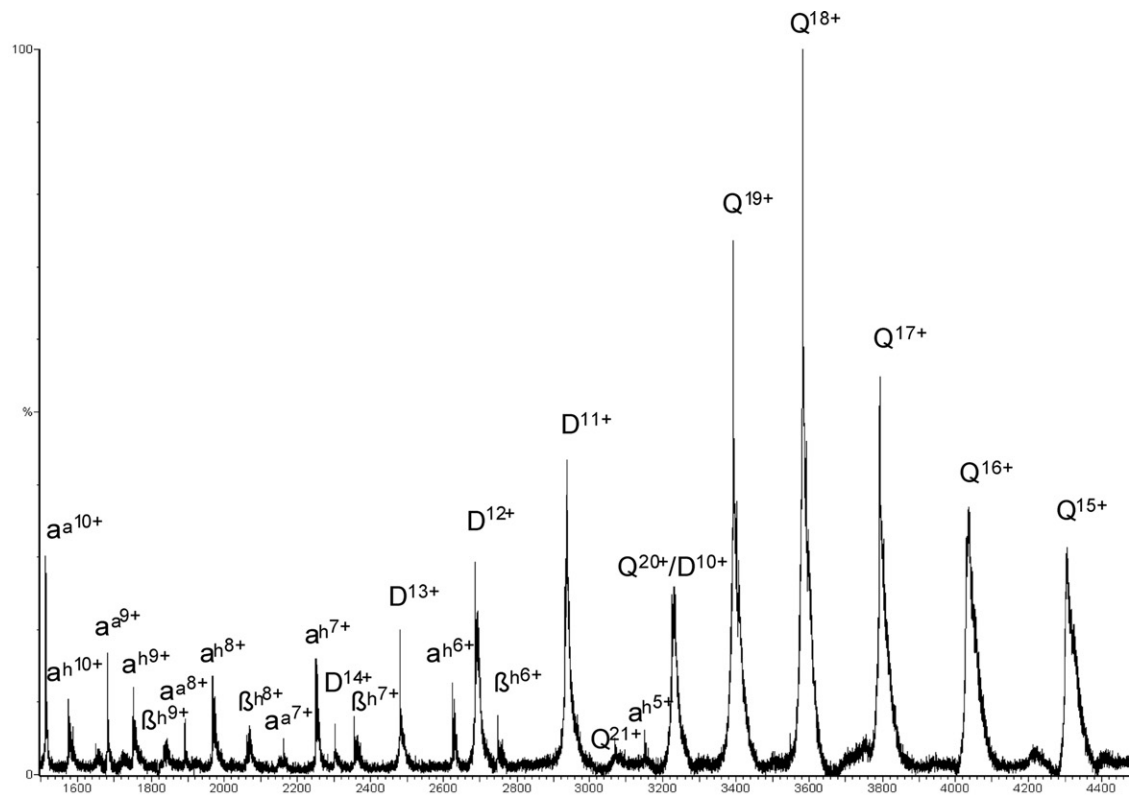
A calibration curve was used to allow the estimation of cross sections for different constituents of hemoglobin in different charge states. Cross sections calculated for equine myoglobin were within 2% of absolute values obtained by DTIMS experiments. These results were reproducible across the three datasets acquired.

Hemoglobin Tetramer Analysis

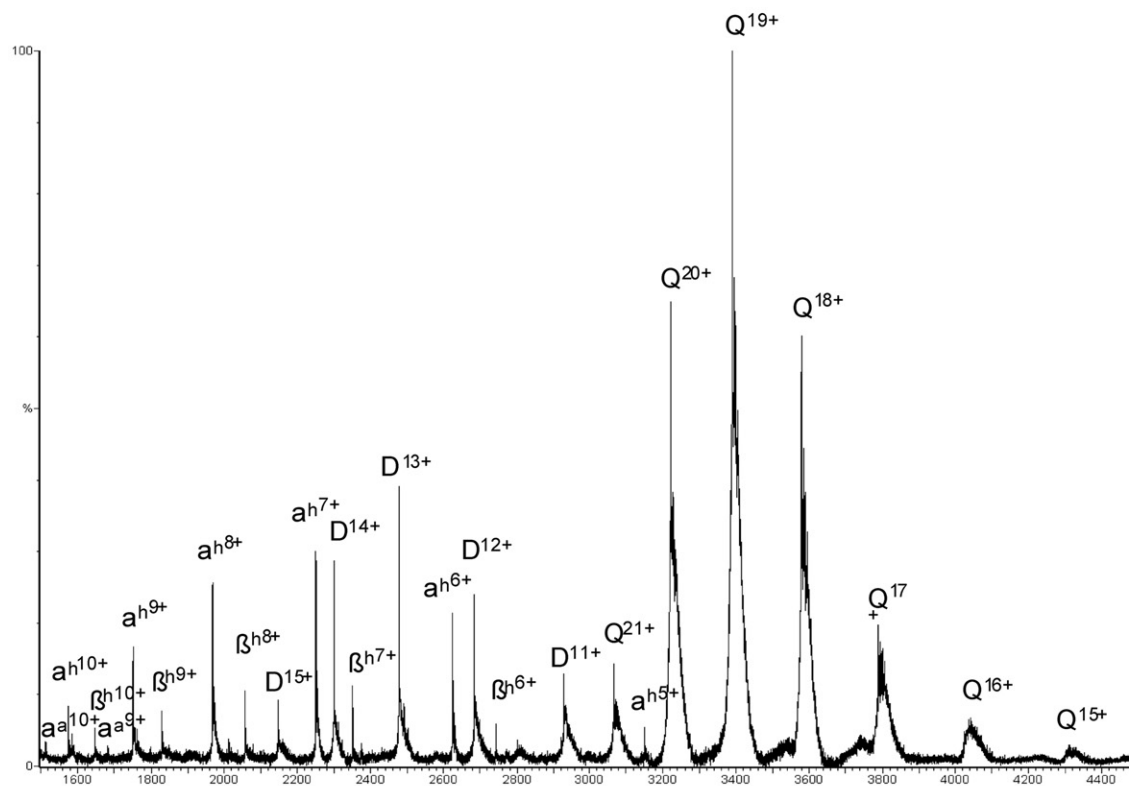
Representative spectra for normal (HbA) and sickle (HbS) hemoglobin analyzed by means of ESI-TOF-MS under nondenaturing conditions are shown in Figures 1a and b, respectively. The data were deconvoluted to give masses of 64,454.7 Da for HbA and 64,395.8 Da for HbS, which were very close to the theoretical masses of 64,453.2 and 64,393.4 Da, respectively [29].

The hemoglobin spectra obtained show the presence of the tetramer $[(\alpha^h\beta^h)_2]$, heterodimer $(\alpha^h\beta^h)$, and apo- and holo-monomer species. The trimer is not seen, as would be expected, because the formation of the hemoglobin tetramer involves the noncovalent association of two $\alpha^h\beta^h$ -dimers. Carefully controlled near-physiological conditions were used in preparing the sample and the absence of any trimer implies that the species observed within the spectra exist naturally in solution. This is consistent with results from isotope labeling studies that showed that nontetrameric ions in the spectrum corresponded to species present in solution [40] rather than products of fragmentation formed during the ESI process [41].

Alpha and beta monomers are observed within the HbA spectrum in both apo- and holo- forms. In a previous study, Griffith and Kaltashov [26] suggested that an α^h monomer first becomes associated with an β^a monomer, to enable the β -chain to incorporate the heme group. This observation was based on the absence of β^h in the spectrum. A subsequent study by Boys and Konermann [31] detected very small quantities of heme-



(a)



(b)

Figure 1. Mass spectra of (a) normal (HbA) and (b) sickle (HbS) hemoglobin analyzed by ESI-TOF-MS under near-physiological conditions, labeled with charge states of tetramer (Q), heterodimer (D), and apo- and holo-monomers (superscripts "a" and "h" refer to apo- and holo-forms, respectively).

deficient dimer and found that both α - and β -monomers were capable of binding heme. The discrepancies observed are thought to be attributable to differences between the commercially prepared and freshly obtained samples used. In the work reported here, in which fresh blood samples were used, β^h was observed in multiply charged states.

It has been reported [42] that without the attachment of the heme group, α - and β -monomers adopt extensively unfolded conformations. Cross sections for various charge states of α - and β -monomers in both the apo- and holo- forms have been estimated and our observations suggest that the predominant conformations of α - and β -monomers in the gas phase are similar to each other and show little change in the absence or the presence of heme (Figure 2). The cross section of each of the molecules studied increases with an increase in charge, thought to be a result of the effects of Coulomb repulsion [16, 43].

The heme-deficient dimer observed in previous studies is not observed here. The existence of both apo- and holo- forms of α - and β -monomers, all of similar cross sections, does not support the need for a β^a to associate with α^h for the β -monomer to recruit heme. Analysis of commercially prepared human hemoglobin (data not shown) does show the presence of a heme-deficient dimer at a molecular mass 32 Da higher than expected; this is in agreement with previous work conducted by the Konermann group [31] on bovine hemoglobin, who attributed the additional mass to the occurrence of oxidative modifications in the commercial protein.

