Multidimensional Liquid Chromatographic Separations for Proteomics

Presented by

Karl Burgess

to

The University of Glasgow

for the degree of

Doctor of Philosophy

April 2008

Division of Biochemistry and Molecular biology

Faculty of Biomedical and Life Sciences

University of Glasgow

…we do what we must because we can
for the good of all of us except for the ones who are dead
but there's no sense crying over every mistake
you just keep on trying until you run out of cake
and the science gets done and you make a neat gun
for the people who are still alive…

…Now these points of data make a beautiful line
and we're out of beta, we're releasing on time
so I'm GLaD I got burned
think of all the things we learned
for the people that are still alive…

…anyway this cake is great, it's so delicious and moist
look at me still talking, when there's science to do
when I look out there it makes me glad I'm not you
I've experiments to run, there is research to be done
on the people who are still alive
and believe me I am still alive
I'm doing science and I'm still alive…

From ‘Still Alive’ by Jonathan Coulton

1. Acknowledgements

Proteomics is the study of phenotype and therefore the effect of the environment on an organism. As such, despite the declaration that follows the acknowledgements, this thesis is the culmination of the effect of the myriad interactions with family, colleagues and friends on the author during, and before, the project. And, while I was the vehicle through which these interactions were made manifest, credit should be equally applied to those who influenced and assisted the research.

1.1. Academic Acknowledgements

I would therefore like to thank my supervisors, Dr. Andrew Pitt and Dr. Kenneth Cook, whose unwavering support and enthusiasm were fundamental to my own enjoyment of the research, and whose near limitless knowledge of their own fields meant that the training I received on this project was second-to-none. Much of the funding, advice and instrumentation used on the project was supplied by Dionex corporation, and I am therefore indebted, again to Dr. Cook, but also to Dr. Remco Swart, Dr. Alex Adam, Robert van Ling, Bas Dolman, Dr. Gurmil Gendeh, Lloyd Billingham, Matthew Morse, Evert-Jan Sneekes, Soren Dantoft and especially Dr. Terry Sheehan for his support and greenlighting of the funding. I am also indebted to Dr. Julia Smith, Dr. Laura Main, Dr. Annabelle McIntyre and Dr. Ralf Hartmer at Bruker Daltonics for their assistance and the early release software that I was privileged to test. Dave Miller, Ian Reah, Martin O'Gorman and Susan Mason at NonLinear Dynamics have been excellent partners in a collaboration on development of the Progenesis LC-MS software. Dr. Neil Inglis and Prof. David Smith at the Moredun Institute, Edinburgh were also wonderful collaborators on rapid peptide separations using monolithic columns. The FT-ICR MS expertise of Dr. Colin MacKay and Stefan Weidt at the SIRCAMS facility in Edinburgh allowed rapid progress to be made in the field of top-down analysis. In relation to which, I would like to thank the other principle members of the RASOR consortium, and especially Prof. Walter Kolch and Dr. Nick Morrice, for their advice and support. I am particularly indebted to Dr. Martin Lynch, Dr. Bernard Ramsahoye, Dr. Christina Nuala, Dr. Christian Preisinger and Dr. Alex von Kriegsheim who provided me with the biological samples used in the project. Fahsai Kantawong, Marek Kuzbicki and Dr. Kamburapola Jayawardena provided essential practical assistance with labelling technologies and gel electrophoresis. Finally, I am indebted to Kit-Yee Tan for providing me with the thesis guidelines, Dr. Richard Goodwin and Dr. Richard Burchmore, who proofread much of this manuscript, as well as providing a fantastic level of support during the project.

1.2. Personal Acknowledgements

I would like to thank my parents, grandparents and the remainder of my antecendants for providing the genetic material as well as (in the former case) most of the initial environmental effects instrumental in the generation of my current phenotype. In

addition I would like to thank (in no particular order) Kit-Yee Tan, Kirsty Steele, Ulf Trapp, Katrina Thomson, Michael Lang, Emma Carrick, Pia Totten, Matthew Wylie, Neil Inglis (another one), Richard Goodwin, Richard Burchmore, Christina Nuala, Samuel Burchmore, Annie Burchmore, Colin MacKay, Heather Allingham, Robert Pavulans, Kurt Dereli, Jason Mitchell, Amanda Bradley, Lorraine Aitchison, Chiew Hoon Neo, Fahsai Kantawong and Sam Lawn for occasionally dragging me out of the basement and forcing me to enjoy myself. If I haven't mentioned you here, it's probably due to thesis-induced brain damage, and not because I'm singling you out for omission.

This project could also not have been completed without the almost constant application of all the music on my mp3 player, and I would therefore like to thank James Maynard Keenan, Hope Sandoval, Ihsahn, James Euringer, Robert Smith, Tori Amos, Xytras, Vorphalak, Layne Staley, Tom Morello, Peter Murphy, Andrew Taylor, Mikael Stanne, Sakis Tolis, Ian Brown, Brian Warner, Saul Williams, Stephen Jenkins, Byron Roberts and Krystoffer Rygg for their services to Science, and especially Jon Crosby, for making the evening of the day I finished writing this thesis a fantastic one.

Dedication:

To Colonel Karel Václav Falk, who dispensed with the usual children's stories and went straight for subatomic physics.

2. Declaration

I hereby declare that the thesis that follows is my own composition, that it is a record of the work done by myself, and that it has not been presented in any previous application for a higher degree.

Karl Burgess

3. Abstract

Sample complexity is one of the key challenges facing contemporary proteomic analysis. A variety of methods is commonly employed to reduce this complexity, both at an intact protein and digested peptide level. For complex lysates containing many thousands of proteins, orthogonal (mutually independent), multidimensional separation methods must be employed to provide sufficient resolution to characterize the appropriate number of different species. The most common of these methods are two dimensional gel electrophoresis (2DGE) of proteins and multidimensional liquid chromatographic separation (MuDPIT) of peptides, which rely on isoelectric focusing followed by mass separation in the former, and ion exchange followed by reversed phase separation in the latter. These methods have significant drawbacks in terms of sample bias, sample preparation and reproducibility, and therefore a new methodology that combines the positive aspects of both separation technologies in an automatable, reproducible form is highly desirable.

New developments in column technology have allowed rapid improved-resolution separation of intact proteins in complex samples, coupled to improved methodology for peptide and protein identification. The separation of complex protein mixtures using both on- and off-line 2D liquid chromatography using derivitized polystyrene-divinylbenzene (PS-DVB) pellicular ion-exchange resins and PS-DVB monolithic reversed-phase columns is described. Proteolytic digestion of the fractions followed by rapid liquid chromatography - tandem mass spectrometry was used to complete the analysis. An alternative methodology, relying on direct analysis of the second dimension eluents by top-down (mass spectrometric analysis of intact proteins) methodology, using an Apex IV 12 T Fourier-transform ion cyclotron resonance mass spectrometer (Bruker Daltonics) and an HCT ion trap (Bruker Daltonics) equipped with electron transfer dissociation has allowed in-depth analysis of intact proteins. Sample types investigated to establish the utility of the methodology include bacterial lysates (*Bordetella parapertusis*, and *Escherichia coli*), a eukaryotic parasite (*Leishmania donovani*), and transformed human cell lines.

These developments lead to a multidimensional intact protein separation methodology suitable for small-sample proteomics (minimum effective protein load of 200 µg) and good reproducibility (1.5% variation in the ion exchange dimension, 0.5% variation in the reversed phase dimension). Analysis of the digested fractions gave good coverage of the proteome as well as a high predicted dynamic range, capable of detecting proteins with codon adaptation index scores ranging from 0.22 to 0.99, out of a logarithmic scale from 0 to 1. Proteins representing low (8 kDa) and high (500 kDa) molecular mass and extremes of predicted pI were identified, as well as a number of membrane proteins. Resolution of the overall protein separation was such that single protein species often occurred in one or two fractions for both the ion-exchange and reversed phase separations, with the fractions varying in complexity. Separation of modified proteins in the ion-exchange dimension demonstrated separation of isoforms.

Most analyses coupled anion-exchange chromatography in the first dimension to monolithic reversed phase separation in the second dimension. To provide alternative first dimension methodologies for specific sample types, external gradient chromatofocussing (based on separation by isoelectric point) and high-pH ion-pair reversed phase were evaluated as additional techniques. In particular the pISEP (CryoBioPhysica) technique for chromatofocussing constituted an effective orthogonal separation methodology for separation, and its use in a proteomics context was compared to that of the original anion-exchange, reversed phase technology.

To rapidly analyse the large number of fractions generated by the technique, bottom-up protocols required improvements in the throughput of peptide separations. PS-DVB monoliths were employed for rapid peptide separations and conditions for analysis were evaluated and optimized. The implementation of a parallel 200 µm monolith system for tryptic peptide separations ensured minimal sample loss and improved sample throughput with little loss of sensitivity. For simple mixtures, reversed-phase separation times could be reduced to a few minutes without significantly affecting data content, although rapid scanning capability is essential due to the very narrow peak widths.

Quantitation is of paramount importance to any proteomic technique, and liquid chromatographic separation of intact proteins provides flexibility for differential analysis of complex samples. Three categories of quantitative analysis were evaluated for two dimensional intact protein separation by liquid chromatography: ultraviolet-absorbance maps, isotopic labeling and label-free computational analyses. UV-absorbance maps of wild-type and pentamidine resistant *Leishmania donovani* were compared using standard two dimensional gel electrophoresis analysis software and resulted in the identification of a small number of quantitiative differences. Modifications to the isotope coded affinity tag (ICAT) and isotope tags for relative and absolute quantitation (ITRAQ) protocols were developed to allow protein labeling prior to separation, and the related methodologies ExacTag, isotope coded protein label (ICPL) and stable isotope labels for amino acids in culture (SILAC) were evaluated. Finally, label-free techniques have been employed for protein quantitation by liquid chromatography (LC)/Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR-MS).

A particular niche for liquid chromatographic separation of intact proteins is in top-down analysis, in which mass spectrometric analysis of intact proteins is performed. The separation technology was directly coupled to high-resolution FT-ICR MS for analysis of standards, mixtures, and cellular lysates, leading to the identification of an intact *Leishmania* protein and post-translational modification (PTM) mapping of histone H4. In addition, the recently developed electron transfer dissociation (ETD) fragmentation methods coupled with proton transfer reaction (PTR) for charge reduction allowed analysis of standards and isotopically labeled cellular lysates using an ion trap instrument.

In summary, we have developed a general proteomic methodology with unique flexibility as well as applicability to top-down analysis. It has been applied to multiple complex samples in a variety of analysis conditions including quantitation methodologies and state-of-the-art mass spectrometric techniques. In general, the method equals other separation methods in terms of protein identification rates, is substantially more reproducible and automated, but is more time-consuming. Currently, two dimensional intact protein separation by liquid chromatography using polystyrene-divinylbenzene-based columns is particularly applicable to top-down

analysis of heavily modified proteins. However, there is considerable remaining room for optimization and improvement of methodologies, and further development will enhance the technique for general use.

Table of Contents

List of Abbreviations

2DGE	Two Dimensional Gel Electrophoresis
2DLC	Two Dimensional Liquid Chromatography
AC	Alternating Current
ALS-PAGE	Acid Labile Surfactant-PolyAcrylamide Gel Electrophoresis
ANOVA	Analysis of Variance
BCA	Bicinchoninic Acid
BSA	Bovine Serum Albumin
C18	Octadecyl Silica
CAI	Codon Adaptation Index
CHAPS	3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate
CHCA	Alpha-cyano-4-hydroxycinnamic acid
CID	Collision Induced Dissociation
cICAT	Cleaveable ICAT
CIEF	Capillary IsoElectric Focussing
CMC	Critical Micelle Concentration
CRK-L	CRK-Like Protein
ddH_2O	Double Distilled Water
DiGE	Differential Gel Electrophoresis
DMEM	Dulbecco's Modified Eagle Medium
DMF	Dimethyl Formamide
DNA	Deoxy-Ribonucleic Acid
ECD	Electron Capture Dissociation
EGF	Epidermal Growth Factor
EMBL	European Molecular Biology Laboratory
EMBOSS	European Molecular Biology Open Software Suite
ESI	Electrospray Ionisation
ETD	Electron Transfer Dissociation
FA	Formic Acid
FAB	Fast Atom Bombardment
FASTA	FAST-ALL
FFT	Fast Fourier Transform
FT	Fourier Transform
GELC	SDS-PAGE followed by digestion and LC-MS
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HETP	Height Equivalent to a Theoretical Plate
HFBA	Heptafluoro Butyric Acid
HPLC	High Performance Liquid Chromatography
ICAT	Isotope Coded Affinity Tag
ICPL	Isotope Coded Protein Label
ICR	Ion Cyclotron Resonance
IEX	Ion Exchange
IMAC	Immobilised Metal Affinity Chromatography
IRES	Internal Ribosome Entry Site
IRMPD	Infra-Red Multi Photon Dissociation
iTRAQ	Isotope Tagging for Relative and Absolute Quantitation

ISD	In-Source Dissociation
KMP	Kinetoplastid Membrane Protein
LC	Liquid Chromatography
LIT	Linear Ion Trap
MALDI	Matrix Assisted Laser Disorbtion Ionisation
MGF	Mascot Generic Format
MIF	Migration Inhibitory Factor
MOWSE	Molecular Weight Search
MuDPIT	Multidimensional Protein Identification Technique
MS	Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
m/z	Mass to Charge Ratio
NCBI	National Center for Biotechnology Information
NGF	Nerve Growth Factor
NHS	*N*-Hydroxysuccinimide
NOG	N-Octyl Glucoside
NPS	Non Porous Silica
OMP	Outer Membrane Protein
ORF	Open Reading Frame
PAGE	Poly-Acrylamide Gel Electrophoresis
PBS	Phosphate Buffered Saline
PEEK	Poly Ethyl Ethyl Ketone
PERL	Pathologically Eclectic Rubbish Lister
pISEP	pI Separation
PMF	Peptide Mass Fingerprint
PSD	Post-Source Dissociation
PS-DVB	Polystyrene DiVinyl Benzene
PTM	Post-Translational Modification
RF	Radio Frequency
RNA	Ribonucleic Acid
RP	Reversed Phase
RPMI	Roswell Park Memorial Institute
RSCU	Relative Synonymous Codon Usage
RSD	Relative Standard Deviation
SAX	Strong Anion Exchange
SCX	Strong Cation Exchange
SDS	Sodium Dodecyl Sulphate
SILAC	Stable Isotope Labelling by Amino Acids in Cell Culture
SIMS	Secondary Ion Mass Spectrometry
TCA	TrichloroAcetic Acid
TIC	Total Ion Chromatogram
ToF	Time of Flight
UV	Ultraviolet
WAX	Weak Anion Exchange
WCX	Weak Cation Exchange
XIAP	X-Linked Inhibitor of Apoptosis

1. Chapter 1: Introduction

1.1. Proteomics

The Human Genome Project and its sister projects for other organisms have been a tremendous success. Vast amounts of data pertaining to the 'blueprints' of living organisms have been obtained [1]. Comparative genomics has not only provided information about the genes involved in pathologies such as bacterial infections, metastasis and teratogenesis, but has also allowed us to predict the coding sequences that will be translated to proteins. Following on the heels of the genome projects, the next step has been the development of *proteomics* [2]. Whereas the focus of the genome projects was the identification of every gene in a genome, the intent of proteomics is to study every protein in the mixture of all the proteins expressed by a particular genome. Whereas genomics studies an organism at the level of the genotype, proteomic techniques allow us to directly study the transient phenotype of an organism - the products of the genome when stimulated by the environment at a particular instant.

The study of the proteome presents us with several technical challenges: the complexity and heterogeneity of the proteome, the instability of individual proteins, and the difficulty of performing quantitative analyses. The template for the proteome, the genome, contains a vast number of protein coding sequences. The human genome, for example contains approximately 30,000 genes, most of which will be translated to protein. Complexity is increased at the transcript level by splice variants. Additionally, once translated from mRNA, proteins are commonly *post-translationally modified* by factors such as acetylation [3], methylation [4], glycosylation [5], phosphorylation [6], truncation events [7] and sulphation [8], vastly increasing the complexity of the array of proteins expressed even in a single cellular state.

Proteins perform the thousands of different tasks, such as structural maintenance, molecular trafficking and transcriptional regulation, required to maintain a functioning cell, and are required to be heterogeneous in structure and physiochemical properties. This provides many difficulties for researchers. Not all proteins will: remain soluble under the same conditions, remain stable and intact under the same conditions, or be effectively analysed by the same instrumentation. A further complication is that many proteins have half lives of a few minutes or hours [9], and the structure of a protein is also carefully

controlled within the cell, as denatured proteins are usually unable to perform their normal functions. In addition, the dynamic range of proteins is also a challenge as expression levels range from less than ten to millions per cell [10, 11], leading to the signals from the abundant proteins often swamping those from rare ones.

Data derived from proteomic analyses can be enhanced by direct comparison of multiple phenotypes. For this reason, comparative quantitation (reviewed in [12]) is an important aspect of proteomics. Obtaining abundance data from a heterogeneous, extremely complex mixture is not trivial, and several methods have been employed with variable success. Despite this, proteomics has developed into a highly successful area of research, which provides the unique ability to identify and quantify hundreds or thousands of proteins simultaneously. Indeed, proteomic studies have provided invaluable insight into many important areas of molecular biology, for example the functionality of the centrosome, [13], flagellar motility in bloodstream trypanosomes [14] and protein degradation using the ubiquitin ligase assembly [15].

1.2. Proteomic Analysis Techniques

Attempts to overcome the challenges of heterogeneity, instability and dynamic range of proteins lead to the two main requirements of any technique that approaches a 'proteomic' analysis technique: the ability to reduce sample complexity to manageable levels, and the ability to identify proteins once they are separated. The most popular methods for separating proteins are gels and chromatography columns, and in both cases, the principle of their action is to split a complex mixture of proteins into multiple lower complexity fractions. The most effective tool for protein identification has become the mass spectrometer [16] coupled to in-silico genome analysis, e.g. the MASCOT program [16]. In simple terms, a mass spectrometer is a device for detecting the number of ions present with a particular mass. In essence, mass spectrometers separate or filter ions by mass (and are described in more detail in section 1.6).

1.3. Protein Separation

'And the king said, "Bring me a sword." So a sword was brought before the king. And the king said, "Divide the living child in two, and give half to the one and half to the other." Then the woman whose son was alive said to the king, because her heart yearned for her son, "Oh, my lord, give her the living child, and by no means put him to death." But the other said, "He shall be neither mine nor yours; divide him."'

1 Kings 3

The 'gold standard' methods of current protein separation technologies are currently two-dimensional sodium dodecyl sulphate poly-acrylamide gel electrophoresis (SDS-PAGE)

(2DGE) (first described in [17], examples of the technique are described in [18-20]), and shotgun two-dimensional liquid chromatography (2DLC) of peptides [21-23], epitomised by the multidimensional protein identification technique (MuDPIT) approach of Yates [24].

1.3.1. 2DGE

Protein samples to be analysed by 2DGE are separated in the first dimension by isoelectric focusing (separation by based on the isoelectric point of protein species), usually using immobilized pH gradients, and in the second by transferring the separated proteins to a conventional SDS poly-acrylamide gel electrophoresis gel for separation by size. Proteins migrate to specific spots on the gel [17] and can be visualized by staining with a dye (e.g. coomassie blue, sypro orange or silver staining). After staining, individually focused gel 'spots' can be excised, subjected to proteolytic digestion and identified using mass spectrometry. 2DGE offers high resolution separation, with more than 2,500 spots being potentially visualized on a single 2DE gel, and while the number of proteins identified often falls well short of this number [25-27] it possesses a wide dynamic range [20] and most importantly it maintains information on splice variation and post translational modification [28].