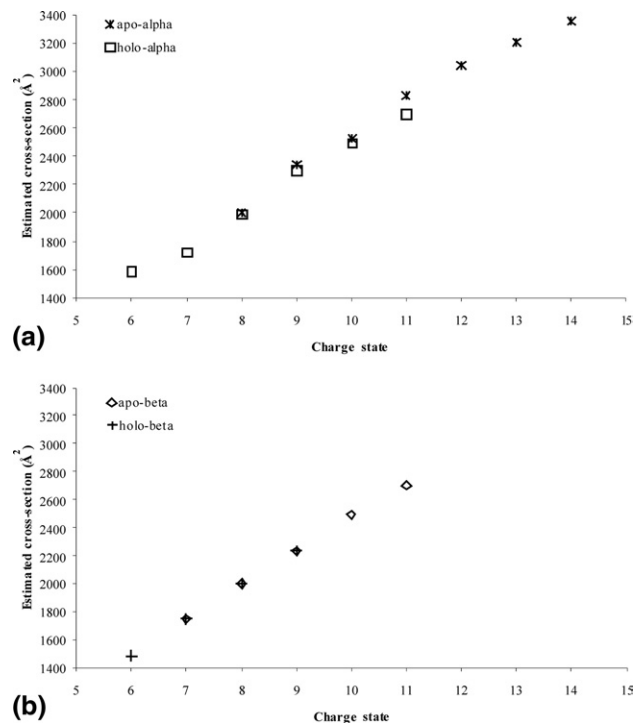


Figure 2. Average estimated cross sections for charge states of (a) apo- α and holo- α monomers and (b) apo- β and holo- β observed within three datasets.

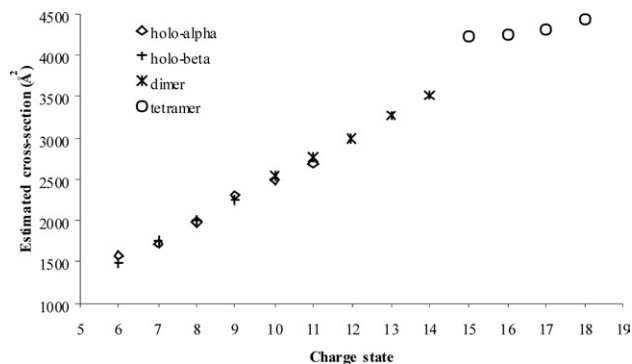


Figure 3. Average estimated cross sections for holo- α , holo- β , heterodimer, and HbA tetramer, from three datasets.

We observe many more charge states of α^a than of β^a . The number of charges accepted on a protein is related to the number of exposed basic sites on the protein's surface. A more folded protein has fewer of its basic sites exposed than an unfolded conformation and thus cannot accept as many charges. This may suggest that the α -chain adopts more unfolded conformations in the gas phase than is possible for the β -chain but, alternatively, the absence of higher charge states of β^a in the spectra may be ascribed to the different desolvation behavior of α - and β -monomers. The α -chain ionizes preferentially over the β -chain because of its greater nonpolar character, thereby competing more effectively for charge [41].

By estimating the cross sections of monomer, dimer, and tetramer, a picture of the assembly process can be obtained (Figure 3). The $[M + 12H]^{12+}$ charge state of dimer has an estimated cross section of 3001 \AA^2 . The $[M + 6H]^{6+}$ charge states of α^h and β^h have estimated average cross sections of 1583 and 1488 \AA^2 , respectively. If these two globin monomers come together to form a dimer, one would expect that the cross section of that dimer would be approximately the addition of the cross sections of the two constituent parts and, indeed, that is the case here. One would further expect that the cross section observed would be slightly smaller than the sum of the monomer subunits because the contact area on both of the monomers would be compacted and contribute less to the overall cross section. The data are in agreement with this.

The average estimated collision cross sections for HbA and HbS, for four different charge states, are illustrated in Figure 4. The data indicate a difference in cross section between normal and sickle-cell hemoglobin and a variation in cross section with charge state.

For HbA and HbS for the charge states studied, the cross sections observed for HbS are somewhat larger than those of HbA. Secondary, tertiary, and quaternary structural considerations make it difficult to determine what the charge state of a molecule should be, theoretically, within a particular solvent at a particular pH. It is clear from previous work that the lowest charge states observed under near-physiological conditions are most

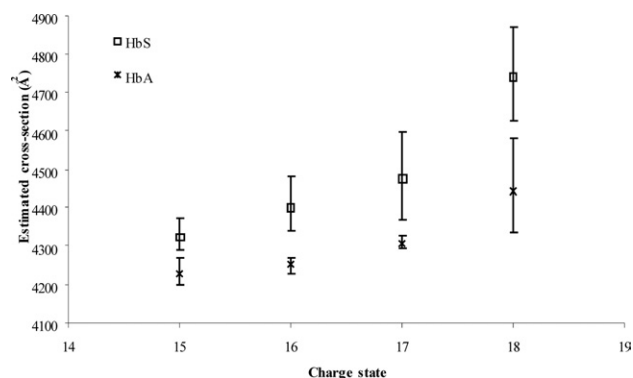


Figure 4. Estimated cross sections for HbA and HbS tetramers for three different charge states, showing averaged values from three datasets with corresponding errors.

representative of the native protein structure [7]. The $[M + 18H]^{18+}$ charge state, for HbA and HbS, is suspected to be representative of a tetrameric structure that is beginning to denature.

The reproducibility of the cross sections estimated for the $[M + 15H]^{15+}$, $[M + 16H]^{16+}$, and $[M + 17H]^{17+}$ charge states of HbA, between the three replicate datasets, is $\pm 1\%$, believed to be representative of the reproducibility capabilities of the experiment. The larger deviation in estimation of cross section for the same charge states for HbS, of $\pm 3\%$, may reflect the presence of a broader population of conformations of the HbS molecule of similar cross section.

The rotationally averaged cross sections for HbA and HbS calculated from X-ray crystallographic structures were 3313 and 3733 Å² for the PA and 4343 and 4775 Å² for the EHSS, respectively. Values estimated experimentally for the $[M + 15H]^{15+}$, $[M + 16H]^{16+}$ and $[M + 17H]^{17+}$ charge states of HbA and HbS fall between these two theoretical approximations and agree with the X-ray observation that HbS has a larger cross section than that of HbA. Gas-phase conformations, although illustrative of solution-phase structures under controlled conditions, have been shown previously to be smaller than those predicted by EHSS approximations [6]; a more compact conformation is thought to be adopted in the gas phase because increased intramolecular interactions cause polar side chains to collapse onto the protein's surface [14].

Conclusions

This work demonstrates the use of TWIMS to probe gas-phase conformations of three-dimensional protein structure and noncovalent complexes.

TWIMS has been successfully used to analyze hemoglobin tetramers. Cross sections calculated for intact hemoglobin tetramers are comparable to those estimated from published X-ray crystallography data and conformational differences are observed between the HbA and HbS molecules. Nontetrameric species observed, including apo- and holo- forms of α - and

β -monomers and $\alpha^h\beta^h$ -dimers, are naturally present in equilibrium in solution and are not products of fragmentation during the ESI process.

Both α - and β -monomers have cross sections similar to each other, suggesting that they maintain a similar fold in the gas phase. Apo- and holo- forms of the monomers also have similar cross sections, suggesting that α - and β -monomers can retain a folded structure in the absence and the presence of the heme group. Extensively disordered monomer structures are not observed.

A heme-deficient dimer has not been observed and the results do not suggest the requirement for association of β^3 with α^h for the β -monomer to recruit heme. The results, obtained on fresh blood samples rather than commercially prepared samples, do not support the hypothesis that a heme-deficient dimer is an essential intermediate in the tetramer assembly process.

Acknowledgments

The authors thank Dr. Nicholas Jackson and Yvonne Elliot, University Hospitals Coventry and Warwickshire NHS Trust, for providing samples and Prof. Michael T. Bowers for providing cross-sectional measurements for equine myoglobin.

References

- Heck, A. J. R.; van den Heuvel, R. H. H. Investigation of Intact Protein Complexes by Mass Spectrometry. *Mass Spectrom. Rev.* **2004**, *23*, 368–389.
- Chernushevich, I. V.; Thomson, B. A. Collisional Cooling of Large Ions in Electrospray Mass Spectrometry. *Anal. Chem.* **2004**, *76*, 1754–1760.
- Tahallah, N.; Pinkse, M.; Maier, C. S.; Heck, A. J. R. The Effect of the Source Pressure on the Abundance of Ions of Noncovalent Protein Assemblies in an Electrospray Ionization Orthogonal Time-of-Flight Instrument. *Rapid Commun. Mass Spectrom.* **2001**, *15*, 596–601.
- Chowdhury, S. K.; Katta, V.; Chait, B. T. Probing Conformational Changes in Proteins by Mass Spectrometry. *J. Am. Chem. Soc.* **1990**, *112*, 9012–9013.
- Mohimen, A.; Dobo, A.; Hoerner, J. K.; Kaltashov, I. A. A Chemometric Approach to Detection and Characterization of Multiple Protein Conformers in Solution Using Electrospray Ionization Mass Spectrometry. *Anal. Chem.* **2003**, *75*, 4139–4147.
- Hoaglund-Hyzer, C. S.; Counterman, A. E.; Clemmer, D. E. Anhydrous Protein Ions. *Chem. Rev.* **1999**, *99*, 3037–3080.
- Scarff, C. A.; Thalassinos, K.; Hilton, G. R.; Scrivens, J. H. Travelling Wave Ion Mobility mass Spectrometry Studies of Protein Structure: Biological Significance and Comparison with X-ray Crystallography and Nuclear Magnetic Resonance Spectroscopy Measurements. *Rapid Commun. Mass Spectrom.* **2008**, *22*, 3297–3304.
- Kaltashov, I. A.; Eyles, S. J. Studies of Biomolecular Conformations and Conformational Dynamics by Mass Spectrometry. *Mass Spectrom. Rev.* **2002**, *21*, 37–71.
- Pringle, S. D.; Giles, K.; Wildgoose, J. L.; Williams, J. P.; Slade, S. E.; Thalassinos, K.; Bateman, R. H.; Bowers, M. T.; Scrivens, J. H. An Investigation of the Mobility Separation of Some Peptide and Protein Ions Using a New Hybrid Quadrupole/Travelling Wave IMS/oa-ToF Instrument. *Int. J. Mass Spectrom.* **2007**, *261*, 1–12.
- Kanu, A. B.; Dwivedi, P.; Tam, M.; Matz, L.; Hill, H. H. J. Ion Mobility-Mass Spectrometry. *J. Mass Spectrom.* **2008**, *43*, 1–22.
- van der Heuvel, R. H.; Heck, A. J. Native Protein Mass Spectrometry: From Intact Oligomers to Functional Machineries. *Curr. Opin. Chem. Biol.* **2004**, *8*, 519–526.
- Kaddis, C. S.; Lomeli, S. H.; Yin, S.; Berhane, B.; Apostol, M. I.; Kickhoefer, V. A.; Rome, L. H.; Loo, J. A. Sizing Large Proteins and Protein Complexes by Electrospray Ionization Mass Spectrometry and Ion Mobility. *J. Am. Soc. Mass Spectrom.* **2007**, *18*, 1206–1216.
- Ruotolo, B. T.; Giles, K.; Campuzano, I.; Sandercock, A. M.; Bateman, R. H.; Robinson, C. V. Evidence for Macromolecular Protein Rings in the Absence of Bulk Water. *Science* **2005**, *310*, 1658–1661.
- Shelimov, K. B.; Clemmer, D. E.; Hudgins, R. R.; Jarrold, M. F. Protein Structure In Vacuo: Gas-Phase Conformations of BPTI and Cytochrome c. *J. Am. Chem. Soc.* **1997**, *119*, 2240–2248.

15. Shelimov, K. B.; Jarrold, M. F. Conformations, Unfolding, and Refolding of Apomyoglobin in Vacuum: An Activation Barrier for Gas-Phase Protein Folding. *J. Am. Chem. Soc.* **1997**, *119*, 2987–2994.
16. Valentine, S. J.; Anderson, J. G.; Ellington, A. D.; Clemmer, D. E. Disulfide-Intact and -Reduced Lysozyme in the Gas Phase: Conformations and Pathways of Folding and Unfolding. *J. Phys. Chem. B.* **1997**, *101*, 3891–3900.
17. Myung, S.; Badman, E. R.; Lee, Y. J.; Clemmer, D. E. Structural Transitions of Electrosprayed Ubiquitin Ions Stored in an Ion Trap over ~10 ms to 30 s. *J. Phys. Chem. A.* **2002**, *106*, 9976–9982.
18. Giles, K.; Pringle, S. D.; Worthington, K. R.; Little, D.; Wildgoose, J. L.; Bateman, R. H. Applications of a Travelling Wave-based Radio-Frequency-Only Stacked Ring Ion Guide. *Rapid Commun. Mass Spectrom.* **2004**, *18*, 2401–2414.
19. Wildgoose, J. L.; Giles, K.; Pringle, S. D.; Koeniger, S. J.; Valentine, R. H.; Bateman, R. H.; Clemmer, D. E. A Comparison of Travelling Wave and Drift Tube Ion Mobility Separations. In *Proceedings of the 54th ASMS Conference on Mass Spectrometry and Allied Topics*, Seattle, WA, 2006.
20. Murayama, M. Structure of Sickle Cell Hemoglobin and Molecular Mechanism of the Sickling Phenomenon. *Clin. Chem.* **1967**, *13*, 578–588.
21. Daniel, Y. A.; Turner, C.; Haynes, R. M.; Hunt, B. J.; Dalton, R. N. Rapid and Specific Detection of Clinically Significant Haemoglobinopathies Using Electrospray Mass Spectrometry-Mass Spectrometry. *Br. J. Haematol.* **2005**, *130*, 635–643.
22. Shackleton, C. H. L.; Falick, A. M.; Green, B. N.; Witkowska, H. E. Electrospray Mass Spectrometry in the Clinical Diagnosis of Variant Hemoglobins. *J. Chromatogr. B Biomed. Appl.* **1991**, *562*, 175–190.
23. Wild, B. J.; Green, B. N.; Stephens, A. D. The Potential of Electrospray Ionization Mass Spectrometry for the Diagnosis of Hemoglobin Variants Found in Newborn Screening. *Blood Cells Mol. Dis.* **2004**, *33*, 308–317.
24. Wild, B. J.; Green, B. N.; Cooper, E. K.; Laloz, M. R.; Erten, S.; Stephens, A. D.; Layton, D. M. Rapid Identification of Hemoglobin Variants by Electrospray Ionization Mass Spectrometry. *Blood Cells Mol. Dis.* **2001**, *27*, 691–704.
25. Shimizu, A.; Nakanishi, T.; Miyazaki, A. Detection and Characterization of Variant and Modified Structures of Proteins in Blood and Tissues by Mass Spectrometry. *Mass Spectrom. Rev.* **2006**, *25*, 686–712.
26. Griffith, W. P.; Kaltashov, I. A. Highly Asymmetric Interactions between Globin Chains during Hemoglobin Assembly Revealed by Electrospray Ionization Mass Spectrometry. *Biochemistry* **2003**, *42*, 10024–10033.
27. Griffith, W. P.; Kaltashov, I. A. Protein Conformational Heterogeneity as a Binding Catalyst: ESI-MS Study of Hemoglobin H Formation. *Biochemistry* **2007**, *46*, 2020–2026.
28. Boys, B. L.; Konermann, L. Folding and Assembly of Hemoglobin Monitored by Electrospray Mass Spectrometry Using an On-line Dialysis System. *J. Am. Soc. Mass Spectrom.* **2007**, *18*, 8–16.
29. Ofori-Acquah, S. F.; Green, B. N.; Davies, S. C.; Nicolaidis, K. H.; Serjeant, G. R.; Layton, D. M. Mass Spectral Analysis of Asymmetric Hemoglobin Hybrids: Demonstration of Hb FS ($\alpha_2\gamma\beta^S$) in Sickle Cell Disease. *Anal. Biochem.* **2001**, *298*, 76–82.
30. Versluis, C.; Heck, A. J. R. Gas-Phase Dissociation of Hemoglobin. *Int. J. Mass Spectrom.* **2001**, *210–211*, 637–649.
31. Boys, B. L.; Kuprowski, M. C.; Konermann, L. Symmetric Behavior of Hemoglobin α - and β -Subunits during Acid-induced Denaturation Observed by Electrospray Mass Spectrometry. *Biochemistry* **2007**, *46*, 10675–10684.
32. Williams, J. P.; Scrivens, J. H. Coupling Desorption/Electrospray Ionisation and Neutral Desorption/Extractive Electrospray Ionisation with a Travelling-Wave Based Ion Mobility Mass Spectrometer for the Analysis of Drugs. *Rapid Commun. Mass Spectrom.* **2008**, *22*, 187–196.
33. Scrivens, J. H.; Thalassinou, K.; Hilton, G.; Slade, S. E.; Pinheiro, T. J. T.; Bateman, R. H.; Bowers, M. T. Use of a Travelling Wave-based Ion Mobility Approach to Resolve Proteins of Varying Conformation. In *Proceedings of the 55th ASMS Conference on Mass Spectrometry and Allied Topics*, Indianapolis, IN, 2006.
34. Shvartsburg, A. A.; Smith, R. D. Fundamentals of Traveling Wave Ion Mobility Spectrometry. *Anal. Chem.* (November 6, 2008). doi:10.1021/ac8016295 [Epub ahead of print].
35. Thalassinou, K.; Grabenauer, M.; Slade, S. E.; Hilton, G. R.; Bowers, M. T.; Scrivens, J. H. Characterization of Phosphorylated Peptides Using Traveling Wave-based and Drift Cell Ion Mobility Mass Spectrometry. *Anal. Chem.* (December 4, 2008). [Epub ahead of print].
36. Mesleh, M. F.; Hunter, J. M.; Shvartsburg, A. A.; Schatz, G. C.; Jarrold, M. F. Structural Information from Ion Mobility Measurements: Effects of the Long-range Potential. *J. Phys. Chem.* **1996**, *100*, 16082–16086.
37. Shvartsburg, A. A.; Jarrold, M. F. An Exact Hard-Spheres Scattering Model for the Mobilities of Polyatomic Ions. *Chem. Phys. Lett.* **1996**, *261*, 86–91.
38. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.
39. Jarrold, M. F. Unfolding, Refolding, and Hydration of Proteins in the Gas Phase. *Acc. Chem. Res.* **1999**, *32*, 360–367.
40. Hossain, B. M.; Konermann, L. Pulsed Hydrogen/Deuterium Exchange MS/MS for Studying the Relationship between Noncovalent Protein Complexes in Solution and in the Gas Phase after Electrospray Ionization. *Anal. Chem.* **2006**, *78*, 1613–1619.
41. Kuprowski, M. C.; Boys, B. L.; Konermann, L. Analysis of Protein Mixtures by Electrospray Mass Spectrometry: Effects of Conformation and Desolvation Behavior on the Signal Intensities of Hemoglobin Subunits. *J. Am. Soc. Mass Spectrom.* **2007**, *18*, 1279–1285.
42. Leutzinger, Y.; Beychok, S. Kinetics and Mechanism of Heme-induced Refolding of Human α -Globin. *Proc. Nat. Acad. Sci. U. S. A.* **1981**, *78*, 780–784.
43. Badman, E. R.; Hoaglund-Hyzer, C. S.; Clemmer, D. E. Monitoring Structural Changes of Proteins in an Ion Trap over ~10–200 ms: Unfolding Transitions in Cytochrome c Ions. *Anal. Chem.* **2001**, *73*, 6000–6007.