The vast majority of proteomics experiments in the literature were performed using 2DGE. Despite this there are a number of issues which limit the value of this approach [29-31]. In practise, 2DGE is difficult to reliably reproduce and automate, time consuming to run, manually intensive, and biased towards high abundance 'average' proteins (isoelectric point of 4-7, mass between 10,000 and 100,000). Finally, there are a number of issues with the separation of hydrophobic proteins [32], especially membrane proteins, due to poor solubility in non-ionic lysis buffers and inefficient transfer from the isoelectric focussing strip to the SDS-PAGE gel.

1.3.2. 2DLC

Shotgun two-dimensional liquid chromatography (2DLC) uses two orthogonal liquid chromatography separations to separate peptides. Protein samples are first proteolytically digested and the peptide species thus generated are usually separated by (cat)ion exchange followed by reversed phase chromatography (described in section 1.5.4). The reversed phase eluent is usually directly coupled to an electrospray mass spectrometer (described in section 1.6.1.1), and the analysis of the eluted peptides is performed on-line [24]. Shotgun 2DLC is easily automated, but the digestion prior to separation greatly increases the complexity of the mixture that needs to be separated. In addition, peptides from post-

translationally modified or differentially spliced proteins are eluted along with unmodified ones, making it much more difficult to identify modified peptides and relate these to the distribution of species in the intact protein, although quantitative approaches such as metabolic labeling (SILAC [33]) and isotope coded tags (cICAT [34], iTRAQ [35], ^{18}O labelling [36]) are assisting in unravelling these complexities. Finally, peptides from high-abundance proteins are spread throughout the separation space, potentially masking the detection of lower abundance proteins.

An ideal approach, that would combine the ease of automation of liquid chromatography with the retention of key PTM information, would be high resolution separation of intact proteins using multi-dimensional liquid chromatography, followed by top-down (direct analysis of intact proteins) or bottom-up analysis (analysis of proteolytically digested proteins). This approach has, until recently, been hampered by the lack of robust, high-resolution chromatographic methods for protein separation that are suitable for use with the small size and high complexity of proteomic samples (see for example [37]). Recently, new chromatography resins and methodologies have been introduced that overcome some of these difficulties. Separation by chromatofocussing (described in detail in section 4.1.2.2) followed by reversed-phase separation has been commercially available for some time and has been used successfully in a number of studies, (e.g. [38-40]). Other liquid-phase multidimensional separation techniques that have been applied to proteins include: capillary isoelectric focussing (CIEF), based on separation of molecules by isoelectric point using a pH gradient generated between the CE cathode and anode [41]; 2-dimensional capillary electrophoresis, based on separation of species by mass-to-charge ratio [42, 43], chromatofocussing coupled to reversed phase (RP) separation (see section 1.5.4.1) on PS-DVB monoliths and digestion on a MALDI (see section 1.6.1.2) target plate [44], size exclusion - RP [45, 46] separation; and weak anion exchange (see section 1.5.4.2) fractionation followed by conventional RP separation, coupled to both top-down and bottom up approaches [47].

1.4. Protein Identification

The method commonly employed for identifying proteins using a mass spectrometer is 'peptide mass fingerprinting' (PMF) [48]. Several naturally occurring proteases, such as trypsin or asp-N, have very specific points at which they cleave protein chains (for further details see section 1.7.2). It is therefore possible to predict, given the sequence of a protein, what fragments it will be digested into for any given protease, and as the accurate masses of the constituent amino acids can easily be calculated from their chemical composition, it

is also possible to generate a list of theoretical masses generated by the proteolytic cleavage of a protein. This is known as an 'in-silico digest'.

By physically digesting a protein and analysing it using a mass spectrometer, it is possible to obtain a spectrum containing a set of fragment masses. By matching these masses with the predicted masses generated by in-silico digestion, it is possible to generate a statistical score of the closeness of a match of the theoretical masses to the predicted masses. Identification of an unknown protein is a matter of predicting the masses generated by proteolytic cleavage for all predicted protein sequences encoded by an organism, usually obtained from a genome database, and searching experimentally derived masses against the theoretical list.

To provide additional data it is common to fragment the peptides in the MS, which can provide sequence data (see section 1.6.3.2). By matching fragments generated experimentally to empirically determined rules for fragmentation, it is possible to obtain additional statistical confirmation of the identity of a protein.

However, there are three drawbacks to the PMF method. The first is the previously mentioned post-translational modifications – these will affect the mass of some of the peptides, immediately reducing the statistical significance of a match. This can be accounted for if the modification is known, or by taking the potential for any modification into account. The latter is not usually feasible due to the increase in computing power required for identification, as each additional modification allowed for requires an additional mass to be searched for, for each peptide in the in-silico digested database. In addition, the statistical score required for confident matching increases for every potential modification included because the number of potentially matching random peptides rises proportionally along with the number of potential mass matches. It is therefore normal to select a small number of expected modifications to include in the search.

Secondly, many peptides are rarely detected by MS. Some are too large or small to be identified within the mass range of a particular mass spectrometer, some simply fail to ionise under the conditions applied in the MS source, and some may be affected by sample constituents that suppress their ionisation.

The third drawback is that if the observed peaks represent digestion fragments of more than one protein, *i.e.* the sample contained a mixture of proteins, then the statistical confidence is compromised. Reducing the complexity prior to MS therefore becomes the

key to identifying proteins and is the driving force behind the research contained within this thesis.

1.5. Liquid Chromatography

The term 'liquid chromatography' (LC) was coined by Tswett in 1906 in his book 'Physical chemical studies on chlorophyll adsorptions' [49]. Since then, LC has grown to become one of the most commonly employed separation technologies available to researchers. The term 'chromatography' comes from its initial discovery as a means of separating different coloured dyes using paper *i.e.* 'paper chromatography'. A drop of black ink (which is a blend of different coloured inks mixed to produce black) or macerated spinach, which contains several chlorophyll pigments, was placed at one end of a piece of paper. When the strip of paper was dipped in a pool of solvent, different dyes travelled at different rates up the paper, separating as they go. Separation is a function of capiliary action and the interaction of the dye molecules with the paper and solvents.

Modern LC works in much the same way as paper chromatography. A sample of solution (the ink in the example above) passes through a column containing a packing material (the paper), and the rate at which each sample component travels is based on their physiochemical properties. The liquid passing through the column is known as the *mobile phase* and the packing material of the column is known as the *stationary phase [50]*. Stationary phases which separate based on different physiochemical properties are available. The most common are 'reversed phase' and 'ion exchange' columns [51].

1.5.1.1. Factors in Chromatographic Separations

Chromatographic separation is affected by a number of factors: the solvent flow rate, the amount of packing material, the viscosity of solvents, the temperature of the system, mass transfer and diffusion.

The viscosity of the solvent significantly affects the backpressure which the pump must overcome to perform chromatography. In standard reversed phase separations, the gradient from aqueous to organic results in significant changes to viscosity as the separation progresses. This is due to the lower viscosity of the organic solvents (e.g. acetonitrile) used to perform the separation. At high concentrations of acetonitrile, the viscosity of the overall solvent decreases significantly, resulting in lower backpressure. The pump rate on flow-sensing systems must be adjusted to compensate for the lower backpressure, which would otherwise result in higher flowrates. Temperature affects the viscosity of the

solvent, with higher temperatures leading directly to a reduction in backpressure. Due to the higher temperature stability of PS-DVB monolithic columns, temperature is raised significantly to ameliorate the effects of the higher flow resistance in these columns [52]. Most traditional silica resins, degrade rapidly at temperatures higher than 60ºC (reviewed in [53]).

A common column preparation methodology for improved resolution is an extension of the total column length. This directly increases the number of 'theoretical plates' (see section 1.5.2) as well as the overall capacity of the column. However, the increase in separation capability is coupled with an increase in backpressure that, unless ameliorated by other methods such as increasing the temperature or reducing the flow rate, may result in failure of the pump. The type of stationary phase, as well, affects the type of separation obtained. The octadecyl silica (C18) stationary phase used for most reversed phase separations is so-called due to the long aliphatic carbon chains bonded to the silica surface. Due to the size of this lipid layer, it takes some time for samples to transition from the mobile phase to the stationary phase. The transition from mobile phase (the buffer liquid) to stationary phase (the functional group associated with the packing material of the column) is known as *mass transfer*, and slow or poor mass transfer results in broad peaks and poor chromatography. The length of the C18 chain results in strong interaction with hydrophobic sample components. Less hydrophobic conventional stationary phases, such as C8 and C4 are used to separate especially hydrophobic sample components that would fail to elute under organic conditions in normal chromatographic time scales.

Diffusion of sample in the mobile phase is a significant factor in peak broadening, and diffusion into and out of derivatised porous silica structures is one of the reasons that traditional stationary phases provide relatively poor resolution compared to monolithic stationary phases. Increasing flow rate is a means of substantially increasing the resolution of a separation as it minimises diffusional mixing. However, in situations where separations are applied directly to MS, the increase in solvent flow may result in poor desolvation, leading to poor sensitivity.

1.5.2. Chromatography Theory

The process by which chromatographic separation occurs has been the source of considerable debate [54-59], but there are two principle theories to explain the separation properties of columns: Plate Theory (formulated by Synge and Martin in 1947 for which

they received the Nobel Prize in Chemistry [60]) , and Rate Theory, (formulated by Van Deemter in 1956 [61]).

1.5.2.1. The Plate Theory of Chromatography

Plate theory describes a separation column as a series of imaginary plates which defines an arbitrary unit of separating power (see Figure 1).

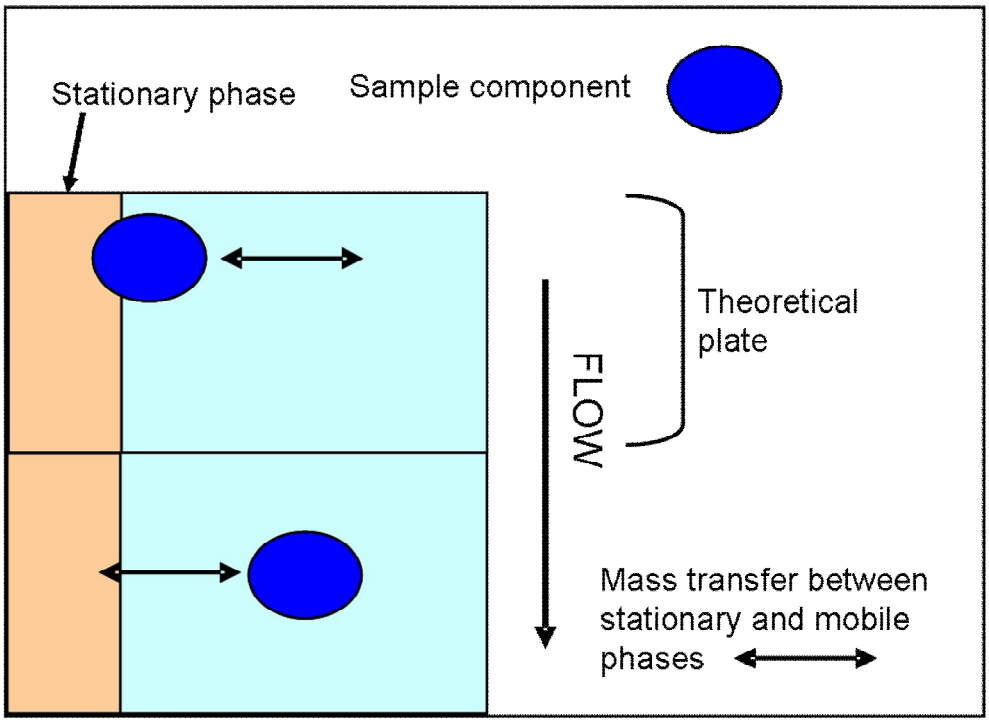


Figure 1: Plate theory. As the mobile phase flows through the column from plate to plate, sample components (blue) are in equilibrium between interaction with the stationary phase (orange) and the mobile phase (green). The length of time a component spends in the mobile phase dictates its speed of movement down the column.

In each plate sample can freely move from the mobile phase to the stationary phase. The length of time a sample component spends in each phase is an equilibrium whose point of balance is dependant on the properties of the component and the composition of the mobile phase. The mobile phase also constantly moves down through the column, carrying sample from one plate to the next. In this way, sample constituents which interact weakly move rapidly from plate to plate, since they spend most of their time in the mobile phase. Conversely, components that bind more strongly to the stationary phase will take longer to pass through the column - their speed being dependant on how strongly they interact. This theory lead to the development of several equations to describe the properties of a column, including the 'height equivalent to a theoretical plate (HETP)'. HETP is calculated by dividing the length of the column by the total number of theoretical plates in the column. The number of theoretical plates in a real column is given by N, where:

$$N = 5.55_tR^2/w_{1/2}^2 = 16_tR^2/w^2 \text{ [60]}$$

- $_tR$ is the retention time of a given chromatographic peak
- $w_{1/2}$ is the half-height peak width
- w is the baseline peak width

1.5.2.2. The Rate Theory of Chromatography

Plate theory has a number of limitations that make it an inexact model of the behaviour of real columns. One example is that mass transfer between the stationary phase and the mobile phase is assumed to be instantaneous. The shape of peaks eluted from the column is also affected by the motion of the components within the stationary phase. Rate theory is a development of plate theory which retains the concept of imaginary separation 'plates' [61], but uses significantly different calculations to determine the height equivalent to a theoretical plate. This calculation is known as the Van Deemter equation. The Van Deemter equation describes the separation of a column as follows:

$$HETP = Au^{1/3} + B/u + Cu \text{ [61]}$$

Where:

- A is the *eddy diffusion*, i.e. the path of the components through the column which will be random and is affected by the packing of the stationary phase;
- B is the *longitudinal diffusion*, i.e. the motion of the particles back and forth in the column, this is minimised by *u*, the flow rate of mobile phase through the column;
- C is the *resistance to mass transfer*, i.e. the ease at which particles move between the stationary and mobile phases, this is multiplied by the flow rate since any resistance to mass transfer will be minimised by a low flow rate that allows time for movement of particles between stationary and mobile phases.

For practical purposes, high resolution columns will have very small plate heights, packed with a stationary phase exhibiting low mass transfer resistance even at higher flow rates. These properties are highly dependant on the material used to construct the stationary phase.

1.5.2.3. Chromatography Systems

Separations are normally performed using a *chromatography system*, or *high performance liquid chromatography* (HPLC) system. A typical system consists of a gradient pump, autosampler, column and detector (see Figure 2).

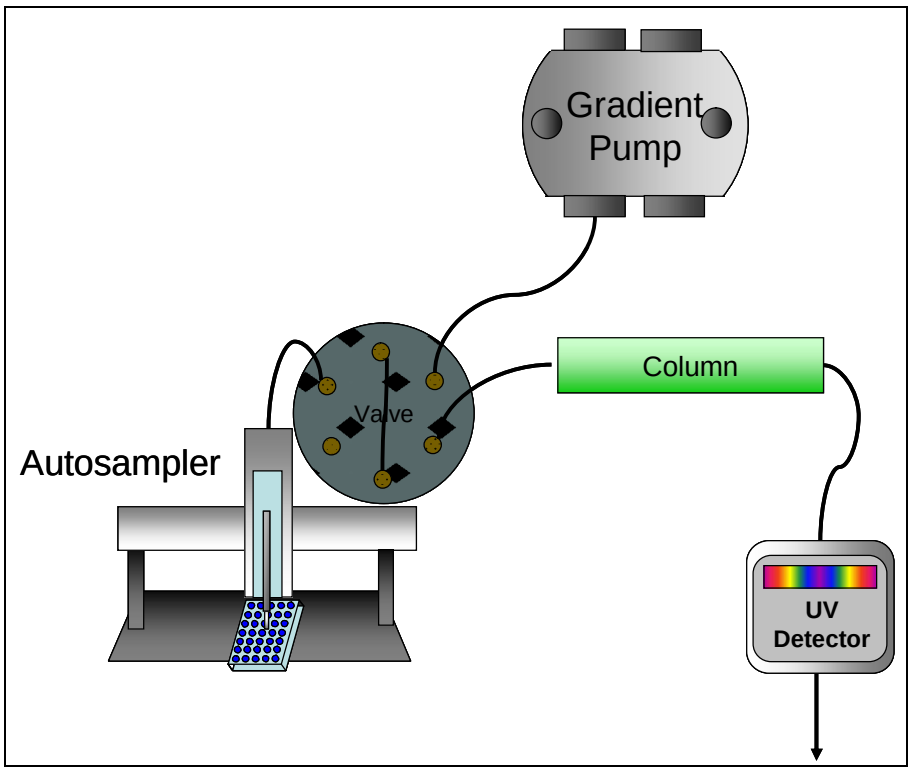


Figure 2: Schematic of a basic HPLC system. Samples are loaded by the autosampler, and the valve allows the aspirated sample to enter the HPLC flow. Separation is performed on the column when a gradient is applied to it from the gradient pump. Eluted components are detected by the UV detector.

Gradient pumps are accurate pumps capable of supplying a range of flow rates at high backpressures (up to 400 bar is common, although some commercially available systems are capable of backpressures up to 870 bar [62]). They are also equipped with the ability to mix two or more solvents using a 'proportioning valve'. Although some direct flow systems exist, most HPLC systems employ a 'split' whereby an average solvent flow is supplied by the pump which is then split into both a waste flow and the desired flow rate, e.g. a 1/100 split from a 300 µL/min flow rate would result in a 3uL/minute flow rate applied to the column and 297 µL/min diverted to waste. An autosampler is a robotic system designed for automatic application of samples. Normally incorporating an injection valve and syringe for aspiration, they are coupled into a loading flow, which may be either derived directly from the gradient pump, or from an additional pump. The most important part of a chromatography system (and the only essential one – samples may be loaded by hand, and gravity fed columns are common) is the separation column. Columns are tubes (commonly either fused silica or poly-ethyl-ethyl-ketone (PEEK)) packed with stationary phases in the form of porous beads [63], non-porous beads [38] or microporous polymer

[64], and range from 20µm in diameter to several centimetres. A variety of chemistries exist (as described in 1.5.4) to assist in separation of specific types of compound. Although for proteomics work, a mass spectrometer is the primary detector, it is useful to apply a secondary method of detection to allow analysis of pre-MS dimensions, or as a confirmation of HPLC function. The most common alternative detector is based on UV absorbance, and is a non-destructive method of detecting proteins and peptides prior to fraction collection and/or analysis by MS. The separate components of a chromatography system are connected by either fluidic connections or valves. HPLC valves allow more complex separations to be performed automatically, for example a valve may be employed to select one or other of two separation columns, or to switch between high flow rate trapping and low flow rate elution of samples. More complex systems may include an isocratic pump for loading samples, additional gradient pumps to allow multidimensional separations and a column oven to maintain temperature stability.