A Comparison of Labeling and Label-Free Mass Spectrometry-Based Proteomics Approaches

Vibhuti J. Patel,^{†,§} Konstantinos Thalassinos,^{†,§} Susan E. Slade,[†] Joanne B. Connolly,[‡]
 Andrew Crombie,[†] J. Colin Murrell,[†] and James H. Scrivens^{†,*}

Department of Biological Sciences, University of Warwick, Coventry CV4 7AL, United Kingdom, and Waters Corporation, Atlas Park, Simonsway, Manchester M22 5PP, United Kingdom

Received February 2, 2009

Abstract: The proteome of the recently discovered bacterium *Methylocella silvestris* has been characterized using three profiling and comparative proteomics approaches. The organism has been grown on two different substrates enabling variations in protein expression to be identified. The results obtained using the experimental approaches have been compared with respect to number of proteins identified, confidence in identification, sequence coverage and agreement of regulated proteins. The sample preparation, instrumental time and sample loading requirements of the differing experiments are compared and discussed. A preliminary screen of the protein regulation results for biological significance has also been performed.

Keywords: proteomics • quantification • iTRAQ • label-free • methanotroph • mass spectrometry

Introduction

Since the mid-1990s, mass spectrometry-based strategies have been the mainstream method for protein identification.¹ There remain, however, a number of issues to be tackled. Intrinsic characteristics of proteomes raise a number of experimental challenges. By nature, proteomes are large and complex. A single gene can often give rise to multiple, distinct proteins due to alternative splicing, sequence polymorphisms and post-translational modifications. Protein databases generated from the genome of an organism may, therefore, not be a true reflection of the potential protein complement.² There has been significant progress in the development of new approaches to tackle these issues, but technical challenges persist.

An ideal approach would enable the comprehensive characterization of proteomes in a high-throughput manner. Currently, the techniques involved can be complex, costly and involve time-consuming data analysis. A low number of replicate experiments conducted—often due to a lack of sample availability—means that reproducibility is a concern. In addition, any given technique may only yield information on a fraction of the relevant peptides in any single analytical run.³

An established proteomics approach is based on the separation of proteins via one- or two-dimensional polyacrylamide gel electrophoresis (PAGE). Proteins are digested within the gel, and the resulting peptides extracted for MS analysis. Drawbacks associated with PAGE include dynamic range, insufficient resolving power to fully separate all proteins within a sample, and restricted sample throughput.⁴

Nongel-based techniques have been developed for the analysis of complex proteomic samples: so-called ‘shotgun’ experiments, where a whole proteome is digested without prior protein separation. Typically, the resulting peptides are separated by strong cation exchange (SCX) chromatography before reversed-phase LC-MS/MS analysis,⁵ an example of an approach known as multidimensional protein identification techniques (MudPIT). This method has been shown to provide increased proteome coverage compared to gels, although it still suffers from problems with reproducibility and dynamic range. This approach has gained popularity within proteomic studies in preference to gels.²

In addition to providing a profile of what proteins are present within a system at a given time, information on the expression levels of these proteins is increasingly required. Techniques in comparative and quantitative proteomics have, therefore, also been developed significantly in recent years. Relative quantification can be performed on proteins separated by two-dimensional PAGE, using image analysis software, sometimes incorporating selective labeling approaches such as difference gel electrophoresis (DiGE).⁶ This approach is subject to the restrictions imposed by the gel methods.

A number of labeling approaches can also be incorporated into ‘shotgun’ type experiments. These include stable isotope labeling by amino acids in cell culture (SILAC),⁷ isotope dilution,⁸ stable isotope labeled peptides,⁹ radiolabeled amino acid incorporation,¹⁰ chemically synthesized peptide standards,¹¹ tandem mass tags (TMT),¹² isotope-coded affinity tags (ICAT),¹³ and more recently, isobaric tags for relative and absolute quantification (iTRAQ).¹⁴ The iTRAQ system is now commercially available with eight isobaric tags,¹⁵ having only initially been available with four tags, and has been widely used in proteomic studies.¹⁶

Most label-based quantification approaches have potential limitations: complex sample preparation, the requirement for increased sample concentration, and incomplete labeling.

* To whom correspondence should be addressed. Tel: +44(0)24-7657-4189. Fax: +44(0)24-7652-3568 E-mail: j.h.scrivens@warwick.ac.uk.

[†] University of Warwick.

[§] Both authors contributed equally to this work.

[‡] Waters Corporation.

There has, therefore, recently been a focus in the area of nonlabeled quantification in order to address some of these issues.¹⁷

Nonlabeled techniques which have been developed include peptide match score summation (PMSS)¹⁸ and spectrum sampling (SpS),¹⁹ both of which can be combined with statistical evaluation to detect differentially expressed proteins.²⁰ Another approach utilizes a protein abundance index (PAI),²¹ which can be converted to exponentially modified PAI (emPAI) for absolute protein quantification.²²

It has been observed that electrospray ionization (ESI) provides signal responses that correlate linearly with increasing concentration,²³ but there have been concerns regarding the nonlinearity of signal response.²⁴ Previous works have introduced quantitative, label-free LC-MS-based strategies for global profiling of complex protein mixtures.^{25,26} More recently, a simple LC-MS-based methodology was published which relies on changes in signal response from each accurate mass measurement and corresponding retention time (AMRT) to directly reflect concentrations in one sample relative to another,²⁷ which has since been developed into a label-free system capable of relative and absolute quantification.^{28,29} All detectable, eluting peptides and their corresponding fragments are observed via rapid switching between high and low collision energy during the LC-MS/MS experiment, giving a comprehensive list of all ions that can subsequently be searched.³⁰

In this work, three proteomics approaches have been used to identify and relatively quantify the proteins within a bacterium when grown under different substrates. Samples have been analyzed both qualitatively and quantitatively by: (i) one-dimensional PAGE; (ii) MudPIT incorporating iTRAQ tags; and (iii) a data-independent, alternate scanning LC-MS method enabling label-free quantification. Comparisons have been made regarding experimental considerations such as ease of use, amount of biological sample required, time required to prepare samples for analysis and total instrument time. The data obtained have been evaluated with respect to number of protein identifications, confidence of the assignments, sequence coverage and agreement of regulated proteins. All approaches have been carried out using equivalent instrumentation, enabling the results to be more directly compared. The organism used in this study is the methanotrophic bacterium *Methylocella silvestris*, an environmentally important organism involved in global methane cycling. Unlike other methanotrophs, *M. silvestris* is able to grow on multicarbon compounds as well as on methane.³¹ In this work, cultures of *M. silvestris* were grown on acetate and methane.

Materials and Methods

Bacterial Growth and Sample Preparation. *M. silvestris* was grown in fermenter cultures on diluted nitrate mineral salts (NMS) medium with methane or acetate (5 mM) as previously described.³¹ Cells, grown to late exponential phase ($OD_{540} \sim 1.0$), were harvested by centrifugation (17 700g, 20 min, 4 °C), washed in growth medium, resuspended in 0.1 M PIPES buffer (piperazine-*N,N'*-bis[2-ethanesulfonic acid], pH 7.0), and frozen in liquid nitrogen. Subsequently, frozen cells were thawed and resuspended in PIPES buffer containing 1 mM benzamidine and broken by four passes through a French pressure cell at 125 MPa (4 °C) (American Instrument Co., Silver Spring, MD). Cell debris and membranes were removed by two centrifugation steps (13 000g, 30 min, 4 °C, followed by 140 000g, 90 min, 4 °C), and the supernatant, containing soluble cytoplasmic

proteins, was used for analysis. A protein assay was conducted on the soluble extract, using a Micro BCA Protein Assay Kit (Pierce Protein Research Products, Thermo Scientific, Cramlington, U.K.) according to the manufacturer's protocol.

Protein Separation by Gel Electrophoresis. Proteins were resolved by 1D SDS-PAGE (14 μ g per lane) and stained with Coomassie Blue. Thirty to 40 slices were excised from each lane, and subjected to tryptic digestion. All processing of the gel plugs was performed by a MassPrep robotic protein handling system (Waters Corporation, Manchester, U.K.) using the manufacturer's protocol. In brief, the gel plugs were destained, the disulfide bonds were reduced by the addition of dithiothreitol, and the free cysteine residues were alkylated with iodoacetamide. The gel plugs were washed prior to a dehydration step, followed by the addition of trypsin (Promega, Southampton, U.K.), and incubated for 4.5 h. The resultant tryptic peptides were extracted twice and transferred to a cooled 96-well microtiter plate; if necessary, they were stored at -20 °C.

iTRAQ Labeling and Strong Cation Exchange Chromatography (SCX). Labeled quantification was carried out using the iTRAQ 4-plex labeling kit (Applied Biosystems, Warrington, U.K.). Protein extracts from the two growth conditions were digested and labeled according to the manufacturer's standard protocol, and the samples pooled and lyophilized. A total of 400 μ g of protein from each growth condition was labeled, giving a total protein loading of 800 μ g. As SCX was carried out offline, the potential for sample losses is higher. A larger initial protein loading was therefore used in order to minimize such losses and optimize the number of proteins identified by this approach. A total of 200 μ g of acetate-grown sample was labeled with the 114 reporter tag, and 200 μ g with the 116 reporter tag. Two hundred micrograms of the methane-grown sample was labeled with the 115 tag and 200 μ g with the 117 tag. As per the manufacturer's protocol, a maximum of 100 μ g of protein was labeled per vial of iTRAQ label, that is, two vials were used per label. The labeling of one growth condition with two different iTRAQ tags provides the means for an internal control to monitor labeling efficiency. The labeled tryptic peptides were partially resolved using a PolySULFOETHYL A SCX column, 2.1 mm \times 20 cm, 5 μ m particles, 300 Å pore size (PolyLC, Columbia, MD), using a stepwise gradient of KCl, adapted from Link et al.³² from 2.5–50% salt solution over a period of 75 min. In total, 64 fractions were collected.

In-Solution Tryptic Digestion. A total of 100 μ g of soluble protein extract was resuspended in 1 mL of 0.1% RapiGest (Waters Corporation, Milford, MA) and concentrated using a 5 kDa cutoff spin column. The solution was then heated at 80 °C for 15 min, reduced with DTT at 60 °C for 15 min, alkylated in the dark with iodoacetamide at ambient temperature for 30 min, and digested with 1:50 (w/w) sequencing grade trypsin (Promega, Southampton, U.K.) for 16 h. RapiGest was hydrolyzed by the addition of 2 μ L of 15 M HCl, centrifuged, and each sample diluted 1:1 with a 50 fmol/ μ L glycogen phosphorylase B standard tryptic digest to give a final protein concentration of 500 ng/ μ L per sample and 25 fmol/ μ L phosphorylase B.

LC-MS/MS Acquisition for Gel-Separated Samples. Peptides extracted from the digested gel were transferred to a nanoACQUITY system (Waters Corporation). A 6.4 μ L aliquot of extract was mixed with 13.6 μ L of 0.1% formic acid and loaded onto a 0.5 cm LC Packings C18 5 μ m 100 Å 300 μ m i.d. μ -precolumn cartridge. Flushing the column with 0.1% formic acid desalted the bound peptides before a linear gradient of solvent B (0.1%

formic acid in acetonitrile) at a flow rate of approximately 200 nL/min eluted the peptides for further resolution on a 15 cm LC Packings C18 5 μm , 5 \AA , 75 μm i.d. PepMap analytical column. The eluted peptides were analyzed on a Micromass Q-ToF Global Ultima (Waters Corporation) mass spectrometer fitted with a nano-LC sprayer with an applied capillary voltage of 3.5 kV. The spectral acquisition scan rate was 1.0 s with a 0.1 s interscan delay. The instrument was calibrated against a collisionally induced dissociation (CID) spectrum of the doubly charged precursor ion of [Glu¹]-fibrinopeptide B (GFP, Sigma Aldrich, St. Louis, MO), and fitted with a GFP lockspray line. The instrument was operated in data dependent acquisition (DDA) mode over the mass/charge (m/z) range of 50–2000. During the DDA analysis, CID experiments were performed on the three most intense, multiply charged peptides as they eluted from the column at any given time. Once these data have been collected, the next three most intense peptides are selected, and this process repeated.

LC-MS/MS Acquisition for iTRAQ Samples. Fractions collected from the SCX separation of iTRAQ-labeled peptides were snap-frozen on dry ice and lyophilized to dryness. The samples were resuspended in 20 μL of 0.1% formic acid and transferred to a CapLC system (Waters Corporation). A 6.4 μL aliquot of extract was mixed with 13.6 μL of 0.1% formic acid and loaded onto a 0.5 cm LC Packings C18 5 μm , 100 \AA , 300 μm i.d. precolumn cartridge. Flushing the column with 0.1% formic acid desalted the bound peptides before a linear gradient of solvent B (0.1% formic acid in acetonitrile) at a flow rate of approximately 200 nL/min eluted the peptides for further resolution on a 15 cm LC Packings C18 5 μm , 5 \AA , 75 μm i.d. PepMap analytical column. The eluted peptides were analyzed on a Micromass Q-ToF Global Ultima (Waters Corporation) mass spectrometer fitted with a nano-LC sprayer with an applied capillary voltage of 3.5 kV. The spectral acquisition scan rate was 1.0 s with a 0.1 s interscan delay. The instrument was calibrated against a CID spectrum of the doubly charged precursor ion of GFP, and fitted with a GFP lockspray line. The instrument was operated in data dependent acquisition (DDA) mode as described above.

LC-MS Configurations for Label-Free Analysis. Nanoscale LC separations of tryptic peptides for qualitative and quantitative multiplexed LC-MS analysis were performed with a nano-ACQUITY system (Waters Corporation) using a Symmetry C₁₈ trapping column (180 μm \times 20 mm, 5 μm) and a BEH C₁₈ analytical column (75 μm \times 250 mm, 1.7 μm). The composition of solvent A was 0.1% formic acid in water, and solvent B was 0.1% formic acid in acetonitrile. Each sample (total protein 0.5 μg) was applied to the trapping column and flushed with 1% solvent B for 5 min at a flow rate of 15 $\mu\text{L}/\text{min}$. Sample elution was performed at a flow rate of 300 nL/min by increasing the organic solvent concentration from 3 to 40% B over 90 min. All analyses were conducted in triplicate. The precursor ion masses and associated fragment ion spectra of the tryptic peptides were mass measured with a Q-ToF Premier mass spectrometer (Waters Corporation) directly coupled to the chromatographic system.

The time-of-flight analyzer of the mass spectrometer was externally calibrated with NaI from m/z 50 to 1990, with the data postacquisition lockmass-corrected using the monoisotopic mass of the doubly charged precursor of GFP, fragmented with a collision energy of 25 V. The GFP was delivered at 500 fmol/ μL to the mass spectrometer via a NanoLockSpray

interface using the auxiliary pump of a nanoACQUITY system at a flow rate of 500 nL/min. The reference sprayer was sampled every 60 s.

Accurate mass data were collected in data independent mode of acquisition (LC-MS^E) by alternating the energy applied to the collision cell between a low energy and elevated energy state. The spectral acquisition scan rate was 0.6 s with a 0.1 s interscan delay. In the low energy MS mode, data were collected at constant collision energy of 4 eV. In elevated energy MS mode, the collision energy was ramped from 15 to 35 eV during each integration.

Data Processing for DDA Acquisitions. The uninterpreted MS/MS data from the gel-separated and iTRAQ-labeled samples were processed using ProteinLynx Global Server (PLGS) v2.3. The data were smoothed, background subtracted, centered and deisotoped. All data were lockspray calibrated against GFP using data collected from the reference line during acquisition.

Data Processing for Label-Free Acquisitions. The LC-MS^E data were processed using PLGS v2.3. The ion detection, data clustering and normalization of the data independent, alternate scanning LC-MS^E data has been explained in detail elsewhere.³³ In brief, lockmass-corrected spectra are centroided, deisotoped, and charge-state-reduced to produce a single accurately mass measured monoisotopic mass for each peptide and the associated fragment ion. The initial correlation of a precursor and a potential fragment ion is achieved by means of time alignment.

Database Searches. All data were searched using PLGS v2.3 against a *M. silvestris* database (<http://genome.ornl.gov/microbial/msil>). Fixed modification of carbamidomethyl-C was specified, and variable modifications included were acetyl N-terminus, deamidation N, deamidation Q and oxidation M. For the iTRAQ data, variable modifications for the isobaric tags were specified. One missed cleavage site was allowed. Search parameters specified were a 50 ppm tolerance against the database-generated theoretical peptide ion masses and a minimum of one matched peptide.

For the LC-MS^E data, the time-based correlation applied in data processing was followed by a further correlation process during the database search that is based on the physicochemical properties of peptides when they undergo collision induced fragmentation.³⁴ The precursor and fragment ion tolerances were determined automatically. The protein identification criteria also included the detection of at least three fragment ions per peptide, at least one peptide determined per protein and the identification of the protein in at least two out of three technical replicates. By using protein identification replication as a filter, the false positive rate is minimized, as false positive protein identifications, that is, chemical noise, have a random nature and as such do not tend to replicate across injections. This approach rules out systematic search events errors due to the repeated ambiguity of a particular spectrum and the subsequent sequence assignment by a search algorithm, as could be the case with peptide-centric searches.

Protein Quantification Using iTRAQ Labeling. PLGS was also used for quantitative evaluation of MS/MS data generated from the analysis of the iTRAQ-labeled peptides. A relative quantification was conducted using a merged data set comprising the results from the database search. Concentration ratios of iTRAQ-labeled proteins were calculated based on signal intensities of reporter ions observed in peptide fragmentation spectra, with the relative areas of the peaks corresponding to proportions of the labeled peptides.¹⁴

Table 1. Total Protein Identifications for the Three Experimental Approaches

	gel-based	iTRAQ	label-free
Total number of protein identifications	389	384	425
Single-peptide identifications	154	206	4
Proportion of identifications which were from a single peptide	40%	54%	0.9%
Identifications with more than one peptide	235	178	421

Protein Quantification Using Label-Free System. Relative quantitative analysis across conditions was performed by comparing normalized peak area/intensity of each identified peptide. Normalization of the data was conducted by the use of an internal protein digest standard. In brief, peak areas/intensities are corrected using those of the internal protein digest. Intensity measurements are typically further adjusted on those components, that is, deisotoped and charge-state reduced accurate mass retention time pairs, that replicate throughout the complete experiment. Next, the redundant, proteotypic quantitative measurements provided by the multiple tryptic peptide identification from each protein were used to determine an average, relative protein fold-change. The algorithm performs binary comparisons for each of the conditions to generate an average normalized intensity ratio for all matched proteins. Proteins with a likelihood of quantification smaller than 0.05 were considered to be significantly regulated. The entire data set of differentially expressed proteins was further filtered by considering only the identified peptides that replicated two out of three technical instrument replicates. A likelihood of regulation higher than 95%, as reported by the quantification algorithm, was considered.