1.5.3. Chromatography Columns

1.5.3.1. Stationary Phase Fabrication

The interaction chemistry of a column can be a property of the packing material itself, or, more commonly, can be a property of functional groups covalently linked to the packing material through a process known as 'derivatisation' [65]. A common example is the 'C18' column, which consists of a piece of fused silica tubing packed with silica beads. The beads are coated with a hydrophobic layer of saturated 18-carbon chains. This is an example of a 'reversed phase' column, which separates on the basis of hydrophobicity - sample components will interact with the hydrophobic carbon chains with a strength proportional to their hydrophobicity. Historically, reversed phase columns are a development from 'normal phase' columns. Normal phase chromatography is based on the use of a polar stationary phase (usually silica). Compounds interact with the stationary phase with a strength proportional to their polarity, until they are eluted by a gradient running from non-polar to polar in nature. The nomenclature, 'reversed' phase results from 'reversing' the nature of the stationary phase (coating the silica with a hydrophobic surface), and the gradient (running from polar to non-polar as the gradient progresses) [66].

A more recent innovation has been the development of 'monolithic' columns [67]. Monolithic columns consist of a piece of tubing packed with a single unit of stationary phase bonded to the walls of the tube. A 'Crunchy™ Bar'-like microporous, rigid structure provides a continuous flow path through randomly generated microchannels. In general,

monolithic columns are prepared by first flushing silica tubing with NaOH, water and additional solvents depending on the type of stationary phase to be packed. Helium is then forced through the tubing to clear any remaining dust. The monolithic stationary phase material is mixed, degassed, and pumped into the tubing under pressure, where it either polymerises naturally or requires additional energy input to fix. Pores are formed with the the activation of porogens (pore forming agents) included in the original mixture.

Two types of monolithic column for proteomics work are currently commercially available: silica based, and polystyrene divinyl benzene (PS-DVB).

1.5.3.2. Silica Based Monolithic Columns

Silica based monolithic columns are a recent development from silica bead resins [64]. The monomer unit used to produce silica monoliths is one of a variety of alkoxysilanes (originally tetramethoxysilane [68]). This is polymerised under conditions of acid and base catalysis. Pores result from syneresis (separation of the mixture into solid and liquid phases, followed by drying) [64] and are a mixture of large (macropores) and small (micropores). High flow rates are provided by the macropores, and diffusion occurs rapidly into the micropores, where most chromatography occurs. Unlike PS-DVB resins, the silica must be derivatised to obtain reversed phase chromatography characteristics. Derivitisation of the packing material to a reversed phase resin is performed by forcing a solution of n-octadecyltriethoxysilane through the column under pressure for several hours [64].

1.5.3.3. Polystyrene DiVinyl Benzene Monolithic Columns

PS-DVB monoliths consist of polystyrene monomer chains, crosslinked by divinyl-benzene [69]. The strength of the polymer is dependant on the amount of crosslinking and hence on the percentage of divinyl-benzene in the mixture.

The production of most PS-DVB monoliths is based on the method described by Premstaller [70]. The fused silica must first be activated by covalent modification of the surface using vinyltrimethoxysilane. The tubing is then filled with a mixture of the monomer (styrene), the cross-linker (di-vinyl benzene), a polymerising radical former (azobisisobutyronitrile (AIBN)), and the inert porogens (1-octanol and tetrahydrofuran (THF)). Polymerisation is initiated using heat or ultraviolet radiation, and the ends of the capillary are pressurised to ensure a uniform structure.

The final three dimensional structure of a monolith is a mass of crosslinked globules, forming a continuous micropore network (see Figure 3).

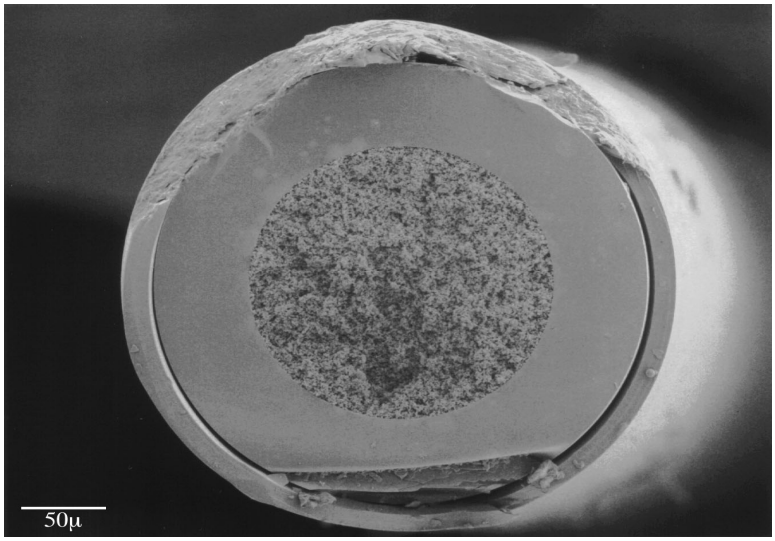


Figure 3: Three dimensional structure of a PS-DVB monolithic column. The polymerisation mixture, along with porogenic compounds, is packed into the outer shell of activated fused silica (the thick solid inner band). The outer band is a polyacrylamide coating applied to fused silica capilary to enhance robustness and flexibility. Polymerisation initiation results from UV or thermal energy, and the presense of the porogens results in the microporous structure, which provides a large surface area for separations (Ivanov, North Eastern University, Boston).

The principle advantages to PS-DVB monoliths are that, because PS-DVB has reversed phase chemistry as a fundamental property (the surface consists of benzene rings linked by carbon chains), the separation takes place on the surface of resin [71]. Additionally, unlike particulate silica resins, where sample must diffuse to access stationary phase buried in large pores, the microporous structure of PS-DVB monoliths minimises the effects of diffusion path lengths [52]. The rigidity of the microporous structure also allows greater stability and pressure tolerance, allowing higher flow rates, shorter separation times and faster reconditioning [71]. PS-DVB is also highly temperature stable, with a nominal operating temperature of 60ºC, which also improves the column resolution [52].

PS-DVB monolithic columns can also be derivatised by substituting chloromethylstyrene for styrene in the polymerisation mixture. This provides a suitable functional group for substitutions, allowing the preparation of PS-DVB monoliths with a variety of column chemistries [72].

1.5.4. Column Chemistries

There are several types of LC used in proteomics. The most common are reversed phase (RP) chromatography [73] and ion exchange (IE) chromatography [74]. Other more specialist techniques, targeted at phosphoprotein or phosphopeptide enrichment are immobilised metal affinity chromatography (IMAC) [75] and titanium oxide chromatography [76]. There are also a number of types of affinity chromatography, which rely on the use of specific ligands in order to trap desired compounds, such as protein A columns [77], lectin columns [78], and antigen columns [79].

1.5.4.1. Reversed Phase Chromatography

Reversed phase chromatography is based on interaction between weakly to strongly hydrophobic sample molecules and a range of hydrophobic stationary phases. Typically, when performing gradient separations, the mobile phase runs from polar (aqueous) to non-polar (organic) solvent, resulting in the elution of increasingly hydrophobic compounds as the concentration of organic solvent increases. Retention on a column is not simply a matter of sample binding to the stationary phase but rather that each sample component is in an equilibrium between solution and stationary phases and hence the speed of a compound's passage down the column is inversely proportional to its hydrophobicity and to the polarity of the solvent. For example, at a given concentration of organic solvent, hydrophobic species will pass down the column more slowly than hydrophilic species [73]. As the organic solvent concentration of the mobile phase increases, the rate that sample components pass down the column increases, such that even very hydrophobic species are eluted quickly. Under such conditions, more hydrophilic species would pass immediately through the column with little or no interaction with the stationary phase, and it is for this reason that gradient separations are applied, so that elution is gradual for the majority of sample components.

Reversed phase separations are usually the final type of separation performed prior to or concomitant with mass spectrometry, and it can be used as an additional dimension of separation after any other chromatograph technique. This is for two reasons: the first is that reversed phase mobile phases can be mass spectrometry compatible; the second is that reversed phase trapping (*i.e.* loading the sample onto the column in aqueous solvent such that chromatography occurs at an imperceptibly slow pace and the sample is retained on the column) can be performed to remove contaminants from a sample, thus combining separation with a useful 'cleaning' step prior to analysis. Non-volatile ionic salts are a

contaminant found at high concentrations in fractions obtained from common first dimension separations such as ion exchange chromatography. These must be excluded from electrospray MS instruments due to ion suppression resulting from their sequestration of the charges in electrospray droplet [80] (the electrospray process is described in more detail in 1.6.1.1).

1.5.4.2. Ion Exchange Chromatography

Ion exchange chromatography is a group of separation techniques based on charge, using four different types of stationary phase [81]. The stationary phases available can be split into two classes: anion and cation exchange chemistry and can be further categorised as 'strong' or 'weak' anion or cation exchangers. Some examples of ion exchange chemistries include quaternary ammonium groups (for strong anion exchange), diethylaminoethyl groups (for weak anion exchange), sulphonic acid groups (for strong cation exchange) or carboxymethyl (for weak cation exchange) groups.

Ion exchange chromatography is based on coulombic interaction of charged analytes with a charged stationary phase [81]. Interaction occurs between the ion exchange stationary phase and overall oppositely charged analytes, and the strength of the interaction is proportional to the charge density of both the analyte and stationary phase chemistry. Weak exchangers are derivitised with groups with a pKa closer to the buffer pH than strong exchangers, and will therefore interact less strongly with equivalently charged species. For example, a weak anion exchanger (the TSKgel DEAE-2SW) may have a pKa of 11.2, whereas a strong anion exchanger (such as the TSK-GEL SAX) may have a pKa of 12.5. Under standard buffer conditions, e.g. pH 8, analytes with a pI of 4 will be negatively charged and will interact strongly with both stationary phases. An analyte with a pI of 8 or higher, however, will be neutrally or positively charged, and will not interact with either stationary phase.

In situations where there is competition for stationary phase functional groups, the strongly charged components will outcompete weaker binding analytes on weak ion exchange resins, whereas the more extreme pKas of strong ion exchangers will tend to interact with more analytes more strongly.

Weakly ionic mobile phases are used to load sample onto an ion exchange column, and elution is performed using a strongly ionic mobile phase, often molar concentrations of sodium chloride, potassium chloride or ammonium sulphate. Careful control of the mobile phase pH is very important, as interaction with the column's stationary phase is dependant

on the charge of the sample components being opposite to that of the stationary phase. For this reason, high pH (~8.0) mobile phase is used for anion (positively charged functional group) exchange chromatography, and low pH (~3.0) mobile phase is used for cation (negatively charged functional group) exchange chromatography.

A pH of 8 for anion exchange columns is commonly employed due to the susceptibility of the fused silica tubing used in HPLC to etching in strongly alkaline conditions, but for systems employing PEEK tubing (which is generally unreactive) throughout, this limit can be increased. In contrast, as described in 1.7.1.4, acidic conditions are a means of precipitating intact proteins from solution, which limits the utility of cation exchange columns for most classes of protein. For peptides, cation exchange columns are commonly used, as the buffer pH is similar if not identical to the acidic conditions applied to the subsequent reversed phase separation.

Ion exchange chromatography is typically used in the first dimension of a 2DLC separation, prior to reversed phase chromatography, because the high salt conditions used for elution interfere with the ionisation of mass spectrometers (as described in section 1.5.4.1).

1.6. Section 2: MS Methods for LC

In section 1.5, LC methods for the separation of proteins prior to mass spectrometry were described. Once samples are separated, they are then analysed by MS, and although there are a number of different types of MS, the principles by which they operate fall into a small number of categories. Analysis by all current MS instruments requires the ionisation of the analyte, the most common methods for which are electrospray ionisation (ESI) [82] and matrix assisted laser disorbtion ionisation (MALDI) [83]. Once ionised, samples are passed to the mass analyser itself, which is a device for separating or filtering ions. Finally, ions are detected by a variety of methods, such as microchannel plates [84] or charge-coupled devices [85]. Each of these methods have applications to the analysis of separated peptides and proteins and will be discussed in detail in this section.

1.6.1. Methods for Biomolecule Ionisation

1.6.1.1. Electrospray Ionisation

Electrospray ionisation was first described by Malcolm Dole in 1968 [82]. It was theorised in 1882 by John Strutt (Lord Rayleigh) that evaporation of a charged droplet of solvent will eventually lead to a point where the repulsion of like charges overcomes the surface

tension of the droplet, and the droplet will disperse [86]. This charge-based dispersion of droplets is the basis of electrospray.

The sample is sprayed from a needle which is maintained at an electric potential relative to an orifice through which ions in the gas phase can enter the MS. Solvent is removed by a combination of its own volatility and the addition of 'curtain gas' – nitrogen gas pumped across the orifice under pressure. In addition, the spray from the needle is usually aimed slightly away from the orifice, and the potential difference between the needle and the curtain plate is employed to attract charged species into the orifice. Nebulising gas coaxial to the spray needle is also commonly employed to improve the spray of ions, and heating of the curtain plate is also commonly used to assist in the evaporation of solvents (see Figure 4). Each peptide may pick up one or more protons (or lose protons to generate anions, if the mass spectrometer is running in negative mode) from the solvent, resulting in multiple charge states, which are detected as separate peaks on the m/z scale of the mass spectrometer, and the number of protons sequestered by a peptide is generally proportional to its size. Individual peaks are often described as the mass-to-charge ratio, followed by the number of charges, e.g. 576(3+).

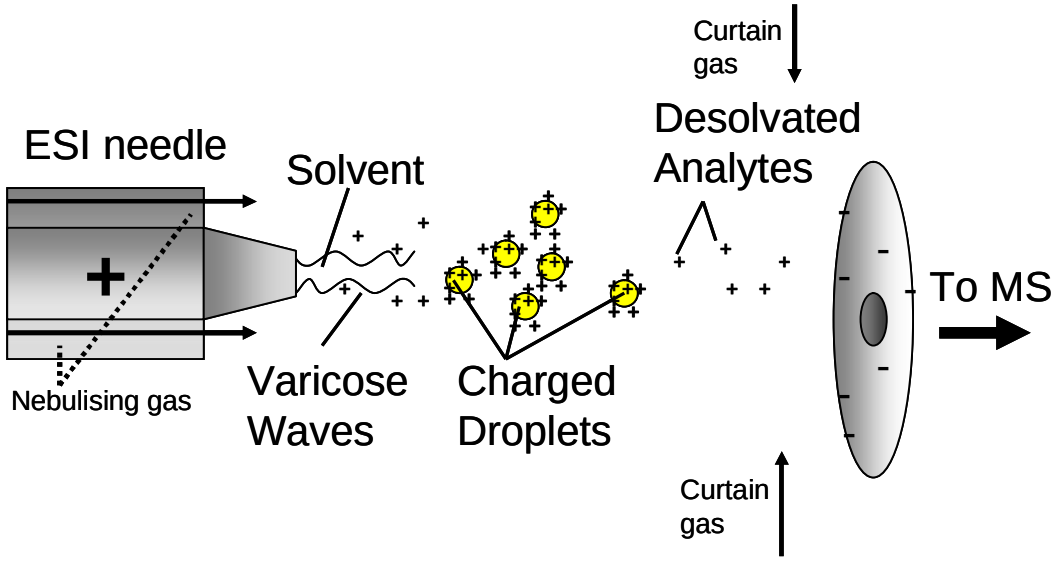


Figure 4: Diagram of an electrospray source. Sample liquid is forced through the front of the needle, where it is exposed to a high voltage. The voltage applied to the flow destabilizes it, first resulting in varicose waves and then droplets as the charge overcomes the surface tension of the fluid. Nebulising (or sheath) gas assists the spray of ions, and (often heated) inert gas is pumped across the orifice to improve dissolution of solvent molecules.

Normally, the electrospray needle is coupled directly to the outlet of an LC system. The mass spectrometer will collect a mass spectrum over a short time period (ranging from a few tenths of a second (for some quadrupole and time of flight type instruments) [63] to 10 seconds (for some Fourier transform ion cyclotron resonance mass spectrometers (FT-ICR-MS))) [87]. Each mass spectrum will contain mass peaks reflecting the composition of the eluent from the LC system at that moment. Thus it is possible to obtain a 'total ion chromatogram' (TIC) which is analogous to the UV chromatograms available using UV detectors by adding together the total intensities of all detected masses at a particular time point. The resolution of the TIC is dependant on the number of spectra obtained, which is dependant on the 'duty cycle' of the mass spectrometer – the time it takes to obtain a single spectrum. It is important to bear in mind that the resolution of the trace refers to the number of acquisition points and hence the smoothness of the chromatogram, as opposed to the resolution of the peaks produced by the LC which is dependant on the properties of the LC system and connections to the mass spectrometer.

1.6.1.2. Matrix Assisted Laser Desorption/Ionisation (MALDI)

"Riding the screaming crest of fettered lons, I shall bring my crystalline chaos where order reigns!" Byron Roberts, from 'Return to the Praesidium of Ys'

Some early success was observed [88] with the analysis of small peptides suspended in a glycerol matrix using secondary ion mass spectrometry (SIMS), and fast atom bombardment (FAB) [89] with beams of charged argon atoms [90]. These methods, however, are generally too energetic to effectively ionise proteins and peptides, and the requirement of a glycerol substrate in peptide samples results in significant ion suppression. An alternative, and now very popular method for ionising proteins and peptides, MALDI, was developed by Tanaka [83] as a soft ionisation alternative to the high energy ionisation methods that preceded it. In MALDI, sample is placed on a metallic target plate, along with *matrix*. The matrix (generally alpha-cyano hydroxycinnamic acid for small peptides and sinapinic acid for larger peptides and proteins) is a photo-ionisable compound that readily transfers its protons to neighbouring species. The matrix is ionised by a laser, and the peptides, once ionised by proton transfer, form a cloud that can be accelerated by a voltage applied to a grid or ring electrode to impel them into the mass spectrometer (see Figure 5). The mechanism for proton transfer has yet to be clearly elucidated but may involve photochemical ionisation (ionisation of the matrix, followed by collision with an analyte molecule), cluster ionisation (where analytes are surrounded by matrix crystals that are photoionised and desolvated, transferring their charge to the

enclosed analyte) or by an energy transfer induced disproportionation method (where a proton is shared between the matrix and a basic site on a biomolecule during crystallisation) [91]. Once in the mass spectrometer, a spectrum of ions is obtained (as decribed in section 1.6.2).

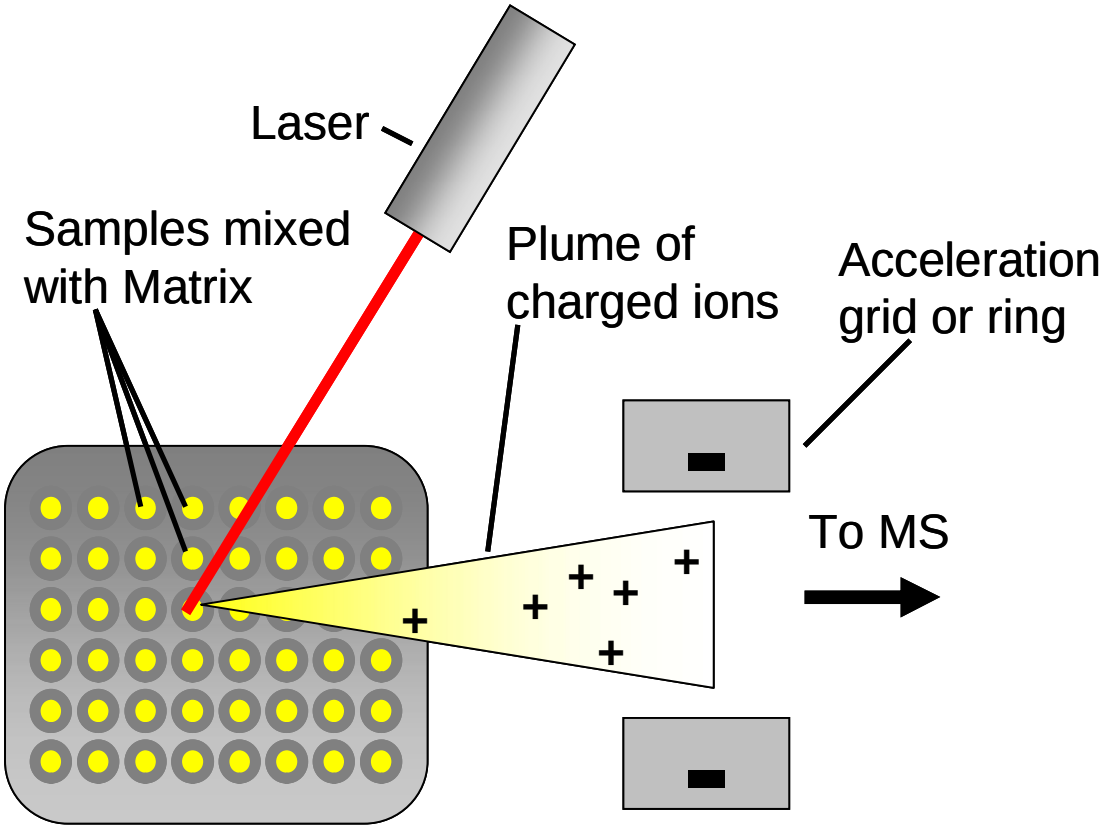


Figure 5: Schematic of the MALDI process. Sample mixed with matrix is spotted in an array on a conductive sample plate. Samples are vapourised and ionized by a laser, forming a plume of ions, which are then accelerated using an oppositely charged grid or ring electrode.

LC/MALDI is a way of sequentially placing LC fractions and matrix on a target plate such that each spot contains a 'snapshot' of the eluent from an LC for a particular time period. It is then possible to extract the total ion count from each spot and develop a TIC of overall ion intensities in an LC run. Unlike electrospray, resolution of an LC/MALDI TIC is not dependant on mass spectrometer duty cycle, but on the number of spots.

Advantages of LC/MALDI over electrospray are dependant on the quality of the software supplied for analysis. For example, unlike electrospray MS software, precursor selection (wherein peaks in a particular spot are selected for fragmentation by MS/MS (see section 1.6.3.2)) can be performed in a number of ways, the most effective of which is to determine the relative abundance (peak area) of a particular peptide throughout an LC/MALDI run and fragment only the most abundant. This is a unique benefit of LC/MALDI that cannot be reproduced by electrospray without repetition of the sample injection. Disadvantages are that analysis times are increased, as the complete separation of

the sample is performed separately to the MS sample analysis. Additionally, the duty cycle of an ESI MS is usually less than 10 seconds on modern instruments, but it is rare for the LC resolution of LC-MALDI spotting to be so rapid, leading to poorer resolution of the TIC, which can have implications for the area calculations involved in quantitation experiments (see section 1.7.3.2).

1.6.2. Mass Spectrometers

Once the sample is introduced into the mass spectrometer, the masses of the sample constituents must be measured. There are a number of very different processes to separate or filter ions based on mass, including quadrupolar mass filters [92], time-of-flight apparatus [93] and quadrupolar ion-traps [92]. Each of these separates ions based on mass and collides them with a detector, which 'counts' the number of ions detected at a particular time by detecting fluctuations in the charge on a detector plate (see section 1.6).

Two other types of mass spectrometer, the Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometer [94] and the orbitrap [95], combine separation and detection by the use of Fast Fourier Transforms (FFT) of ion packet waveforms.

1.6.2.1. Quadrupole Mass Filter

The quadrupolar mass filter was invented by Wolfgang Paul in 1953 [92]. An ion is impelled into the filter by a low DC voltage, and while in the filter experiences no additional forward momentum. On exit of the filter, ions generally impact a detector, unless the quadrupole is a component of a hybrid MS (see section 1.6.2.7). The quadrupole itself consists of two sets of two linked rods. A constant positive DC voltage is applied to one set of rods and a equal magnitude negative voltage is applied to the other set of rods. A supplementary RF voltage is applied to both sets of rods, normally in the MHz range. Due to the asymmetric field thus generated, the profile of the ions passing through the filter will be elliptical. The oscillation of the RF voltage acts to either stabilise or destabilise the trajectory of ions passing through the filter, depending on their mass to charge ratio (m/z) (see Figure 6). While stabilised ions spiral directly through the filter, destabilised ions undergo one of two behaviours: when an ion's m/z is greater than that of the selected ion, it will be too massive to be pulled back to the centre of the quadrupole, resulting in a spiral path of increasing diameter, eventually resulting in collision with one of the rods, or, in the

case of an ion of lower m/z than that selected, the ion will be deflected significantly by the RF voltage and collide rapidly with one of the rods. Additionally, the width of the window of ion stability is inversely proportional to the magnitude of the DC voltage applied to the quadrupoles, so that by reducing the DC voltage, a very wide mass range may be allowed through the filter, or the selectivity may be improved by increasing the voltage [96].

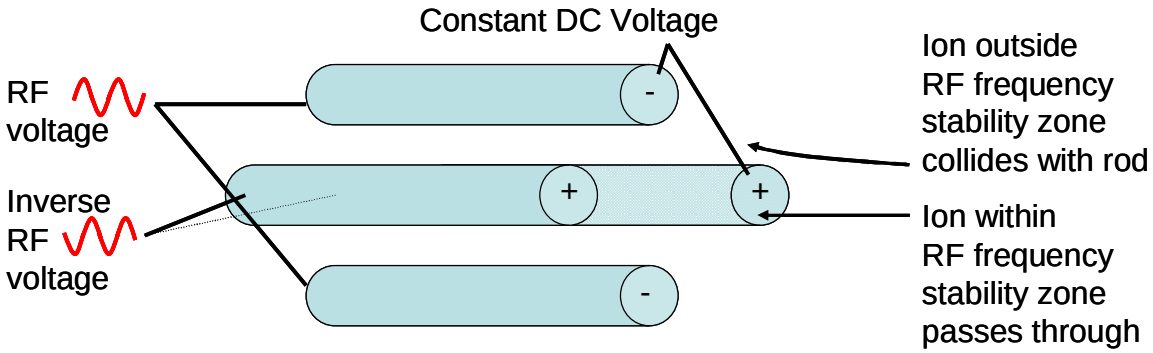


Figure 6: A quadrupole mass filter consists of two sets of two oppositely linked rods, with a constant (positive) DC current on one pair, and a constant (negative) DC current on the other. Filtration is mediated by an RF voltage applied to all four rods, in one phase on one pair of rods, and the opposite phase on the other pair of rods. Stability of a given ion's trajectory is dependant on its m/z value and the RF voltage. Unstable ions collide with one of the rods prior to passing through the filter. A particular RF voltage will provide stability for ions with a particular m/z and instability for other ions.

To optimise resolution for scanning an m/z range, the DC voltage is kept high enough so that the mass filtering will be very selective, but not so high that selected ions may be destabilised by fluctuations in field strength resulting from electrical noise. The particular m/z value selected for can be altered by changing the amplitude of the RF voltage or the frequency of the switching, and an m/z range is scanned b,y gradually increasing the RF voltage or switching frequency so that the stable m/z range moves from high to low or vice versa [96].

1.6.2.2. Spherical Ion Trap Mass Spectrometers

The quadrupolar ion trap was also invented by Wolfgang Paul. It consists of two parabolic end caps and a ring electrode that surrounds them. Filling of the trap is performed by means of a gate voltage, which opens to allow ions to enter the trap, and closes once the trap is filled, to eliminate space charging effects caused by excessive charge density resulting from too many ions. Once in the trap, ions oscillate at a 'secular frequency' based on an RF voltage (the 'fundamental RF'), and in some instruments a DC voltage, applied to the trap through a ring electrode (see Figure 7). Ions are trapped by the quadrupolar electrostatic field generated by the RF voltage on the ring electrode. Trapping is enhanced by the use of 'damping gas', normally helium, which is used to cool trapped ions and

prevent their escape. To eject an ion from the trap, a supplementary RF voltage is applied to the end caps of the trap. If this RF voltage is appropriate for the 'secular frequency' of a specific ion, the ion will be excited, and will be ejected from the trap. Lower amplitude RF voltages can be applied to a specific ion to increase its energy and the force of its collision with the damping gas, resulting in low energy fragmentation [92].

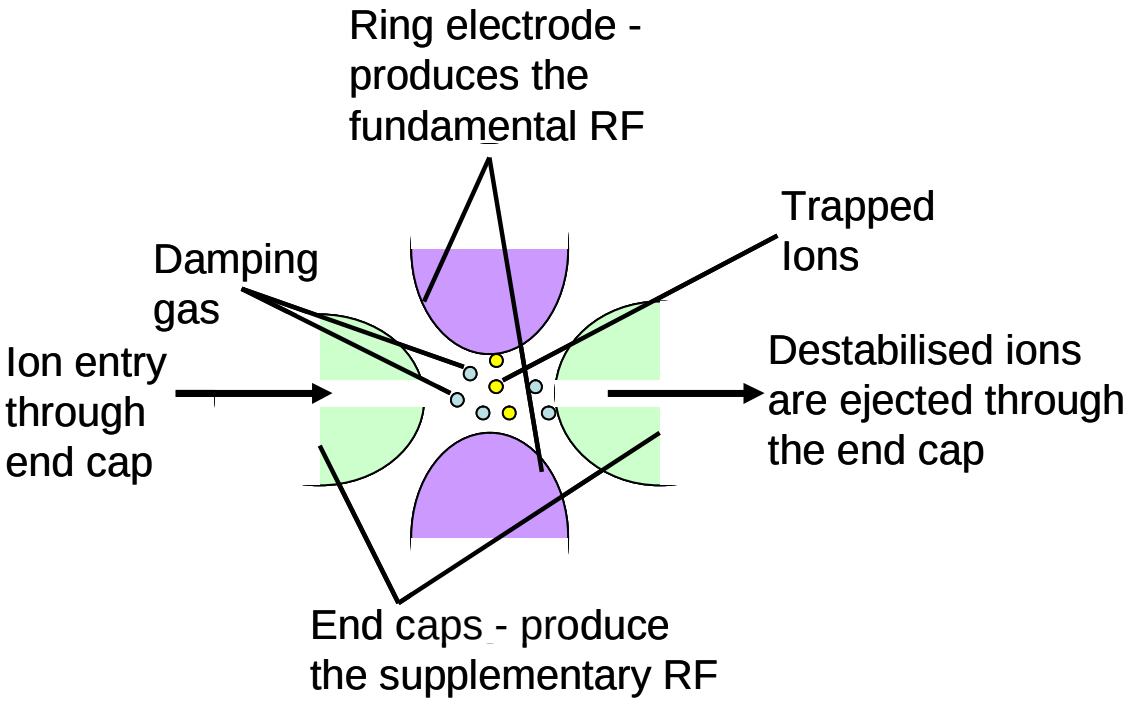


Figure 7: A spherical ion trap consists of two curved end cap electrodes separated by a ring electrode. Ions enter through one of the end caps, and are trapped in a quadrupolar field generated by an RF voltage on the ring electrode. Once trapped, ions are ejected by resonance excitation at a specific frequency related to the m/z of an ion.

1.6.2.3. Linear Ion Trap Mass Spectrometers

The linear ion trap (LIT) (see Figure 8) is simply a modification of the quadrupole [97]. However, its function is similar to that of the spherical ion trap. Rather than alternating negative and positive poles, the four poles of a linear ion trap maintain a constant, equal DC voltage, with a fundamental RF voltage added to enhance stability of the ions. A DC current applies an electrostatic field to each end of the quadrupole to prevent leakage of ions. Ions are trapped by pushing them into the quadrupole through an aperture in one end of the LIT at a DC voltage slightly higher than the trapping field on the end plates, and escape of the ions is prevented by the presence of damping gas. Once trapped, ions are excited by applying a large, supplementary RF voltage from the end plates, until ions of a particular secular frequency are destabilised and are ejected from the trap into the detector. Lower RF voltages applied to the ion trap can be used to fragment the ions [97].

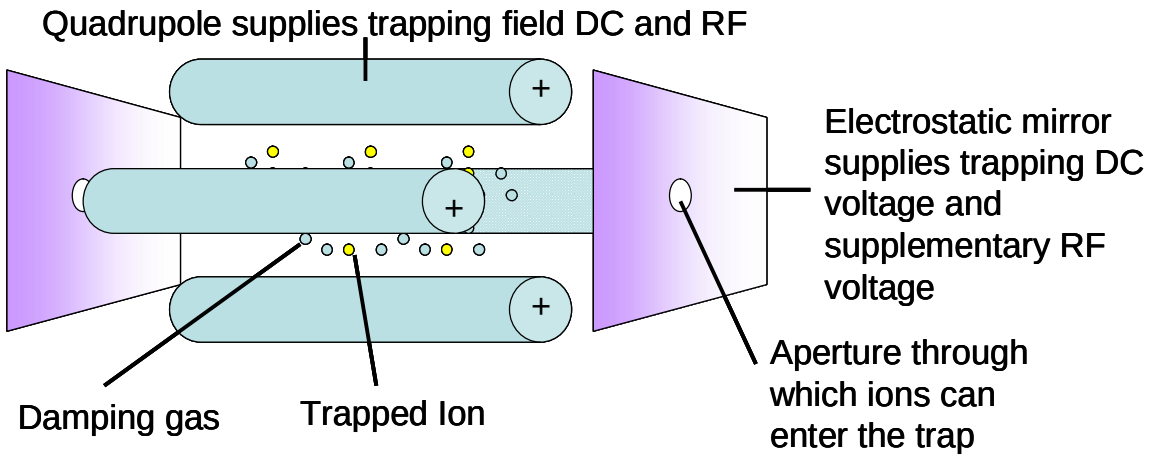


Figure 8: Linear ion trap. Ions are trapped by a combination of DC and fundamental RF voltages applied by the quadrupoles, and a DC voltage supplied by the electrostatic mirrors at the ends of the quadrupole. Ions can enter and leave the trap by the aperture at each end, but must overcome the mirror voltages at the end plates to do so. Once ions are pushed into the trap, they are prevented from leaving by reduction in kinetic energy mediated by collision with damping gas. Ions are ejected by supplementary RF excitation corresponding to their secular (resonance) frequency applied by the end plates.

1.6.2.4. Time of Flight Mass Spectrometers

The distinguishing feature of time-of-flight mass spectrometers is the flight tube – a metal tube through which ions pass, until they hit a detector at the end [93]. The cloud of ions (commonly generated by MALDI or electrospray) is initially accelerated by a strong electric field generated by a grid or ring with a high voltage (see Figure 9). Ions will separate over the length of the tube on the basis of their inertia. Since ions are accelerated with the same force and hence achieve the same kinetic energy during the initial acceleration, they will move at different velocities dependant on their mass to charge ratio. The longer the flight tube, therefore, the better the separation in time of arrival at the detector. However, ions are imparted with a range of kinetic energies during acceleration, and consequently, by the time they hit the detector, a broadening effect is observable on the ion packet, limiting the total resolution obtainable. In an effort to improve the resolution of time-of-flight tubes, the reflectron was developed by Boris Mamyrin, in 1966 [98]. This requires the use of a series of equal charged (with respect to the ions) electrostatic ring mirrors at the end section of the ToF tube. The mirrors apply an electrostatic field across one end of the flight tube, and ions travel towards the end of the tube, slowing down due to the field effect, before finally being reflected back out of the mirror into a detector offset from the source (see Figure 9). When a packet of ions hits the reflector, ions with higher kinetic energy will penetrate the field more deeply before being reflected. Conversely, ions with lower kinetic energy are reflected sooner, which acts to

focus the ions into a smaller packet. There is an upper mass limit to the use of a reflectron, because the range of kinetic energies of large ions contributes less to peak broading than metastable decay resulting from post-source dissociation (as described in 1.6.3.2.5). Since the products of metastable decay result in a variety of ions possessing different masses and kinetic energies, all of which will have different reflection profiles, the reflectron actually reduces the resolution of ions with an m/z of 10,000 and above [99].

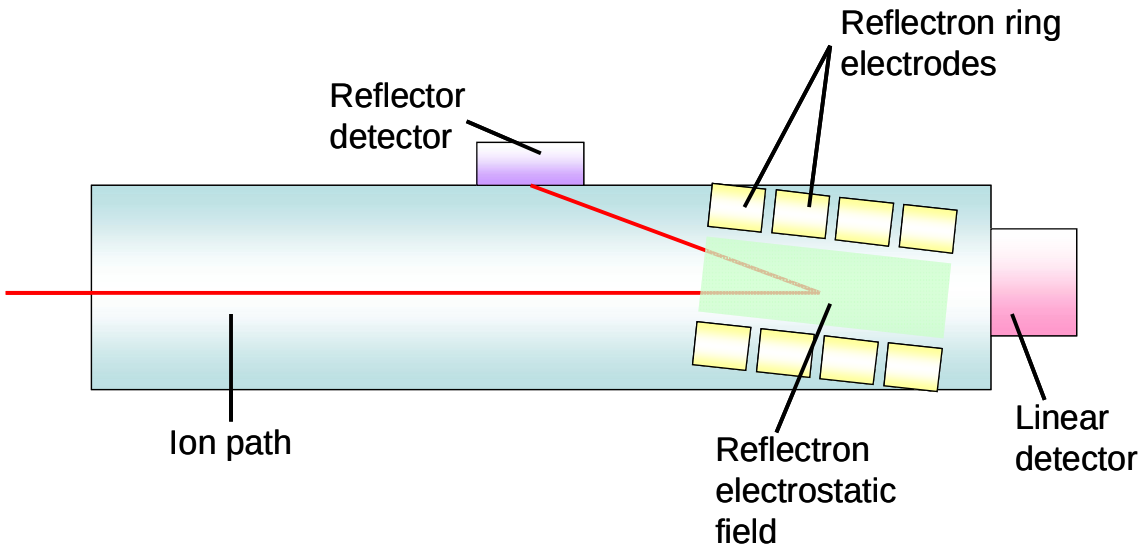


Figure 9: Time of flight tube. Ions pass down the tube until they hit the linear detector (for non-reflectron experiments) or are refocused and reflected by the reflectron, followed by collision with the reflector detector.

1.6.2.5. Fourier Transform Ion Cyclotron Resonance (FT-ICR) Mass Spectrometers

The FT-ICR mass spectrometer is a relatively recent development combining detection of mass and separation of ions [94]. The FT-ICR MS is based on the observation that charged particles in a magnetic field will experience a force proportional to the strength of the magnetic field, the charge on the particle, and their initial velocity (see Equation 1). If the ions are not subject to collision, their path will be bent into gyration perpendicular to the field, with a direction of gyration dependant on the polarity of their charge. The frequency of the gyration (the *cyclotron frequency*) is inversely proportional to the mass to charge ratio of a given ion [94].