Results and Discussion

Protein Identifications. Three distinct experimental approaches have been employed in order to provide profiling and quantitative information regarding the proteome of *M. silvestris*. The numbers of proteins identified via each approach are summarized in Table 1. The total number of nonredundant proteins identified, when single peptide identifications are included, is comparable for all three techniques at 389, 384, and 425 proteins, respectively.

Differences arise, however, when looking at the number of peptides per protein identification. There have been questions raised in the literature regarding the validity of identifications performed using a single peptide, so-called 'one-hit wonders', and whether they should be included in the list of proteins identified.³⁵ Of the gel separation identifications, 154 were from a single peptide as are 206 in the iTRAQ experiment. As a proportion of the total proteins identified, these values are 40% and 54%, respectively. This is typical of many results in the literature.³⁶ In the label-free results, of the 425 identifications, only 4 are from a single peptide: proportionally less than 1%. As the label-free analysis is performed in triplicate, only an identification observed in at least two of the three replicates was taken to be valid; therefore, single-peptide identification in label-free data means that a single peptide was found in at least two of three data sets. Proteins identified by each experimental setup are listed in Supplementary Data, Tables 1–5. Figure 1 shows the overlap of protein identifications between the three approaches; 1a uses all data, including

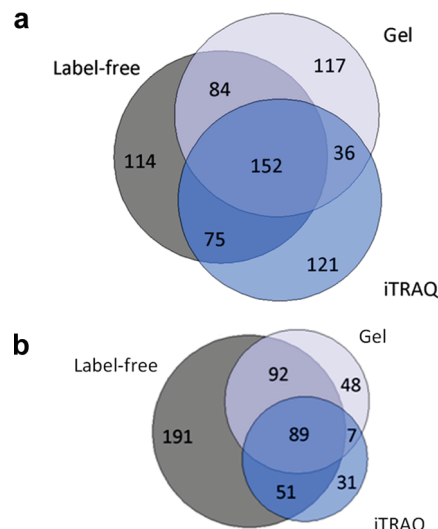


Figure 1. (a) Number of proteins identified by the various experimental approaches, including single-peptide identifications. (b) Proteins identified by the various experimental approaches, identifications based on a minimum of two peptides.

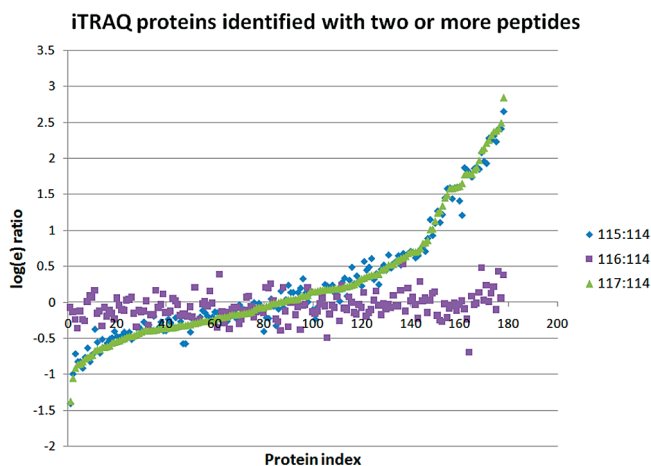


Figure 2. Differential expression as determined by iTRAQ labeling; all tags have been normalized to the 114 label.

single-peptide identifications, 1b illustrates filtered data, with only identifications obtained with two or more peptides. All proteins identified are listed in Supplementary Tables 1–5, giving information on the molecular weight and *pI* of the identifications, and also the number of peptides identified. When including single peptide based identifications, there are a total of 699 proteins identified. Each of the techniques uniquely provides approximately 17% of those identifications. The remaining 49% of the identifications overlapped as shown. To overcome the uncertainty involved in the inclusion of single peptide-based identification, Figure 1b shows the data presented only including identifications made using a minimum of two peptides. This gives a total of 509 protein identifications, of which 9% were unique to the gel-based approach, 6% to iTRAQ, and 38% to label-free. This shows a significant increase in the proportion of unique identifications by the label-free method, a reduction in gel-based unique identifications and a considerable decrease in those uniquely identified by iTRAQ.

A closer inspection of the number of proteins identified with and without the inclusion of single-peptide identifications reveals some interesting observations. As one would expect,

the total number of proteins identified is lower when single-peptide identifications are excluded (509 when excluded, compared to 699 when included), including those identifications common to all three methods (89 when excluded, compared to 152 when included). In contrast, the number of proteins unique to the label-free method, and those common to the label-free and gel methods, has increased. This is because all but four of the proteins identified by the label-free method were done so with two or more peptides, whereas the gel and iTRAQ methods generated a large number of single-peptide identifications. The fact that 152 proteins were independently identified by all three methods provides strong evidence that, although some of these (63 in total) were identified with a single-peptide by one or more technique, they should possibly not be discarded as false-positive identifications. This raises the question as to what should be done with protein identifications based on a single-peptide. While a majority of these are likely to correspond to false-positive identifications, there are a small number that are potentially valid and should also be included in the list of confidently identified proteins, although this is not definitive. Further discussions of results will therefore exclude single-peptide identifications.

Relative Quantification of Identified Proteins. Supplementary Figure 1 shows the 1D SDS-PAGE separation of the *M. silvestris* proteome obtained from different growth conditions. The difference in intensity of staining in a 1D gel is indicative of differential expression, and some representative changes are highlighted. A band on a 1D gel, however, can often contain multiple proteins due to the limitations of the resolving power of this technique. Although the analysis of gel-separated samples provided a comparable number of protein identifications, quantitative analysis using a 1D separation is difficult. Quantification via gel methods is more routinely performed using two-dimensional separations, which were not carried out here. Further results, focusing on differential expression, use only iTRAQ and label-free data.

Figure 2 represents the differential expression of proteins as characterized by iTRAQ labeling, plotted on a \log_e scale; the values are included alongside protein identifications in Supplementary Data, Table 3. Tags 115 and 117, which correspond to methane-grown samples, and tag 116, which corresponds to an acetate-grown sample, were normalized to tag 114, which corresponds to an acetate-grown sample. The values for the 116 sample are clustered close to a line along the x -axis as would be expected since the 114 and 116 samples should be identical. The 115 and 117 samples should also be identical and we would therefore expect good agreement between their ratios, as is observed. This experiment provides a good indication of the reproducibility of the iTRAQ approach. As can be seen by the 115 and 117 trends, distinct up- and down-regulated proteins may be identified in *M. silvestris* when grown under methane as compared to when the organism is grown under acetate. The standard deviation of all the 116:114 ratios is 0.17, providing an indication of what can be considered true up- or down-regulation. If these values are considered to be a normal distribution around a calculated mean of 0, then any proteins with 115:114 and 117:114 ratios within -0.5 and 0.5 cannot be said to be regulated, using the value of three standard deviations to provide filtering parameters. Only those identifications showing ratios outside these values have been accepted as up- or down-regulated.

In the iTRAQ method, samples from different growth conditions are pooled together. Quantification depends entirely upon

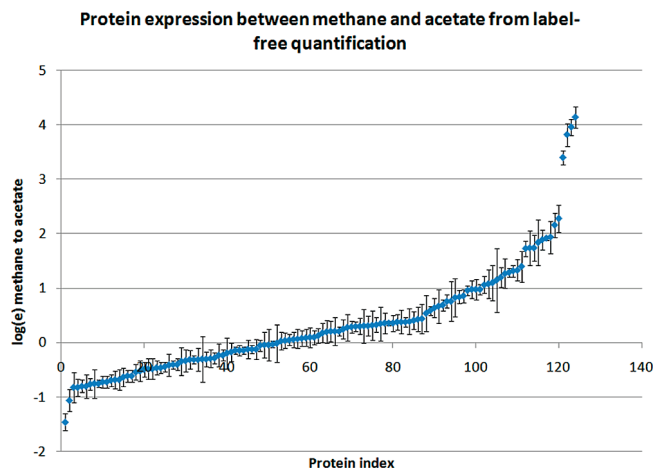


Figure 3. Automated protein-level quantification of regulated proteins using the label-free system; error bars correspond to the automatically generated standard deviation values.

the isobaric tags; if insufficient data is available from the isobaric tags, the protein identification will still be provided in the overall results table, but will not appear in the quantification data. In the label-free system, samples from differing growth conditions are kept separate, so a distinct set of protein identifications is generated for each sample. A total of 231 proteins were identified unique to the acetate-grown sample, 70 were unique to the methane-grown sample and 124 were common to both conditions. Data relating to these 124 proteins were then processed to provide information on relative expression between the samples.

Figure 3 shows the relative protein expression for the regulated proteins (common to acetate and methane substrates) identified using the label-free approach; this is the output from the relative quantification software, which generates peptide signal intensity measurements, using all the peptides identified for any particular protein identification. These represent deisotoped, charge-state reduced and accurately mass measured ion lists, which are used for both qualitative identification and relative quantification.²⁸ \log_e values used as the quantitative measurement can be found in Supplementary Data, Table 6, including indication of proteins assigned to only one of the two growth conditions. Error measurements are automatically generated as standard deviation values, which have been plotted. For an MS^E acquisition, the technical variation with respect to signal intensity has been shown to be 10–15% with highly consistent reproducibility.^{27,38} For the label-free quantitative data, the significance of regulation level was specified at 30%. Hence, 1.3-fold (± 0.30 natural log scale) was used as a threshold to identify significant up- or down-regulation.³⁸ This is typically 2–3 times higher than the estimated error on the intensity measurement.^{27,38,39} Those identifications with relative expression values between -0.3 and 0.3 cannot be taken as regulated; only those identifications outside these values can be said to be regulated.