Equation 1: Force experienced by a charged particle in a magnetic field. Q is the charge on the ion, v is the ion's velocity, and B is the magnitude of a spatially uniform magnetic field [100].

$$Force = qv \times B$$

An FT-ICR MS consists of the source and ion optics (used to generate and transfer the ions) followed by the ICR cell, which is maintained in a very strong magnetic field (ranging from 3Tesla for older instruments to 21Tesla for extremely powerful instruments). Introduction of ions to the cell is performed by impelling them with a low DC voltage, and initial gyration is imparted by a very rapid DC pulse. The FT-ICR cell consists of 6 plates in a box configuration held at a very high vacuum (around 10^{-11} mBar) (see Figure 10). One pair of opposing plates produces electrostatic trapping fields which preclude ions from leaving the cell at either end. A second pair of opposing plates produces an RF voltage, resulting in an electrical field that increases the radius of an ion packet's gyration if the RF frequency equals the cyclotron frequency of the ions. Broadband excitation of the ions in the trap to allow detection is performed by applying a rapidly changing RF frequency in an RF 'chirp'. The final pair of plates is a set of detector plates, which are placed perpendicular to the magnetic field trapping the ions. As ions cycle in the trap, they move towards and away from the plates, inducing an image current in each. The amplitude of the image current is proportional to the number of ions in a packet and their radius of gyration and the frequency of the image current is equal to the ion packet's cyclotron frequency. The image current for all the ion packets in the cell is a complex wave, the free induction decay (FID) wave (or transient), resulting from the addition of all of the waveforms of each constituent ion packet. The fast Fourier transform (FFT) is used to deconvolute each FID to determine the constituent waveforms and hence, from the frequencies of these, the masses present. The resolution of the resulting spectrum is proportional to the time allowed for the ions to cycle, usually several seconds, the magnetic field strength of the instrument, and the initial coherence of the ions generated from the applied pulse of frequencies [100], but is inversely proportional to the number of molecular collisions in the cell.

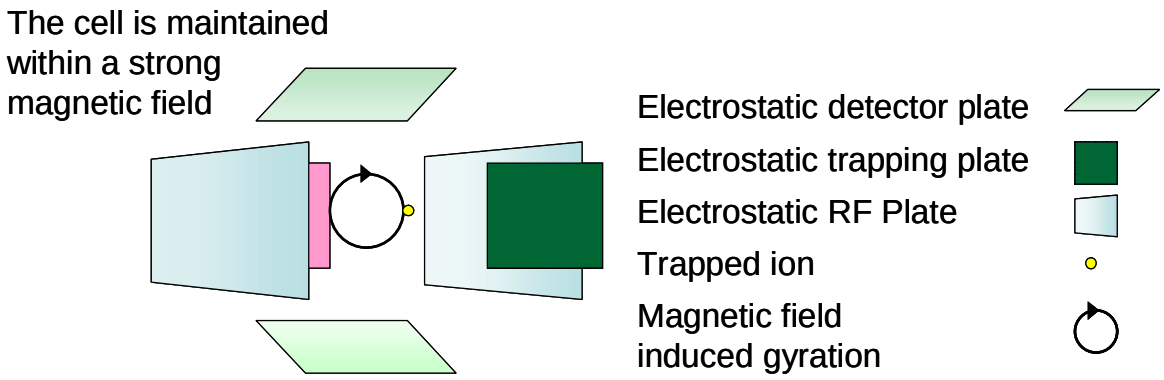


Figure 10: Schematic of an FT-ICR cell. Once impelled into the cell by an electrostatic field, ions are held there by electrostatic forces generated by two trapping plates. Gyration of the ions is induced by the strong magnetic field, and the radius of gyration is increased by resonant excitation induced by the RF signals generated by the RF plates. A mirror current is generated in the detector plates as the ions cycle and gradually decay to their resting state. Frequency of gyration is inversely proportional to the m/z ratio of the ions, and the complex wave generated may be deconvoluted with the use of the fast Fourier transform to determine the constituent waveforms and hence the m/z ratios of the trapped ions.

1.6.2.6. Orbitrap Mass Spectrometers

Also based on FFT deconvolution of waveform data is the orbitrap [95], a new (first described in 2005) mass spectrometer based on the Kingdon trap [101]. The orbitrap is an electrostatic trap, but, unlike the quadrupolar ion trap is based on constant DC voltage, rather than RF. The ion trap itself is toroidal, with a split outer shell electrode, and an inner barrel electrode (see Figure 11 and Figure 12).

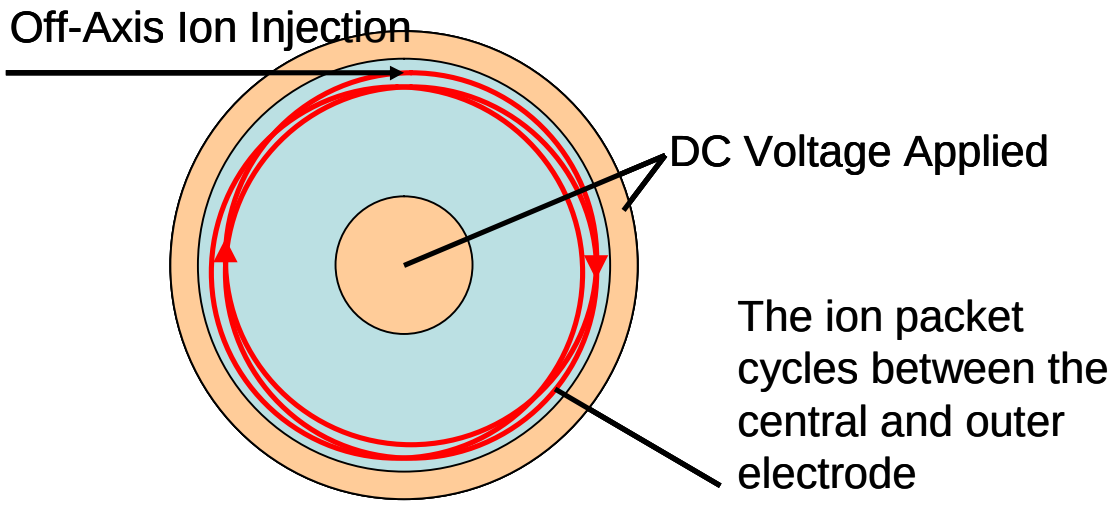


Figure 11: Cross section showing ion injection to the Orbitrap. The orbitrap consists of a central barrel electrode and an outer shell electrode around which ions orbit. Repulsion from the walls of the torus results from a constant DC voltage, applied to both the outer shell and central barrel. Injection of the ions is performed off axis of both the barrel electrode and the central shell (as shown in Figure 12).

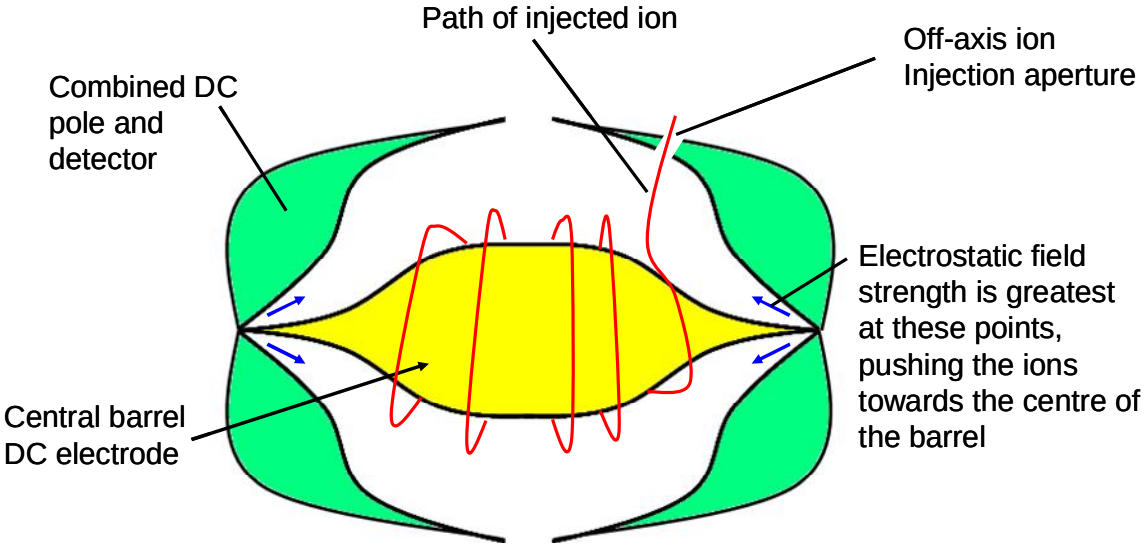


Figure 12: Cross section of the orbitrap, showing the split shell electrode which doubles as both a source of DC voltage and a detector. Ions are injected off axis (see also Figure 11), through a small aperture in the shell, and orbit between the central electrode and the outer shell. Electrostatic field strength is greatest at the ends of the trap, pushing the ions towards the central area of the barrel. Ions

are detected by their production of an image current in the shell electrodes, in much the same way as the two detector plates of an FT.

Ions are injected off axis of both planes of the trap, resulting in the equivalent of the FT-ICR's excited ion orbit without additional excitation. Once in the trap the ions cycle radially between a central barrel electrode and an outer shell electrode with an axial (in plane with the barrel electrode) oscillation frequency proportional to their m/z. The axial oscillation of the ion packets is detected by the amplification of transient mirror currents in the split outer electrode, much like the detector plates on the FT-ICR MS (see section 1.6.2.5). The current that is detected is deconvoluted in the same manner as, and approaches the same resolution of, the FT-ICR MS, but is limited by the axial decay of the ions path back to the centre of the torus after roughly 1.8 seconds [95].

1.6.2.7. Hybrid Instruments

Modern MS instruments often combine two or more mass scanners in a single package, hence the popular electrospray Q-ToF/QStar (Quadrupole-Time-of-Flight), Q-Trap/Thermo Finnigan TSQ triple-quadrupole (three quadrupoles (the last normally configured as a linear ion trap) in series), 4700/4800/Ultraflex (Time of Flight/Time of Flight), Bruker Apex (Quadrupole/FT) and the Thermo-Finnigan LTQ-Orbitrap (linear ion trap/orbitrap).

1.6.3. Top Down and Bottom Up Analysis

The spectra obtained as the end result of any method of mass spectrometry are functionally similar, with certain types of instruments possessing advantages over others from the standpoint of: *sensitivity* (the concentration threshold at which a molecule can be detected), *resolution* (mass divided by peak width at either 10% or 50% height) and *duty cycle* (the time taken to obtain a spectrum) [102, 103]. For all current mass spectrometers, resolution is inversely proportional to mass to charge ratio (m/z) and only the most highly resolving instruments (FT-ICRs and orbitraps) are capable of isotopic resolution of large intact proteins.

Top-Down analysis is the technique of using a mass spectrometer to analyse intact proteins [104]. Identification of proteins by accurate mass alone is complicated by post-translational modifications, isoforms and splice variants – the actual protein's mass may be significantly different than the mass derived from a genome database. Fragmentation methods (see Tandem MS) are normally employed to provide sequence data and information about post-translational modifications, but this remains one of the most challenging fields in mass

spectrometric analysis of proteins. Analysis of pre-digested proteins (bottom-up proteomics) is significantly less challenging than top-down analysis, and is the process by which the vast majority of proteomics data is collected.

1.6.3.1. Bottom up Analysis by Peptide Mass Fingerprinting

The most commonly used peptide mass fingerprinting program is MASCOT [16]. MASCOT is a commercial development of the molecular weight search (MOWSE) algorithm [48]. MOWSE has been through significant development stages since its inception in 1993, but the principle of its function has been maintained. As described (in section 1.4), the MOWSE algorithm compares a list of masses obtained from a mass spectrum with a list of masses generated computationally from a database of proteins. The set of peptides generated by proteolytic digestion of each protein is unique and may be used to statistically identify an unknown set of peptides as belonging to a specific protein. In order to identify an unknown set of peptides coming from a specific protein, two sets of data are required: the list of experimentally derived peptide masses itself, and a database containing mass lists for the theoretical peptides generated from a protein digest of each protein in a database. Pappin's original paper on the MOWSE algorithm [48] states that statistically significant identifications are routinely possible within a mass error tolerance of around 2Da. Given the size of modern genome databases, this value is extremely high, and 100 ppm mass accuracy is commonly used for PMF. For highly accurate mass readings, this can be reduced [105]. For each set of query peptides, the MOWSE algorithm compares fragment molecular weights with database molecular weights:

$DBMw - tolerance - 1 < FMw < DBMw + tolerance + 1$ [48]

Where:

- DBMw is the molecular weight from the database.
- FMw is the query molecular weight.
- Tolerance is the mass tolerance in Daltons or parts per million

Before generating the scores, the database is first sorted in two ways. First each protein is sorted into 10 kDa bins. Next, all peptides predicted to be generated from all proteins in a 10 kDa bin are sorted into 100 Da bins. The contents of each 100 Da bin are then normalised by dividing each peptide mass by the largest peptide mass in each bin.

Normalised value = Peptide mass / Maximum peptide mass per bin [48]

This generates a table of numbers between 0 and 1. Peptides matched from peak lists derived from real fragmentations are matched to their normalised values in the table. These numbers are then multiplied together if they occur in the same protein and their inverse is calculated.

The final score is generated by normalising to an average protein size of 50,000 Da, which reduces the effect of random matches to very large proteins.

The score is therefore calculated as:

Score = $50{,}000/(P_n \times H)$ [48]

Where:

- P_N is the inverted product score
- H is the mass of the protein

A schematic describing the process is shown in Figure 13.

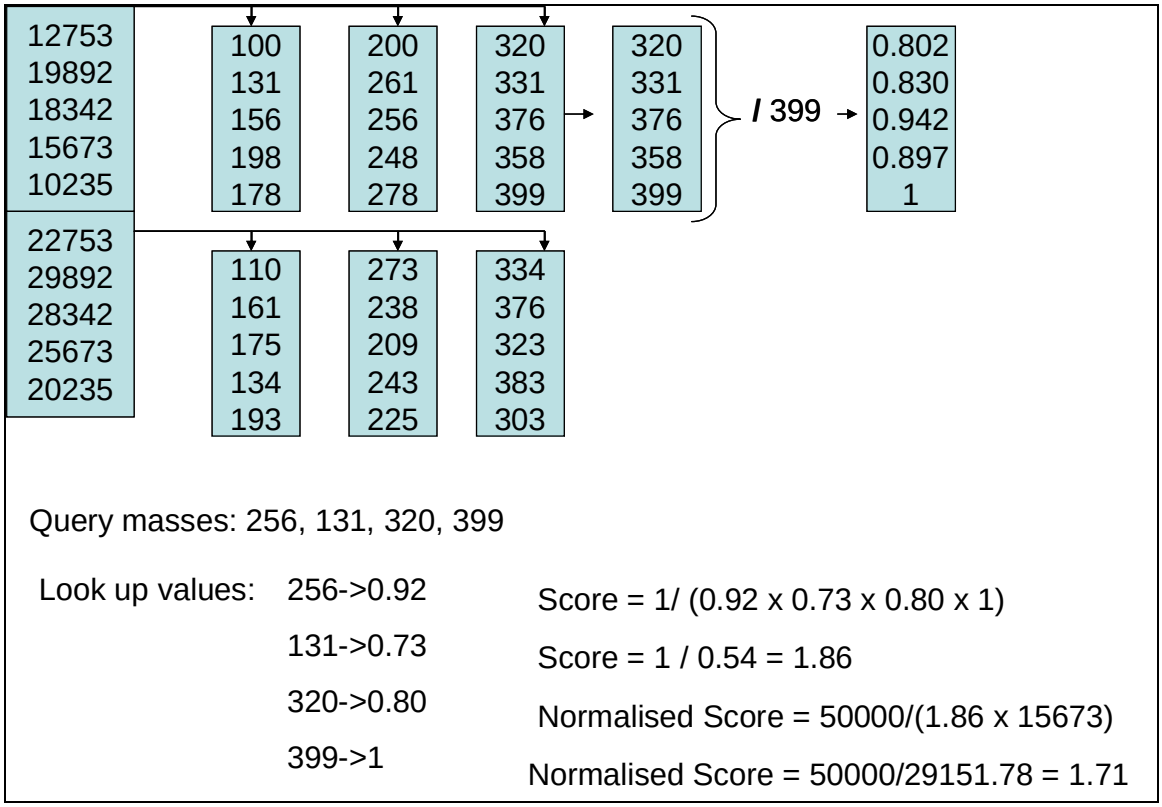


Figure 13: Schematic of the MOWSE scoring algorithm. Each protein in a database is sorted into a 10 kDa bin. Each peptide generated by an *in-silico* digest of a binful of proteins is also sorted, this time into 100 Da bins for each 10 kDa protein bin. To generate a score, each peptide has its mass normalized by dividing the original mass by the largest mass in a particular 100 Da bin. When a query

sequence is matched to a mass, its normalized score is obtained. Any other matches to a particular protein are multiplied together, then the resulting score is inverted to generate a probability-based score. To eliminate random matches to very large proteins, match scores are normalized against an average protein of 50 kDa by multiplying the score and the match's mass together, and dividing 50,000 by this total.

1.6.3.2. Tandem MS

Tandem MS, also known as MS/MS, demonstrates enhanced statistical strength of identifications compared to PMF, and can also, where necessary, allow sequencing of peptides for which the genomic sequence is not known in a process called 'de-novo sequencing' [106]. The principle of MS/MS is the fragmentation of parent peptides. This can be performed by a variety of methods, including collision induced dissociation (CID) [107], electron capture dissociation (ECD) [108, 109], electron transfer dissociation (ETD) [110], infra-red multi photon dissociation (IRMPD) [111], post-source dissociation (PSD) [112] and in-source dissociation (ISD) [113]. The fragmentation of a peptide is partially random, partially directed by the physicochemical properties of the peptide and partially directed by the method of fragmentation. These contributing factors lead to the generation of a variety of fragment ions from a single parent peptide. Once detected, these often form a series of peptide fragment ions separated by mass differences corresponding to single amino acids. The amino acid series have a terminology based on the position of the fragmented bond and the charge remaining on the detected fragment [107, 114] (see Figure 14). Fragments with charges localised to the N-terminal end of the peptide result in a, b and c ion series, while fragments with charges localised to the C-terminal end of the peptide are known as x, y and z ion series. The position of the broken bond denotes the specific type of ion, and the most common type of fragmentation is across the peptide bond, which results in a series of y ions or b ions, depending on their charge localisation. Fragmentations across the Cα-N result in c and z ions, and fragmentations across the Cα-C bond result in a and x ions. Fragmentation spectra are complicated by the fact that many of the ions formed are unstable and readily undergo further rearrangements.

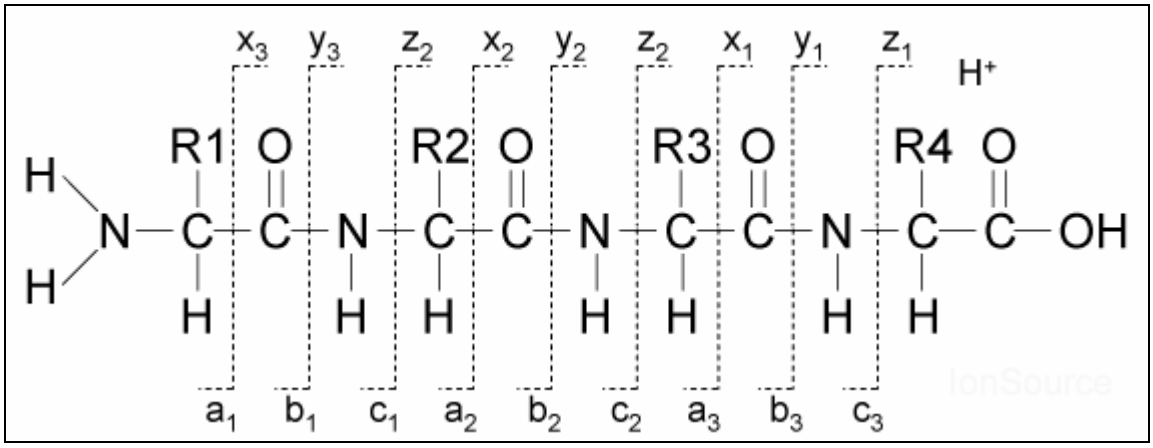


Figure 14: MS/MS peptide fragmentation nomenclature. (from http://www.ionsource.com/tutorial/DeNovo/full_anno.htm).

In addition, peaks corresponding to a double fragmentation across the y/b and x/a bonds, resulting in single amino acids are often observed in the low mass range of the spectrum, and are known as immonium ions [115].

A consideration of collision induced dissociation (CID), infra-red multi photon dissociation (IRMPD), electron capture dissociation (ECD), electron transfer dissociation (ETD), post-source dissociation (PSD) and in-source dissociation (ISD) will now be given.

1.6.3.2.1. Collision Induced Dissociation

CID is the most common method for performing MS/MS, and is based on thermal decomposition of ions. CID is commonly performed by passing ions in the gas phase through a low pressure chamber containing a small amount of gas. Ions are heated or excited, and collisions occur between the ions and the (generally inert) gas. The most common fragmentation in CID occurs across the peptide bond, generating predominantly y and b ion series [107]. Higher energy CID, such as that employed in the collision cells of ToF/ToF instruments, produces internal fragment ions generated from multiple cleavage events, in addition to y and b ions [116].

1.6.3.2.2. Infra-red Multi Photon Dissociation

Infra-red multi photon dissociation (IRMPD) is a type of fragmentation used to provide fragmentation similar to CID on FT-ICR mass spectrometers [111]. A high-powered infrared pulse laser is fired into the ICR cell. Absorbance of photons by the ions in the trap increases the energy of the ions until sufficient energy is absorbed causing bonds to be broken. It is generally a low energy fragmentation process, forming b and y ions [111].

1.6.3.2.3. Electron Capture Dissociation (ECD)

'electrons deify one razorblade
into a mountainrange; lenses extend
unwish through curving wherewhen till unwish
returns on its unself.'
e.e. cummings – pity this busy monster, manunkind

Electron capture dissociation is another type of fragmentation normally used in FT-ICR instruments [108, 109]. Instead of using photons to heat the sample, a metallic filament in the cell is heated in order to produce free electrons. These cycle in the opposite direction to the positively charged proteins or peptides and are captured by the positive ions, causing low-energy, chemically-based, radical fragmentation. ECD's efficiency is proportional to the number of charges on the peptide or protein, and it is therefore especially effective for electrosprayed proteins. Any of the six most common types of fragmentation (see Figure 14) can occur in ECD, but most commonly c and z ions are formed [109]. The fragmentation occurs by capture of an electron by a positively charged functional group, usually on a lysine, arginine, histidine or peptide N-terminus. Rearrangement results in the formation of a radical centre on a closely situated carbonyl carbon atom, followed by radical fragmentation across the N-Cα bond, and the formation of a radical on the Cα [117] (see Figure 15). The overall positive charge of the peptide is reduced by one as a result of electron capture, which can lead to neutralisation of the peptide if it carries only a single positive charge prior to electron capture.

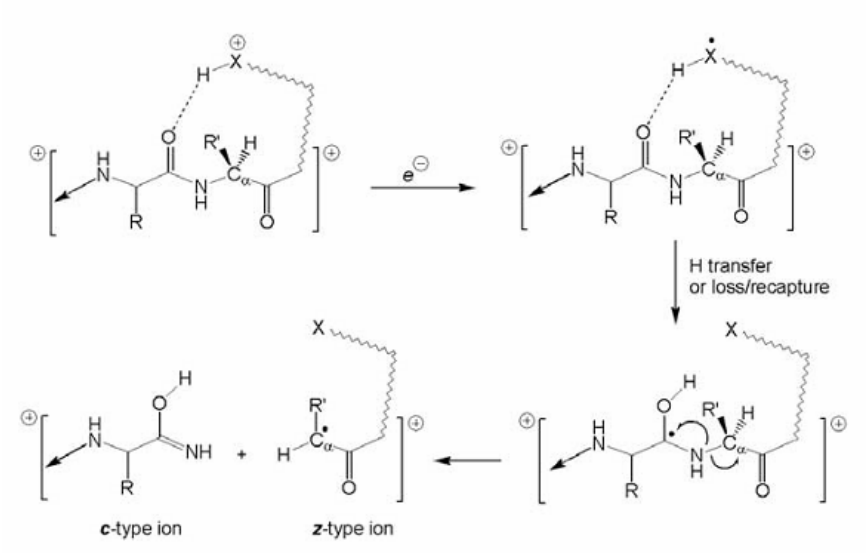


Figure 15: ECD process. An electron is captured by a positively charged functional group in an internally desolvated peptide, forming a radical. The electron migrates to the carbonyl carbon, followed by a rearrangement resulting in breakage of the N-Cα bond and formation of a radical at the Cα carbon. The ion containing the radical forms the z-ion and the other fragment forms the c-ion. From [117].

1.6.3.2.4. Electron Transfer Dissociation (ETD)

Electron transfer dissociation (ETD) is a development from ECD [110]. Implementation of ECD in non-FT ICR instruments is challenging because the generated electrons rapidly escape non-magnetic traps, but it is the best method for improving the fragmentation of highly charged peptides, since it shares with ECD the same property of efficiency in proportion to charge number. Hence an alternative, ETD was developed. To perform ETD, rather than generating a radical electron directly, methane is bombarded with electrons to generate a positively charged radical. This high-energy species is too energetic to provide selective fragmentation, so the methane is reacted with an additional reagent (fluoranthine or anthracine) that forms lower energy radicals. These radicals are then introduced to an ion trap and react with trapped peptides by transferring an electron, resulting in fragmentation in the same way as ECD. This type of fragmentation predominantly generates c and z ion series [118].

Proton Transfer Reaction

Proton transfer reaction is a method for reducing the complexity of ETD spectra. It was developed by Coon [119] as a method of performing the top-down analysis of large peptides and small proteins on a low-resolution ion trap instrument. Prior to this, the major difficulty in performing top-down analysis on such an instrument is that, upon fragmentation of the parent protein, ions with charge states too great for the resolution of the instrument to identify are often generated. These ions are incapable of contributing to the identification of the protein, as their mass cannot be correctly specified. Since many proteins, when electrosprayed, have charge states of 20+ or higher, and many ion trap MSs have difficulty distinguishing charge states of 4+ or higher, this leads to the generation of fragment spectra in which the charge states, and hence mass, of the majority of the peaks cannot be identified.

It was discovered that after ETD fragmentation, a second chemical (a strong proton acceptor) could be added to the reaction mixture to reduce the charge state of the fragment ions. This reagent sequesters protons with a reaction rate proportional to the square of the charge on a peptide, bringing the majority of fragment ions down to a charge state low enough to be identified by the ion trap MS. Coon et al. [119] used deprotonated benzoic acid as a proton acceptor, but recent developments by Hartmer [120] have resulted in a modification to the generation of the ETD reagent for use as a proton acceptor, thus eliminating the requirement for a second chemical ionisation source.

1.6.3.2.5. Post Source Dissociation (PSD)

Post-source decay (PSD) is a form of fragmentation generally used to generate primary structural data on MALDI ToF instruments [112]. Ions fragmented by post-source decay are known as 'metastable ions' because they are stable between the point of precursor ionisation and acceleration, but fragment during the ion packet's travel down the flight tube. A reflectron is required to identify post source decayed ions, because they initially travel at the same rate as their parent ions. Upon reflection and reacceleration, however, their different masses translate to different kinetic energies, resulting in separation prior to detection. In so-called ToF/ToF instruments, a timed ion selector allows isolation of an individual precursor in the first half of the flight tube, before fragmentation and reacceleration in the second half of the flight tube, thus ensuring a spectrum free of peaks except for the precursor peptide and its fragments. The fragment series generated by this type of fragmentation are usually a, y and b ions.

1.6.3.2.6. In Source Dissociation (ISD)

In source decay is similar to post-source decay, in that it is a method for fragmentation normally used with MALDI-ToF instruments [113]. Unlike PSD however, ISD occurs in the few nanoseconds (<200) that occur between laser ionisation and acceleration of the ion packet. The extent of ISD can be controlled to a degree by employing a MALDI-ToF MS equipped with delayed extraction, allowing the time between ionisation and acceleration to be increased. ISD generally produces c, y and z series fragment ions and can be used for top-down analysis of large intact peptides and small proteins. The difficulty of peforming precursor isolation, however, makes this technique only applicable to extremely pure samples, since there is no way to select a particular parent ion m/z for isolation.

Tandem MS produces a complex spectrum of peaks consisting of ion series, immonium ions, internal fragments and potential neutral losses resulting from PTMs. A peptide MS/MS spectrum is complex to analyse, but in general, one or two ion series predominate. For example, the CID on Q-ToF-type instruments generally produces a very strong y-ion series but weak b-ion series. Conversely, the very low energy resonance excitation CID of ion trap instruments generally results in nearly equal intensity y and b ion series. The latter is more challenging to assign sequence to, but provides additional information.

Based on empirical data on the type of tandem MS spectrum generated in a particular instrument, it is possible to assign peptide sequence to the ion series and other fragment ions in a particular spectrum. This is a more reliable means of assigning a peptide peak to a particular protein than PMF, and, with a good sequence, it is sometimes possible to identify confidently a protein on the basis of a single MS/MS spectrum.

1.7. Section 3: Sample Preparation and Modification of Proteins for LC/MS Analysis

1.7.1. Preparation of Biological Samples

Separation of proteins and peptides and their subsequent analysis by MS requires careful sample preparation. Cells are extremely complex and contain a large amount of material besides protein including lipid membranes, nucleic acids, sugars, metabolites and salts. Much of this will interfere with the analysis of the protein component, and it is therefore desirable to remove any contaminants prior to protein separation. In addition, many cell types have tough cell walls that are effective at protecting the inner environment of the cell from lysis. For this reason, specific methods must be employed to extract proteins from a cell. An important consideration when considering preparation methodology is that any manipulation of a cell may result in loss of protein and distortion of the quantity of individual proteins.

1.7.1.1. Selection of a Lysis Buffer

The internal environment of a cell is rigourously controlled, with ionic strength, pH and osmotic potentials subject to maintained feedback systems [121]. It is therefore ideal, once lysed, that the proteins be kept in a buffer as close to those original conditions as possible. Low concentration phosphate or Tris buffers are normally used at pH 7.5 or 8 [122], because the high salt content of isotonic buffers renders the sample unsuitable for ion exchange chromatography. In addition, to optimise protein extraction, detergents and chaotropes are generally added. Because of their amphipathic nature, detergents surround proteins and assist in maintaining them in solution by shielding their hydrophobic residues from the polar solvent [123]. The anionic detergent sodium-dodecyl sulphate (SDS) is the most commonly used in cell lysis, but other surfactants such as the non-ionic detergents n-octyl glucoside (NOG) and Triton X-100 or the zwitterionic detergent CHAPS can also be used [123]. Chaotropes, which function by disrupting the water network around proteins, can also assist in extracting protein [124]. The most common chaotropes used in lysis buffers are urea and guanidine hydrochloride (for example [125-127]). These steps result in unfolding of the protein.

1.7.1.2. Lysis Procedures

Lysis is the procedure of disrupting the cell and all of the organelles to release protein and there are several steps to the lysis procedure for any given cell type. The first is the

selection of an appropriate buffer to both contain the cells prior to lysis and to maintain the released proteins in solution post-lysis. The second is the actual breaking of the cells into the buffer solution.

1.7.1.3. Breaking Cells

The particular method used for lysis depends on the robustness of the cell. For most mammalian cells, which only have a lipid bilayer rather than a cell wall, simple sonication (the use of high frequency soundwaves to create shearing forces which tear open the cell), freeze thawing and vortexing (which also generate shearing forces) will often be sufficient. The efficiency of a sonicator for lysis is dependant on its power. Bath sonicators (the sound induced energy is transferred to the sample by use of a water bath) [128] are effective for easily lysed samples such as most mammalian cells. Probe sonicators, which directly contact the buffer which the cells are suspended in are more powerful and capable of lysis of moderately tough bacteria [129].

For cells with thick cell walls, such as some bacteria or yeast, the use of acid-etched glass beads and a ribolyser (a machine for automatic bead mixing) or vortex is required. Alternatively, a French press (which uses high pressure to squeeze cells through a narrow orifice, resulting in shearing) can be used, or the cells can be frozen in liquid nitrogen and ground in a pestle and mortar.

The process of bursting a cell releases the contents of most organelles in addition to the cytosol. Several macromolecular cell components, such as the proteasome or lysosome, contain proteases that degrade proteins, which is generally undesirable. For this reason it is common to add protease inhibitors to the lysis buffer. In addition, DNAses and RNAses are generally added to break up the stable nucleic acids and render them easily excluded from the solution by precipitation of the protein, since the resulting nucleosides are retained in organic solvents. After lysis, samples are centrifuged, and a large pellet of insoluble matter is usually observed. This contains membranes, sugars and some proteins insoluble in the lysis buffer. This can be either discarded, or retained for further, more stringent resolubilisation to extract proteins with different solubilisation conditions.

1.7.1.4. Purification by Precipitation

Once extracted from the cell, nucleic acids, salts and sugars still remain in the lysis buffer. To obtain a purified pellet of protein a precipitation can be performed upon the cell extract. There are several methods of precipitation, including trichloroacetic acid, polyethylene

glycol, chloroform, ammonium sulphate and acetone precipitation (reviewed in [130]). Most methods involve maintaining the precipitant/sample solution at a low temperature (-20 or -80ºC) in order to decrease the saturation point of the buffer liquid [131-133]. Precipitation using a nonpolar solvent such as acetone, chloroform or polyethylene glycol disrupts the hydrogen bonding around the polar residues in a protein, causing it to drop out of solution [131]. TCA or other acid-based precipitations again rely on the disruption of hydrogen bonds, but in addition, salt bridges between acidic and basic residues will be broken [133]. The change in charge state will also create new salt bridges which additionally change the conformation of the protein, exposing hydrophobic residues to the solvent. 'Salting out' or increasing the ionic strength of a solution until precipitation occurs, is customarily performed using ammonium sulphate. Based on the Hofmeister series, ammonium and sulphate ions are most effective at disrupting the hydrogen bonding of water molecules, which leads to a reduction in the solubility of hydrophobic species (such as most proteins) [132].

It is usual to remove remaining soluble materials by washing precipitated material with fresh precipitant, prior to resuspension of the sample in a buffer suitable for analysis.

1.7.1.5. Resuspension of Precipitated Samples

The choice of resuspension buffer for a protein sample has much in common with the choice of lysis buffer, with some additional strictures. Once precipitated, proteins are generally in a denatured, aggregated state and require detergents and chaotropes to assist in bringing them back into solution. Chaotropes such as urea or guanidinium hydrochloride are commonly used to assist in resuspension of biomolecules [130]. Guanidinium HCl is an extremely effective chaotrope, but its strongly ionic character precludes its use for any process sensitive to ionic strength, such as ion exchange, where the addition of the usual 4-8M of the compound would preclude any binding of the dissolved sample to the column. Urea is a more appropriate chaotrope for protein samples. However, proteins may be modified by the thermal breakdown products of urea (by carbamylation), such that only the purest sources of urea should be used and the solution must be maintained at a temperature of less than 37 degrees [134]. Many of the best detergents for resuspending protein are not compatible with liquid chromatography, in particular, anionic or cationic detergents bind very strongly to ion exchange resins and in some cases cannot be removed [135]. Another consideration for ion exchange is the buffer pH (see 1.5.4.2). Peptides are normally separated using cation exchange columns, and require a pH of < 3 is usually used in the loading and elution steps to maximise column binding, whereas proteins are normally

separated using anion exchange and therefore require a pH of > 8 for optimum binding, and loading buffers significantly different to these values may result in poor binding of the sample to the column.

The critical micelle concentration of a detergent is also an important consideration. At high concentrations in solution, detergents have a tendency to form micelles, spheres of detergent molecules in a single layer with the hydrophobic core protected from water and the hydrophilic groups in contact with the solvent. Large, predominantly hydrophobic molecules, such as proteins often localise to the hydrophobic cores of micelles and will readily pass through a column without interacting with the stationary phase of an LC column. It is therefore important to either use a low enough concentration of detergent that the CMC is not an issue, or to dilute a sample with concentrated detergent to a sufficient degree that the solution falls well below the CMC [136]. The choice of detergent is therefore a critical one: a detergent should ideally be non-ionic or zwitterionic, such that it fails to interact with ion exchange columns. It must also have a high enough CMC such that it can be used in high enough concentration to solubilise proteins but fail to carry those proteins through the column in micelles.

1.7.2. Proteolysis for Bottom Up Analysis

For all proteomics techniques except top-down analysis (see section 1.6.3), proteolysis is performed at some stage, whether that stage is prior to separation, for example in shotgun 2DLC, or as a final stage prior to MS analysis (in LC separation of intact proteins (as described in this thesis) or 2DGE (see section 1.3.1)).

Proteolysis is most commonly performed using trypsin, which cuts at the carboxyl side of lysine and arginine residues, excepting those immediately followed by a proline residue, due to steric hindrance and the requirement to break two bonds; the peptide and the proline R-group bond. In its native state, trypsin readily cleaves itself, and additionally exhibits chymotryptic activity (which generally cleaves at large hydrophobic residues) after long incubation times. It is thus advantageous to modify purified trypsin such that chymotryptic activity and self lysis is minimised.

A number of specific endoproteases exist, and may be used for specific purposes, such as for proteins possessing few trypsin digestion sites [16, 48]. For example, Asp-N cuts at the amine side of aspartic acid and cysteine and Glu-C cuts at the carboxyl side of glutamine and aspartic acid. Lys-C and Arg-C cleave proteins at the carboxyl side of lysine and arginine respectively and can be used to generate larger peptides than trypsin. Most PMF

software provides the option to use alternative proteases. Chymotrypsin is a protease popular for *de-novo* sequencing (where a sequence is generated manually without the use of fingerprinting by reading the fragmentation series). It is rather unspecific, preferentially cleaving at aromatic residues, followed by long chain charged residues such as methionine and lysine, except before proline due to steric hindrance and the requirement to cleave two bonds.

Additionally, several small molecule protein cleaving agents exist. The most commonly used is cyanogen bromide, which cleaves after methionine residues [137]. Small molecule agents have several benefits over proteases, the most obvious of which is that they do not lose their activity due to self-protease activity. They can also access folded proteins and proteins trapped in a gel more readily than the relatively slow diffusion process that trypsin must undergo and work in extreme conditions, for example highly acidic or basic buffers. The drawback to the use of cyanogen bromide is that it is extremely toxic.