Both iTRAQ and label-free allow profiling and relative quantitative data to be concurrently collected. The ability to do this, particularly in a high-throughput manner, is desirable but often difficult. In total, 79 confident identifications (i.e., more than one peptide) are common between the two methods, which is a much larger overlap than previous studies comparing methods of quantification.^{14,40} A scatter plot comparing the regulation as assessed by the two methods is shown

Table 2. A Comparison of the Experimental Requirements for Each of the Approaches, and the Information Obtained from the Data Generated

	1D-SDS-PAGE	iTRAQ	label-free
Protein loading	14 μ g	100 μ g per iTRAQ labeling vial; 800 μ g total loading	0.5 μ g for each of 3 technical replicates
Number of overnight steps	2	5	1
Samples to analyze by MS	30–40 fractions	30–60 fractions	1 per growth condition
Reverse-phase LC and MS acquisition	30–40 h	30–60 h	2 h
Total analysis time	4 days	6 days	Less than 3 days
Total instrument time	30–40 h	30–60 h	6 h per sample
Size of data file	300 MB \times 40 (1.2 GB)	300 MB \times 40 (1.2 GB)	6 GB \times 3 (18GB)
Number of proteins confidently identified (with more than one peptide)	235	178	421
Average number of peptides per protein (including single-peptide identifications)	5	5	12
Average sequence coverage	15%	11%	45%

in Supplementary Figure 2a. There is reasonable correlation, with an R^2 value of 0.69, with one distinct outlier. If the overall trend of regulation is compared, all of the common identifications for which quantification data is available are in agreement, bar the outlier. Twenty-one proteins are indicated to be up-regulated in the methane sample compared to acetate and 6 are indicated to be down-regulated; the remaining proteins show no distinct differential expression when filters for both data sets are applied. If the one distinct outlier is removed from the data set, then the correlation improves significantly (R^2 value 0.80), as shown in Supplementary Figure 2b. The outlier, corresponding to the enzyme citrate synthase, presented down-regulation in the methane-grown sample according to the label-free analysis, but up-regulation according to iTRAQ. Interrogation of the raw data showed good correlation between all three replicates of the label-free acquisition in both growth conditions. In the iTRAQ data, however, there was a disparity in the data from the isobaric tags. Five peptides were used for identification, with quantification data available for four of these. Three peptides showed down-regulation in the methane growth condition; the one peptide which indicated up-regulation was the shortest of the five (four residues), the others matching at least eight residues within the assigned MS/MS spectrum. If the short peptide is removed, there is down-regulation of citrate synthase within the filtering parameters, and in-line with the label-free data, suggesting that this was a misassignment by the software. MS/MS spectra of the matched sequences and isobaric tags are shown in Supplementary Figure 3. Although this is only one anomalous data point, it indicates potential problems if single-peptide identifications are used to provide quantitative data from an iTRAQ experiment.

The label-free approach differs from iTRAQ in that each growth condition is analyzed independently, while in iTRAQ, samples from different conditions are pooled together. Of the 425 nonredundant identifications obtained by the label-free method, 231 were unique to the acetate-grown sample and 70 unique to the methane-grown. From these 301 proteins, 54 were also identified by iTRAQ, and were compared with the iTRAQ quantification list. Of these, 25 were distinctly regulated and all showed agreement, that is, were shown by iTRAQ to be up-regulated in whichever growth condition the label-free method had exclusively assigned. This has been represented as a comparative table in Supplementary Data, Table 7. The 29 identifications which fall outside the iTRAQ filtering parameters for accepted regulation levels, as described earlier, have been highlighted.

Comparison of Experimental Approaches. A summary of the methodology for all three experimental systems and the results obtained from each can be seen in Table 2. There is a stark difference in the total amount of protein required for the three setups: up to 1 mg for iTRAQ, 14 μ g for the gel-based method, and less than 1 μ g for label-free. Although the injection amount for the LC-MS/MS analysis is comparable between all three techniques, this does not relate to the total amount of protein required for an adequate data set. In the gel-based and iTRAQ approaches, the amount indicated is necessary to generate enough peptides over 30–60 fractions for MS analysis. With the employed label-free method, the amount loaded directly for LC-MS analysis is sufficient for a full qualitative and quantitative data set. Sample requirement can be an important consideration when performing proteomic studies, as it can be a challenge to generate a suitable amount from biological systems. If less sample is required for a single experiment, additional analyses can be carried out, which will add confidence to the results obtained.⁴¹ It has previously been shown that even three replicate MudPIT experiments may not provide full coverage of all the proteins within a sample.⁴²

An ideal method for proteomic analysis would enable comprehensive and high-throughput studies, making experimental and instrumentation time an important factor when considering which approach to utilize. Both the gel-based and iTRAQ setups require up to 60 h of MS data acquisition time, based upon our chosen number of bands cut from the gel or fractions from the strong cation exchange chromatography, and upon the gradient setup in the reversed-phase chromatography. The analytical time could be shortened by choosing fewer fractions, or reducing the reversed-phase gradient, but this may also reduce peptide recovery and/or separation. The label-free experiments require 6 h of instrument time (2 h per replicate). In addition to this, preparing samples for iTRAQ requires a number of days, including overnight steps. This issue can make the approach less suitable for a routine analysis setup when compared to the label-free method.

The average number of peptides identified per confident protein assignment for the gel-based and iTRAQ analyses is 5, compared to an average of 12 for the label-free method. The gel-based approach gives an average sequence coverage of 15%, higher than the iTRAQ average of 11% which is slightly lower than previous work.² The average sequence coverage for the label-free data is 45%. An increased number of peptides and higher sequence coverage can confer more confidence in identifications obtained.

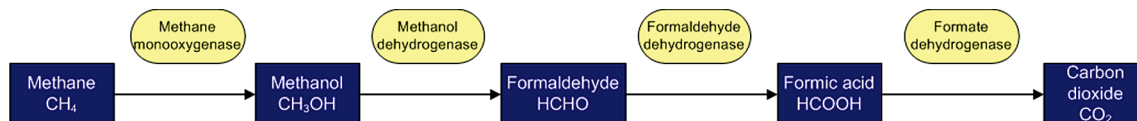


Figure 4. The pathway of methane oxidation in methanotrophic bacteria.

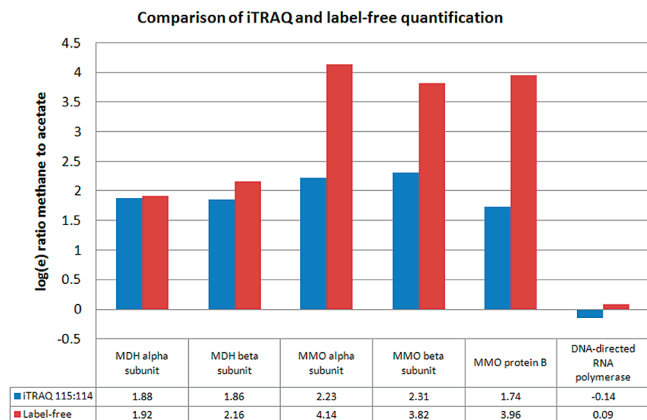


Figure 5. iTRAQ and label-free quantification data for proteins identified as key metabolic enzymes within *M. silvestris*; a housekeeping protein has been included as indication of a biological marker.

Biological Significance of Results Obtained. Bacteria from the methanotroph family utilize a common pathway to process methane in order to use it as a carbon and energy source, an overview of which is given in Figure 4. As the *Methylocella* genus has been recently identified and is relatively uncharacterized, it is difficult to make predictions about potential biochemical changes which would be seen on a growth substrate other than methane. It could, however, be suggested that some down-regulation of the enzymes in the methane oxidation pathway would be seen. Our study identified the key enzymes methane monoxygenase (MMO) and methanol dehydrogenase, with quantitative data from both iTRAQ and the label-free approach indicating a significant down-regulation when *M. silvestris* was grown on acetate. MMO is a multimeric protein with subunits α , β and γ .³⁷ The α and β subunits show up-regulation in the methane growth samples, as does the accessory MMO Protein B. The γ subunit shows significant up-regulation on methane when analyzed by iTRAQ; using our data filtering (more than one peptide, more than one replicate for label-free), this subunit is only seen in the methane growth condition for the gel and label-free analyses, as is the accessory MMO Protein C. There is also up-regulation of the α and β subunits of methanol dehydrogenase in the methane-grown samples, which is the second enzyme in the methanotroph methane oxidation pathway.

The quantification data relating to these enzyme identifications has been shown in Figure 5. DNA-directed RNA polymerase has also been included as a housekeeping protein, and as such should not display up- or down-regulation regardless of growth substrate. Such proteins can provide a check, on the biological level, for the significance of differential proteomic data.