A recently developed alternative to cyanogen bromide, 2-(2'-nitrophenylsulphenyl)-3-methyl-3-bromoindolenine (BNPS-Skatole) cleaves at tryptophan residues in the presence of acid [138].

1.7.3. Protein Abundance and Quantitation

Proteomic studies are occasionally dismissed as a 'fishing expedition'. This is rather a disservice, but nonetheless there is no better method for hypothesis generation than a global proteomics experiment. The concept of the 'global proteomics experiment is, however, limited by current separation and analysis technology [139]. This is due in part to the dynamic range of protein expression: some proteins are expressed in the region of tens of copies in a cell (some transcription factors, for example), some (particularly structural proteins and chaperones [10]) are expressed in the millions. It is, obviously, extremely difficult to obtain absolute quantities for each protein expressed by a cell type. This has been performed to date only for the yeast *Saccharomyces cerevisiae*, and then only in the single state of logarithmic growth. The latter research, however, was able to validate a computational method of estimating protein abundance: the codon adaptation index (CAI)[10] as providing a fairly good indication of the abundance of a protein in yeast. When evaluating a new method of separation and analysis for proteomics, it is useful to calculate the codon adaptation index for identified proteins, as an estimate of the dynamic range of the technique.

1.7.3.1. The Codon Adaptation Index (CAI) as a Measure of Protein Abundance.

There are 64 possible combinations of three DNA bases, and only 20 amino acids commonly expressed in proteins. After excluding three codons signalling for transcriptional termination (stop codons), 61 possible combinations remain. It therefore follows that some amino acids are coded for redundantly – by multiple codons. Leucine, for example, is coded for by six different codons (see Figure 16). It was observed [11] that the nucleotide sequences of high abundance proteins were biased towards similar alternative codons. For example, one or two codons that code for leucine may be more commonly used in high abundance proteins than others.

First Position	Second Position: U				Second Position: C				Second Position: A				Second Position: G				Third Position
	code	Amio Acid	code	Amio Acid	code	Amio Acid	code	Amio Acid									
U	UUU	phe	UCU	ser	UAU	tyr	UGU	cys	U								
	UUC		UCC		UAC		UGC		C								
	UUA	leu	UCA		UAA	STOP	UGA	STOP	A								
	UUG		UCG		UAG	STOP	UGG	trp	G								
C	CUU	leu	CCU	pro	CAU	his	CGU	arg	U								
	CUC		CCC		CAC		CGC		C								
	CUA		CCA		CAA	gln	CGA		A								
	CUG		CCG		CAG		CGG		G								
A	AUU	ile	ACU	thr	AAU	asn	AGU	ser	U								
	AUC		ACC		AAC		AGC		C								
	AUA		ACA		AAA	lys	AGA	arg	A								
	AUG	met	ACG		AAG		AGG		G								
G	GUU	val	GCU	ala	GAU	asp	GGU	gly	U								
	GUC		GCC		GAC		GGC		C								
	GUA		GCA		GAA	glu	GGA		A								
	GUG		GCG		GAG		GGG		G								

Figure 16: Codon usage table, demonstrating the multiple redundancy of codons to amino acids. From (http://citnews.unl.edu/croptechnology/lessonImages/960324911.gif)

$$CAI = \exp\left(\frac{\sum_i \sum_j n_{ij} x_{ij}}{\sum_i \sum_j n_{ij}}\right) \quad [11]$$

Where x is the total number of times the jth codon in the ith codon family has been used in the reference set of proteins, and n is the number of times the codon appears in the query gene. Further development of the equation was performed by Adam Eyre-Walker in 1996

[140], who altered the calculation to exclude the effects of amino acid composition on the calculation.

$$CAI = \exp\left(\sum_i \frac{\sum_j n_{ij} x_{ij}}{\sum_j n_{ij}} \Big/ m \right) \text{[140]}$$

Where m is the number of codon families used in the calculation.

By observing the codons used in the open reading frames (ORFs) of a selection of high abundance proteins, it is possible to produce a metric that can be used to evaluate any other ORF in a genome. By totalling the number of common and uncommon codons, the abundance of the protein for which it codes can be estimated. The CAI calculation produces a score of between 0 and 1 and is pseudo-logarithmic – thus a protein with a CAI of 0.5 will be roughly 10 times as abundant as a protein with a CAI of 0.4. The CAI is a general measure of abundance, and is calculated from an unchanging genomic dataset. It cannot, therefore, be used for quantitation in comparative studies, because phenotype is not taken into account.

1.7.3.2. Quantitation Techniques for Liquid Chromatography-Based Proteomics

The development of quantitative methods for proteomics has been an intense area of research. The two most common types of quantitative method are isotopic labelling, and direct quantitation by MS trace comparison.

1.7.3.3. Isotope Coded Affinity Tagging (ICAT)

ICAT is a saturation labelling technique that involves labelling proteins with two different mass tags, a ‘heavy’ and a ‘light’ tag at cysteine residues [141, 142]. The heavy tag is isotopically differentiated from the light tag by 8Da (original reagent) or 9Da (latest generation cICAT reagent).

The latest generation cICAT reagent consists of four parts: a reactive group that binds to the protein, the aliphatic carbon chain tag itself (either 9 x C_{12} or 9 x C_{13}), an acid cleavable linker and a biotin tag which allows purification. Samples are initially reduced to expose all cysteine residues to chemical modification, followed by labelling with either the light or the heavy tag and digestion. The two samples are then combined and purified away

from excess reagent using cation exchange chromatography. A further stage of affinity chromatography using a streptavidin column is required to trap only those peptides labelled with biotin, which is then cleaved off with addition of concentrated TFA. The final mixture of cysteine containing peptides is separated using shotgun 2-D liquid chromatography, and analysed using MS. An algorithm searches the resulting data for peaks in the peptide MS spectra differing by 9Da. Obtaining the area of each peak allows relative quantitation between control and test samples to be performed.

There are some limitations to the ICAT method. Approximately 8% of proteins contain no cysteine residues, and therefore will never be seen in a conventional ICAT experiment [143]. Additionally, because there are only two reagents, comparison of more than two samples can only be done by successive runs against a standard, which can introduce significant experimental error. Finally, due to the rarity of cysteine residues, only a few peptides are labelled in most proteins, which results in reduced confidence of quantitative analysis.

1.7.3.4. O_{18} Labelling

The tryptic digestion of peptides is a hydrolysis reaction, requiring the incorporation of water molecules. The use of O_{18} labelled water in the digestion buffer will result in the incorporation of a single O_{18} labelled carboxyl group at the carboxyl end of each tryptic peptide [36] [144]. Thus a control sample may be digested with normal water and a test sample may be digested in the presence of O_{18} labelled water, resulting in peptides differing in mass by 2 Da. These may be quantified in the same manner as iCAT peptides. The main drawback to this type of quantitation is that the labelled peptides will occur at the same point as +2 and subsequent isotopic peaks of the unlabelled peptide. This significantly limits the dynamic range of the technique, although attempts have been made to improve quantitation through the use of isotope matching software [145].

1.7.3.5. Isotope Tagging for Relative and Absolute Quantitation (iTRAQ)

The protocol for ITRAQ [146, 147] labeling is similar to that of ICAT. Samples to be quantified are reduced and alkylated (to prevent opportunistic chemical modification, such as oxidation of cysteine residues, and to prevent labelling of cysteine residues). They are then digested and labelled with the iTRAQ reagents. The nature of the label itself, however is significantly different: while the ICAT reagent targets cysteine residues, the iTRAQ tag reacts with the amino terminus of peptides and lysine side chains. Additionally, rather than

the two different tags available in iCAT, four different iTRAQ reagents are available with masses of 114, 115, 116 and 117, and a newly available kit raises this number to eight, with a slight increase in the mass of the tag from 144 to 304. When attached to the parent peptide, iTRAQ tags are isobaric (possessing identical masses) and only a single peak is observed in each mass spectrum. Quantitation is observable only in MS/MS spectra, where the tags readily cleave, and the fragment ions appear in an unpopulated region in the MS/MS trace. By comparing the areas of the 114, 115, 116 and 117 Da peaks, it is possible to obtain relative quantitation of peptides from multiple different samples. To obtain absolute quantitation, a standard with known abundance must be labelled with one of the tags and added to the mixture. This then serves as a calibrant for the remaining peaks. Drawbacks to the methodology are that, with the standard protocol, labelling is performed at a relatively late stage in the sample preparation process – after reduction, alkylation and digestion, and therefore significant experimental error may be introduced by the manipulations at each stage. Additionally, generation of intense reporter ions for quantitation along with high quality sequence data for identification of a peptide is difficult to achieve.

iTRAQ reagents are highly reactive in water and are therefore reacted with the sample peptides in a solution of 70% ethanol. This step limits the use of iTRAQ for intact protein labelling because of potential organic solvent precipitation of proteins.

1.7.3.6. ExacTag

ExacTag is a new labelling technique from PerkinElmer, based on an MS/MS reporter ion, much like iTRAQ, but is intended for use on intact proteins. The reagent itself relies on thiol chemistry, and therefore, like ICAT, is intended for use on cysteine residues. An alternative version of the protocol exists that relies on blockage of cysteine, derivitisation of lysine residues to a thiol group, and labelling of the lysine residues for greater coverage of quantitation. The reporter ions used in ExacTag have two benefits: first, up to ten different tags are available, thus allowing a significant number of replicates to be performed, or the analysis of ten different states simultaneously; the second benefit is that the reporter ions range from 451 to 478 m/z, and are therefore likely to be in within the mass range for fragments on an ion trap instrument. This advantage, of course, consequently leads to a significant disadvantage, as the total mass of the ExacTag is 1046Da, compared to, for example, the iTRAQ modification, which is only 244 Da. This leads to the generation of very large labelled peptides, especially if post-labelling tryptic

digestion at lysine residues is blocked due to inability of the enzyme to recognise the labelled residue, resulting in missed cleavages.

1.7.3.7. Stable Isotope Labelling with Amino Acids in Cell Culture (SILAC)

The technique of SILAC is based on the addition of isotopically labelled amino acids to culture medium [33]. Cells are grown in medium containing C13 and/or N15 labelled amino acids (commonly lysine or arginine). Carbon starvation is sometimes employed to assist uptake of labelled amino acids. These labelled amino acids are taken up and incorporated into cellular proteins. Once cells are lysed, the lysates mixed together, and analysed by MS, it is possible to observe the relative abundances of a given labelled and unlabelled peptide or protein pair in terms of the spectral abundance of their individual mass peaks. The principle drawback of SILAC is that the technique works only with cells that can be grown on medium containing the labelled amino acid and that are auxotrophic for the relevant amino acid. The most commonly used amino acids for SILAC labelling are lysine and arginine, based on the principle that after a tryptic digestion every peptide (excepting the C terminal peptide) can be used for quantitation. Other amino acids suitable for SILAC labelling include leucine (which was used in the original report of the technique [33]), and labelled methionine [148] and tyrosine are commercially available. Lysine is commonly used in $^{12}C_6$ and $^{13}C_6$ versions for a mass shift of 6, but can be supplemented by $^{14}N_2$ and $^{15}N_2$ labels for a total mass shift of 8. Triplet labels can be implemented with the use of labelled arginine, as the 6 and 10Da mass shifts between $^{12}C_6{}^{14}N_4$ arginine, $^{13}C_6{}^{14}N_4$ arginine and $^{13}C_6{}^{15}N_4$ arginine can be clearly visualised on most MS instruments.

1.7.3.8. Direct Quantitation by Comparison of Ion Abundance (Label Free Quantitation). DeCyder MS, MSight and Progenesis LC/MS

Both DeCyder MS, MSight [149] and Progenesis LC/MS are developments of comparison software for 2-dimensional gel electrophoresis. DeCyder MS is based on the DeCyder software and relies on the generation of two dimensional maps of peptide intensity, where mass is plotted against time, with ion abundance displayed as spot intensity. Once generated, these ‘pseudogel’ images can be overlaid and compared. Peaks are detected, and can be de-isotoped, deconvoluted and their charge state can be determined. After this stage, runs are overlaid and warped to provide peak matches. Comparison statistics are performed on the matched peaks, with rigour increasing with the number of replicated runs. MSight is public domain software based on Melanie, which, like DeCyder is designed for comparison

of 2D gels. MSight is capable of the same kind of pseudogel generation and matching, but lacks the peak processing facilities of DeCyder MS. Progenesis LC/MS is a recently developed piece of software produced by NonLinear Dynamics. It is a development of their SameSpot algorithm for 2-dimensional gel electrophoresis analysis. It, like MSight, lacks peak processing facilities, but provides excellent statistics, including principle component analysis, of the quantitative differences observed.

Unlike the chemical labelling techniques described above, where quantitation is performed in a single separation, direct MS quantitation relies on the comparison of mass spectrometry data, normally the different intensities of peaks between runs. It is therefore entirely reliant on the reasonable reproducibility of the mass chromatograms, and variation in processing between sample can have significant deleterious effects on the statistics.

1.8. Complex sample sources for intact protein separation

1.8.1. *Bordetella parapertussis* and *bronchiseptica*

Bordetella sp. are Gram-negative pathogenic bacteria responsible for respiratory infection in humans. There are several recognised species, but only four: *B. pertussis, B. parapertussis, B. bronchiseptica and B. holmsesii* are pathogenic in mammals [150]. *Bordetella pertussis* is the causative agent of whooping cough in humans, and the other members of the *Bordetella* family are sufficiently closely related that *B. parapertussis* (which infects humans and sheep) and bronchiseptica (which infects dogs and pigs) can be used as model organisms to study the molecular biology of pertussis virulence in humans [151].

Virulence in *Bordetella* is mediated by two-component regulatory systems (a membrane bound, self phosphorylating kinase that transfers its phosphoryl group to a response regulator, which is generally a transcription factor). Comparative genome analysis has resulted in the identification of a putative novel two-component regulatory protein, termed Bho AS. Production of Bho AS null mutants and murine virulence assays confirmed that the null mutant *Bordetella brochiseptica* is avirulent.

1.8.2. Leishmania

More than 12 million people worldwide are infected with *Leishmania*. It results from infection with *Leishmania* parasites. [152]. Cutaneous leishmaniasis is a pathology

classifiable into four stages: growth of a painless nodule followed by development of a central crust and occasionally ulceration, initiation of healing, and development of a mottled scar. The parasite persists in two forms: a promastigote in the fly vector (*Phlebotomus* in Old World leishmaniasis and *Lutzomyia* in New World leishmaniasis), and an amastigote in the vertebrate host. The two forms may be distinguished by the internal, non-functional flagellum in the amastigote stage and the extended, motile flagellum in the promastigote stage [153].

Treatment of leishmaniasis is performed using chemotherapy. The most commonly used drugs are arsenical compounds, but pentamidine is an alternative, second-line therapy based on interference with the protazoans' DNA synthesis by modification of the kinetoplast (mitochondrial genome) and fragmentation of the mitochondrial membrane [154]. The mechanism of resistance to pentamidine-based drugs is characterised by decreased uptake (18 to 75 fold), and increased efflux [155], but the mechanism by which this occurs is not currently known.

Drug resistance in *Leishmania* is characterised by downregulation of many membrane transporter proteins, including glucose transporters [152]. The preparation of *Leishmania* cell lines in which expression of glucose transporters is silenced may provide insights into the molecular biology underlying drug resistance in these parasites.

1.8.3. K562 Cells

The K562 leukemia cell line is derived from a patient in blast crisis suffering from chronic myeloid leukaemia (CML) [156]. CML, in almost all cases, results from the Philadelphia chromosome, which is the result of a reciprocal translocation between chromosomes 9 and 22 (originally observed in [157]). The translocation results in a gene fusion between two genes known as BCR and ABL. Normal BCR displays serine/threonine kinase activity [158] and ABL is a tyrosine kinase responsible for cytoskeletal assembly at the neuromuscular junction [159]. The bcr/abl fusion protein loses the serine/threonin kinase activity of bcr to result in a constitutively activated abl [160], which produces antiapoptotic, pro-proliferation cellular behaviour.

Glivec, also known as STI571 or imanitab mesylate, is an anti-cancer drug used for treatment of CML [161]. Its mechanism of action is to specificially inhibit tyrosine kinases, and it therefore inhibits the antiapoptotic behaviour of bcr/abl, resulting in cell death. K562 cells, which express the bcr/abl oncogene, are therefore an important model

for the study of leukaemia and cancers in general. Studying the effects of Glivec on the cells may provide significant insight into these diseases.

1.8.4. PC12 Cells

PC12 cells are derived from a pheochromocytoma of adrenal medulla cells in an irradiated rat. The characteristic that renders them of particular interest to researchers comes from their response to neural growth factor (NGF) and epithelial growth factor (EGF). In the presense of NGF, they begin to differentiate into neurons [162] but when stimulated by EGF, they proliferate [163]. Both phenomena are controlled by stimulation of the Ras/Raf/MEK/ERK/MAP kinase pathway, the former in sustained stimulation, the latter in transient stimulation [163]. They are therefore an important target for study, both in cancer (to understand control of proliferation) and to enhance our general understanding of intracellular signalling and differentiation.

1.9. Conclusions

Liquid chromatography and mass spectrometry, combined in the discipline of proteomics, form a whole much greater than the sum of their parts. The convenience of reversed phase liquid chromatography for both contaminant removal and high resolution separation recommend its use for any kind of protein or peptide sample. Unfortunately, a single dimension of separation is only sufficient for very low complexity samples, thus a multidimensional sample preparation technique is the method of choice, whether this is gel or LC based, followed by reversed phase LC of peptides to remove contaminants.

Separation of intact proteins, whether by gel or liquid chromatography techniques has many advantages over peptide separation. Due to the limitations of gel technology, alternative methods for intact protein separation are desirable. The use of liquid chromatography for intact protein separation has met with difficulty due to poor resolution. The use of polystyrene divinylbenzene monolithic columns may alleviate this drawback, providing fast and reproducible separation of protein species. Multidimensional separation using these columns will provide a new technique for high resolution separation of proteins.

Along with rapid peptide separation, protein LC would provide high resolution and high throughput analysis of very complex samples for bottom up studies. For top down analyses, a method of separating intact protein provides the possibility of the analysis of complex samples using high resolution instruments. In addition, the combination of a

multidimensional liquid chromatography protein separation method with quantitation allow elegant and information-rich experiments to be performed.

1.9.1. Aims:

- To develop two-dimensional liquid chromatography of intact proteins as a proteomic technique, using polystyrene-divinylbenzene based columns.
- To apply quantitation methods to the protein separation to allow comparison of different samples.
- To improve the sensitivity and analysis rate of digested protein samples.
- To use intact protein separation in conjunction with top-down mass spectrometry analysis.

2. Chapter 2: Methods and Materials

2.1. Chemicals

General chemicals were obtained from Sigma (UK) at Analar or better purity grade, unless otherwise stated. Acetonitrile (HPLC gradient grade) was obtained from Sigma (Dorset, UK). Protein standards (BSA, cytochrome c, alpha lactoglobulin, beta lactalbumin, carbonic anhydrase, trypsin inhibitor, ovalbumin, myoglobin, lysozyme and alpha casein) were obtained from Sigma (UK). Sequencing grade trypsin was obtained from Promega. The Bradford assay kit was purchased from Bio-Rad laboratories. ddH_2O was obtained from a MilliQ system. Urea and CHAPS (PlusOne grade) was obtained from GE Healthcare.