Conclusions

For any integrated proteomics experiment, a number of important issues need to be considered. These include the need for qualitative (profiling) and quantitative information. Confi-

dence in identification and quantification, reproducibility, sample size, instrument time, sample preparation, cost, and sequence coverage are all important factors that need to be taken into account. The ability to place any changes observed into the context of the biological pathways involved remains a crucial aspect of the research. This study has evaluated the potential applicability of a number of common approaches to profiling and differential proteomics. The experiments have been restricted to a proteomics study of cytosolic proteins, and comparable technology platforms were employed. Good agreement was obtained between the commonly utilized iTRAQ labeled experiment, a gel based study and that based on a label-free LC-MS approach. At the profiling level, when considering all identifications, including those based on single peptides, the number of identified proteins was comparable for all three methods. When requiring more than one peptide for identification, the label-free approach gave superior information particularly when coverage was taken into account. Both the iTRAQ experiment and the label-free approach provided relative quantification data sets, and the agreement between the approaches was better than previously observed in comparisons between different quantitative methods.⁴⁰ This is most likely due to the use of comparable instrumentation, as each method employed high-performance liquid chromatography coupled to a Q-TOF tandem MS acquisition. The label-free experiment does, however, have advantages in terms of sample requirement, sample preparation and instrumental time requirements.

A preliminary screen of the protein regulation results for biological significance shows agreement with previous analysis of the regulation of methane monoxygenase in *Methylocella*.⁴³ This, together with the significant number of identifications provided by all three approaches, and the excellent agreement of two quantitative data sets, indicates the potential for further proteomic studies on this methanotroph.

Acknowledgment. Vibhuti Patel thanks Eleanor Blatherwick (Biological Mass Spectrometry and Proteomics Facility, Department of Biological Sciences, Warwick) for assistance with experimental work. Vibhuti Patel and Konstantinos Thalassinou thank Hans Vissers (Waters) for helpful discussions regarding the manuscript.

Supporting Information Available: Tables of protein identifications from gel-based analysis, acetate growth and methane growth; protein identifications from iTRAQ analysis; protein identifications from label-free analysis, acetate growth and methane growth; quantification data from label-free analysis; qualitative agreement between iTRAQ identified proteins and label-free identification assigned to only one growth condition. 1D SDS-PAGE separation of the *M. silvestris* cytoplasmic proteome under different growth conditions; Correlation of quantification data from iTRAQ and the label-free method for identifications using two or more peptides and when the outlier corresponding to citrate synthase is removed; peptide sequences for precursors and corresponding isobaric reporter ions for outlying protein identification citrate synthase.

This material is available free of charge via the Internet at <http://pubs.acs.org>.

References

- Pandey, A.; Mann, M. Proteomics to study genes and genomes. *Nature* **2000**, *405* (6788), 837–846.
- Domon, B.; Aebersold, R. Mass spectrometry and protein analysis. *Science* **2006**, *312* (5771), 212–217.
- Wilkins, M. R. Guidelines for the next 10 years of proteomics. *Proteomics* **2006**, *6* (1), 4–8.
- Ong, S.-E.; Pandey, A. An evaluation of the use of two-dimensional gel electrophoresis in proteomics. *Biomol. Eng.* **2001**, *18* (5), 195–205.
- Washburn, M. P.; Wolters, D.; Yates, J. R. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol.* **2001**, *19* (3), 242–247.
- Ünlü, M.; Morgan, M. E.; Minden, J. S. Difference gel electrophoresis. A single gel method for detecting changes in protein extracts. *Electrophoresis* **1997**, *18* (11), 2071–2077.
- Ong, S.-E. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* **2002**, *1* (5), 376–386.
- Fairwell, T. Sequence analysis of complex protein mixtures by isotope dilution and mass spectrometry. *Biochemistry* **1970**, *9* (11), 2260–2267.
- Kuhn, E. Quantification of C-reactive protein in the serum of patients with rheumatoid arthritis using multiple reaction monitoring mass spectrometry and ¹³C-labeled peptide standards. *Proteomics* **2004**, *4* (4), 1175–1186.
- Sirlin, J. L. On the incorporation of methionine ³⁵S into proteins detectable by autoradiography. *J. Histochem. Cytochem.* **1958**, *6* (3), 185–190.
- Gerber, S. A. Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100* (12), 6940–6945.
- Thompson, A. Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal. Chem.* **2003**, *75* (8), 1895–1904.
- Gygi, S. P. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* **1999**, *17* (10), 994–999.
- Ross, P. L. Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol. Cell. Proteomics* **2004**, *3* (12), 1154–1169.
- Ow, S. Y. Quantitative shotgun proteomics of enriched heterocysts from *Nostoc* sp. PCC 7120 using 8-plex isobaric peptide tags. *J. Proteome Res.* **2008**, *7* (4), p. 1615–1628.
- Sadowski, P. G. Quantitative proteomic approach to study subcellular localization of membrane proteins. *Nat. Protoc.* **2006**, *1* (4), 1778–89.
- Panchaud, A. Experimental and computational approaches to quantitative proteomics: Status quo and outlook. *J. Proteomics* **2008**, *71* (1), 19–33.
- Allet, N. In vitro and in silico processes to identify differentially expressed proteins. *Proteomics* **2004**, *4* (8), 2333–2351.
- Liu, H.; Sadygov, R. G.; Yates, J. R. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal. Chem.* **2004**, *76* (14), 4193–4201.
- Colinge, J. Differential proteomics via probabilistic peptide identification scores. *Anal. Chem.* **2005**, *77* (2), 596–606.
- Rappsilber, J. Large-scale proteomic analysis of the human spliceosome. *Genome Res.* **2002**, *12* (8), 1231–1245.
- Ishihama, Y. Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Mol. Cell. Proteomics* **2005**, *4* (9), 1265–1272.
- Chelius, D.; Bondarenko, P. V. Quantitative profiling of proteins in complex mixtures using liquid chromatography and mass spectrometry. *J. Proteome Res.* **2002**, *1* (4), 317–323.
- Müller, C. Ion suppression effects in liquid chromatography-electrospray-ionisation transport-region collision induced dissociation mass spectrometry with different serum extraction methods for systematic toxicological analysis with mass spectra libraries. *J. Chromatogr., B: Anal. Technol. Biomed. Life Sci.* **2002**, *773* (1), 47–52.
- Wang, W. Quantification of proteins and metabolites by mass spectrometry without isotopic labeling or spiked standards. *Anal. Chem.* **2003**, *75* (18), 4818–4826.
- Radulovic, D. Informatics platform for global proteomic profiling and biomarker discovery using liquid chromatography-tandem mass spectrometry. *Mol. Cell. Proteomics* **2004**, *3* (10), 984–997.
- Silva, J. C. Quantitative proteomic analysis by accurate mass retention time pairs. *Anal. Chem.* **2005**, *77* (7), 2187–2200.
- Silva, J. C. Absolute quantification of proteins by LCMSE: A virtue of parallel MS acquisition. *Mol. Cell. Proteomics* **2006**, *5* (1), 144–156.
- Silva, J. C. Simultaneous qualitative and quantitative analysis of the *Escherichia coli* proteome: a sweet tale. *Mol. Cell. Proteomics* **2006**, *5* (4), 589–607.
- Cheng, F.-Y. Absolute protein quantification by LC/MSE for global analysis of salicylic acid-induced plant protein secretion responses. *J. Proteome Res.* **2009**, *8* (1), 82–93.
- Theisen, A. R. Regulation of methane oxidation in the facultative methanotroph *Methylocella silvestris* BL2. *Mol. Microbiol.* **2005**, *58* (3), 682–692.
- Link, A. J. Direct analysis of protein complexes using mass spectrometry. *Nat. Biotechnol.* **1999**, *17* (7), 676–682.
- Geromanos, S. J. The detection, correlation, and comparison of peptide precursor and product ions from data independent LC-MS with data dependant LC-MS/MS. *Proteomics* **2009**, *9* (6), 1683–1695.
- Li, G.-Z. Database searching and accounting of multiplexed precursor and product ion spectra from the data independent analysis of simple and complex peptide mixtures. *Proteomics* **2009**, *9* (6), 1696–1719.
- Veenstra, T. D.; Conrads, T. P.; Issaq, H. J. What to do with one-hit wonders? *Electrophoresis* **2004**, *25* (9), 1278–1279.
- Breci, L. Comprehensive proteomics in yeast using chromatographic fractionation, gas phase fractionation, protein gel electrophoresis, and isoelectric focusing. *Proteomics* **2005**, *5* (8), 2018–2028.
- Colby, J.; Dalton, H. Some properties of a soluble methane monooxygenase from *Methylococcus capsulatus* strain Bath. *Biochem. J.* **1976**, *157* (2), 495–497.
- Vissers, J. P. C.; Langridge, J. I.; Aerts, J. M. F. G. Analysis and quantification of diagnostic serum markers and protein signatures for gaucher disease. *Mol. Cell. Proteomics* **2007**, *6* (5), 755–766.
- Chambery, A. Qualitative and quantitative proteomic profiling of embryonic stem cells by means of accurate mass LC-MS analysis. *J. Proteome Res.* **2009**, *8* (2), 1047–1058.
- Usaita, R. Characterization of global yeast quantitative proteome data generated from the wild-type and glucose repression *Saccharomyces cerevisiae* strains: the comparison of two quantitative methods. *J. Proteome Res.* **2008**, *7* (1), 266–275.
- Karp, N. A.; Lilley, K. S. Design and analysis issues in quantitative proteomics studies. *Proteomics* **2007**, *7* (1), 42–50.
- Durr, E. Direct proteomic mapping of the lung microvascular endothelial cell surface in vivo and in cell culture. *Nat. Biotechnol.* **2004**, *22* (8), 985–992.
- Trotsenko, Y. A., et al., Metabolic Aspects of Aerobic Obligate Methanotrophy. In *Advances in Applied Microbiology*; Academic Press: New York, 2008; pp 183–229.

PR900080Y