2.2. Sample Preparation

2.2.1. Bacterial Samples: *Bordetella parapertussis*

A culture of *Bordetella bronchiseptica* (strain BBC17) was grown overnight at 37ºC in Stainer-Scholte SSX media with shaking, diluted into 2 x 500mL vessels and again grown overnight at 37ºC with shaking. Cells were harvested by centrifugation at 12000rpm at 20ºC and washed 4 times with ddH_2O using progressively smaller volumes. A final volume of 10 mL was split into Eppendorf tubes and pelleted at 12000rpm. Cells were stored at -20ºC until freeze dried. Growth of *Bordetella* was performed by Martin Lynch.

To prepare lysates, cells were suspended in 330µl 8M urea, 4% CHAPS, 100mM Tris, followed by vortexing for 1 minute. The solution was sonicated three times for 10 seconds each interspersed with chilling on ice for 30 seconds. Insoluble components were pelleted by centrifugation at 14000xg for 10 minutes. Supernatant was transferred to a clean tube and 1200 µl cold acetone was added to enable precipitation. The sample was vortexed briefly and stored at –20ºC for 24 hours. To produce a pellet of purified protein, the sample was centrifuged at 14000xg for 10 min. The resulting pellet was resuspended in a small volume of 8 M urea, 2% CHAPS, 10 mM phosphate buffer. This provided a total yield of 1.5-2 mg of protein per 10 mg of lyophilised bacteria (determined by Bradford assay (BioRad)).

Further methods development resulted in slight modification to the precipitation protocol for eukaryotic samples (described in section 2.2.2.2).

2.2.2. Eukaryotic Samples: *Leishmania donovani* and *mexicana*

2.2.2.1. Growth

Leishmania donovani and *mexicana* promastigotes were grown in medium 199 (Invitrogen, UK) supplemented with 10% (w/v) heat-inactivated foetal calf serum (Invitrogen, UK) at 25ºC and harvested during the mid-log phase of growth. Cell pellets were collected by centrifugation 1200 x g and were washed twice with PBS. Preparation of *Leishmania* samples was performed by Christina Nuala.

2.2.2.2. Lysis

Cells were suspended in 1 mL 8 M urea, 4% w/v CHAPS, 10 mM Tris-Cl pH8, followed by vortexing for 30 seconds. The solution was sonicated for three periods of three seconds using a probe sonicator, interspersed with 30 seconds chilling on ice. Insoluble components were pelleted by centrifugation at 14000xg at 4ºC for 10 mins. 250 µls supernatant was transferred to each of 4 clean tubes and 1 mL cold acetone was added to each. The samples were vortexed briefly and stored at -80ºC for 24 hours. To wash the pellet, 1mL cold 80% acetone was added, followed by vortexing to disperse the pellet and centrifugation at 14000g at 4ºC for 10 mins. The protein pellet was resuspended in a small volume of 8 M urea, 0.4%CHAPS, 10mM Tris-HCl buffer pH8.

Protein concentrations were determined by Bradford assay (BioRad, UK), as per instructions.

2.2.3. Eukaryotic Samples: SILAC Labeled K562 Cells

K562 cells were grown in 50 mL of RPMI medium in 175 cm^2 cell culture flasks to a cell density of 1-3x10^7 cells/mL. $^{13}C_6$ lysine-labeled cells were stimulated with Glivec for 6 hours. Cells were harvested by centrifugation in 50 mL cell culture tubes at 1,000xg for 2 min at room temperature. Cleared growth medium was removed by using a Pasteur pipette attached to a vacuum pump. Cell pellets were washed twice with ice-cold PBS, followed

by centrifugation at 1,000xg for 2 mins at 4°C. PBS was removed, and 1 mL of ice cold NP-40 lysis buffer was added, with agitation. Lysis was assisted by two periods of incubation at 4ºC for 15 min, interspersed by agitation on ice. The lysate was centrifuged at 25,000xg for 15 min at 4°C. Supernatant was removed and transferred to a fresh tube, followed by BCA protein assay. Lysates were produced by Christian Preisinger.

2.2.4. Eukaryotic Samples: SILAC Labeled PC12 Cells.

PC12 cells were grown on Dulbecco's Modified Eagle Medium (DMEM, Gibco), and either labelled with Arg ^{12}C, ^{13}C (+6Da) and $^{13}C^{15}N$ (+10Da) or left unlabelled. For stimulation, starved overnight, and then stimulated with EGF (^{13}C), NGF ($^{13}C^{15}N$) for 30 minutes or were left unstimulated (unlabelled cells). Pellets were washed with PBS, and cells were lysed in 1% v/v triton X-100, 150 mM NaCl, 20 mM HEPES pH7.5 with 1 mM phenylmethylsulfonylfluorid (PMSF), 2 mM sodium fluoride (NaF), 1 mM sodium vanadate (Na_3VO_4), 5 µg/mL leupeptin and 2.2 µg/mL aprotinin, 1 mM sodium pyrophosphate ($Na_3P_2O_7$), 20 mM β-glycerophosphate as protease inhibitors. Protein concentrations were assayed with the BCA assay (Pierce) and mixed at concentration ratios of 1/1/1. SILAC labeled PC12 lysates were produced by Alex Von Kriegsheim.

2.2.5. Eukaryotic Samples: Histone Preparations

Wild type HCT116 cells were cultured to 60-80% confluence in RPMI 1640 medium with 10% FCS and 1% w/v penicillin/streptomycin. Cells were washed and scraped into nuclear extraction lysis buffer (250 mM sucrose, 50 mM Tris-HCl, pH7.5, 25 mM KCl, 5 mM MgCl, 0.2 mM phenylmethylsulfonyl fluoride, 50 mM $NaHSO_3$, 45 mM sodium butyrate, 10 mM 2-mercaptoethanol, 0.2 % Triton X-100) containing protease and phosphatase inhibitor cocktails (from Roche, Indianapolis, USA, and Sigma, Dorset, UK respectively). Cells were pelleted by centrifugation at 800xg for 10 mins followed by acid extraction with 1 mL of 0.2 M H_2SO_4 with incubation for 1 h at 4ºC. Samples were then centrifuged at 13000xg for 15 minutes, followed by precipitation by adding TCA to 20% of final volume. The pellet was washed in cold acetone containing 1% 50 mM HCl followed by dispersion of the pellet into cold acetone overnight. Samples were resuspended in H_2O, and a BCA protein assay was then performed. Histone samples were produced by Bernie Ramsahoy.

2.3. Ion Exchange Separation.

2.3.1. Bacterial Samples:

Ion exchange was performed on wild-type *B. parapertussis* and Bho AS mutant samples. Before separation, each lysate was diluted from 40 µl to 200 µl with 10 mM phosphate buffer. Sample injection was performed by a FAMOS™ autosampler (LC Packings, NL). Separation was performed on coupled 2mm x 50mm ProPac™ SCX (Dionex, UK) and 2mm x 250mm ProPac™ SAX (Dionex, UK) columns, maintained at a flow rate of 200 µl per minute using a BioLC™ GP50 pump (Dionex, UK). The low ionic strength buffer (A) consisted of 10 mM phosphate buffer, pH 8, and the high ionic strength buffer (B) consisted of 10 mM phosphate buffer, pH 8, 1 M KCl. The gradient ran from 0%-50% B in 20 minutes, with a 5 minute wash at 90% B and 30 minute equilibration at 0% B. Collection was performed into 200 µL glass-lined vials (Chromacol, UK) or 96-well microplates (Greiner Bio-ONE, UK) using a Probot (Dionex, UK) as a fraction collector. Chromatograms were generally of high absorbance and complexity. Filled vials and plates were stored at -20°C until reversed phase separation could be performed.

2.3.2. Eukaryotic Samples:

Ion exchange separation was performed using an UltiMate® 3000 HPLC system (Dionex, UK) at 37ºC using either a 250mm x 1mm ProPac™ PS-DVB pellicular SAX column run at 50µl/minute with a 200µl sample loop. For column reproducibility analysis, ion exchange was done using a 250mm x 2mm ProPac™ PS-DVB pellicular SAX column either alone or coupled to a 50mm x 2mm ProPac™ PS-DVB pellicular SCX column run at 200µl/minute with a 200µl sample loop to improve coverage of basic peptides. Both ion exchange equilibration and elution solutions consisted of 10mM Tris-HCl buffered to pH 8, the elution buffer additionally contained 1M NaCl.

When collecting ion exchange fractions for subsequent reinjection onto reversed phase columns, a 20 minute gradient from 0-400mM NaCl was employed for separation with an 8 minute wash at 1M NaCl followed by 15 minutes of equilibration at 0mM NaCl for a total run-time of 45 minutes. Fractions from ion exchange chromatography were collected in 96 well microplates using a CTC PAL (LEAP technologies) fraction collector.

2.4. Reversed Phase Separations

For all samples not used for top-down analysis, ion exchange fractions were additionally separated via 500 μm x 5 cm reversed phase PepSwift PS-DVB monolith (Dionex, UK) and collected in 384-well AB-1056 microplates (ABGENE, UK) at a flow rate of 20 μl/min. Initially, 60 μl of the 200 μl collected were used for separations, reducing the effective protein from 1 mg to ~300 μg. Subsequent injections were performed using a 200μl loop, fitted to the FAMOS™ autosampler (later replaced by the functionally equivalent WPS-3000 (Dionex, UK)) to load an entire ion exchange fraction. The 200μl loop was switched out during the run to minimize gradient delay.

Monolithic PepSwift PS-DVB capillary columns (50 mm x 0.5 mm) were obtained from Dionex (UK), and were used for all intact protein reversed phase separations. Pumping was performed using a BioLC™ GP50 pump (Dionex, UK) fitted with an Accurate™ splitter (Dionex, UK). Later, this was replaced with an UltiMate 3000™ system (Dionex, UK). Reversed phase buffers consisted of 2% v/v acetonitrile, 0.1% v/v HFBA, 97.9% v/v ddH_2O and 80% v/v acetonitrile, 0.08% v/v HFBA, 19.92% v/v ddH2O. Each fraction was separated using a 20 minute gradient at 20 μl/min from 2 to 52% v/v acetonitrile, with a 4 minute wash step at 72%, followed by an 8 minute equilibration step at 2% v/v acetonitrile. Column temperature was maintained at 60ºC using a column oven. Initial fraction collection was performed into 384-well PCR plates using a Probot (Dionex) fraction collector, and later a CTC autosampler (LEAP technologies)). Plates were dried in a centrifugal vacuum dryer (SpeedyVap, Eppendorf) for ~1 hr, followed by overnight digestion with Promega modified porcine trypsin solution. Twenty μg of trypsin was dissolved in 25mM ammonium bicarbonate dissolved in 10% v/v ACN, and 5 μl was added to each well followed by incubation at 37ºC for 16 h.

2.5. On-line 2 Dimensional Separation of Intact Proteins

To fully automate the process for the generation of intensity maps, the interim ion exchange fraction collection was eliminated by coupling a 1 mm x 250 mm ProPac™ ion exchange column (Dionex, UK) to two valve-selectable reversed phase ProSwift 500 μm x 50 mm columns. A very long gradient (from 4 - 32 hours in duration depending on the first dimension resolution desired) of 0 mM to 400 mM NaCl (details are described in section 2.5.3) was applied to sample trapped to the ion exchange column. Eluted proteins were trapped by a downstream reversed phase column. During this process, a standard reversed

phase gradient (as described in section 2.5.3) was applied to the alternate column (see Figure 17). At the end of each reversed phase gradient, reversed phase columns were switched using a nanovalve, eliminating any loading time and maintaining a smooth ion exchange gradient. Sample washed off the ProPac™ column by increasing salt concentration was trapped by a ProSwift column. While this was occurring, the other ProSwift column was eluted by a short reversed phase gradient, and fractions were collected. Reversed phase columns alternated between trapping and elution during the extended ion exchange gradient.

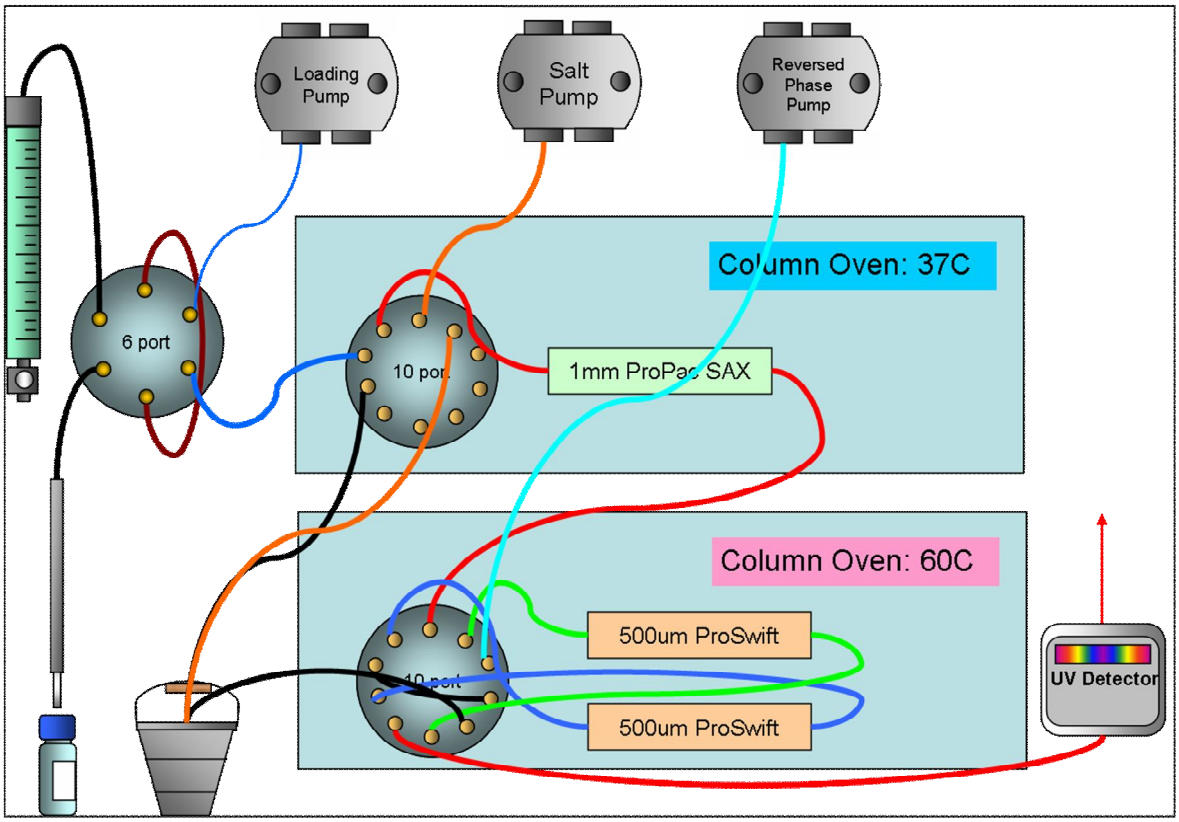


Figure 17: Schematic of online intact protein separation system plumbing. An initial loading step is performed using the isocratic pump. This flow path is equipped with an in-line microfilter to trap any particulates to preclude contamination of the ProPac™ column downstream. During the injection step, proteins are trapped either on the ProPac™ strong anion exchange column, or failing that, on a 500µm ProSwift reversed phase column downstream of it. Post-injection, a very shallow gradient is applied to the ion exchange column, trapping eluted proteins on alternating ProSwift columns while the other column has a standard, 20 minute reversed phase gradient applied to it.

2.5.1. Equipment

HPLC was performed using a Dionex UltiMate® 3000 running Chromeleon® 6.80 and fraction collection was performed using a CTC Pal (LEAP Technologies, USA) robot. Columns consisted of a custom packed 1 mm x 250 mm ProPac™ ion exchange column (Dionex, UK) and two reversed-phase ProSwift 500 µm x 50 mm columns.

2.5.2. Preparation of Mobile Phase and Standards.

Ion exchange phases consisted of 10mM Tris-HCl pH 8.0, with the elution buffer additionally containing 1M NaCl. Aqueous reversed phase buffer consisted of 2% v/v acetonitrile, 0.1% v/v formic acid and 97.9% v/v ddH_2O. Organic reversed phase buffer consisted of 80% v/v acetonitrile, 0.08% v/v formic acid, and 19.92% v/v ddH_2O.

2.5.3. Chromatographic Conditions

The HPLC was configured according to the diagram in Figure 17. The ion exchange column was a ProPac™ PS-DVB pellicular SAX column with dimensions of 1mm x 25cm. Reversed phase columns were PepSwift (poly)styrene di-vinylbenzine monoliths with dimensions of 500 µm x 5cm. Gradients were programmed as shown in Table 1. Flow Rates were 25 µl/min for the loading and ion exchange gradient flows, and 20 µl/min for RP separations. Injection volumes were 180 µl into a 200 µl loop in all cases. Detection of eluted proteins and peptides was performed using UV absorbance at 214 and 280 nm. Ion exchange columns were maintained at a temperature of 37ºC, while reversed phase columns were maintained at 60ºC.

Table 1: Gradient table per 'fraction'

Time (min)	MeCN (%)	NaCl (%)
0.0	2%	n
1.0	2%	*
21.0	56%	*
22.0	72%	*
26.0	72%	*
27.0	2%	*
36.0	2%	n+2.85% (for 16 fraction separations)

* The ion exchange gradient continued between n and n+2.85% at each intervening step.

2.6. MS Analysis of Samples

2.6.1. Direct MALDI of Digested Samples

Direct MALDI spotting was performed using a Genesis Workstation 100 (Tecan) liquid handling robot. Target plates were pre-spotted with a 2.5 mg/mL α-cyano hydroxyl-cinnamic acid (CHCA), 70% v/v Acetonitrile, 29.9% ddH_2O, 0.1% v/v TFA solution. Prior to the matrix drying out, wells containing digested peptides from two dimensional intact protein separations were resuspended in 3 µl 0.1% TFA, 50% acetonitrile, 49.9% ddH_2O and 0.5 µl was overspotted onto the target plate. Two 192 spot target plates (AB 192+6, Applied Biosystems, Warrington, UK), were used to hold the spots from each 384 well plate. Target plates were analysed using a 4700 Proteomics Analyser MALDI ToFToF instrument (Applied Biosystems, Warrington, UK).

2.6.2. Conventional Peptide LC-MALDI of Digested Samples

LC-MALDI was performed using an UltiMate® Plus Dual Gradient system (Dionex, UK), with spotting performed using a Probot (Dionex, UK). Solvents consisted of 2% v/v acetonitrile, 0.1% v/v FA, 97.9% v/v ddH_2O and 80% v/v acetonitrile, 0.08% v/v FA, 19.92% v/v ddH_2O. Gradients performed on the UltiMate® system ran from 2% v/v to 41% v/v acetonitrile with a 20 minute gradient, 4 minute wash step at 72% acetonitrile, and 10 minute equilibration at 2% v/v acetonitrile, for a total run time of 36 minutes. The column used was a 75 µm C18 3 µm particle size, 100Å pore size PepMap (Dionex, UK) run at 300 nl per minute. Column eluents were mixed with a 2.5 mg/mL CHCA, 70% v/v acetonitrile, 29.9% v/v ddH_2O, 0.1% v/v TFA solution concomitant with spotting. 192 spot target plates (AB 192+6, Applied Biosystems, Warrington, UK), were used to collect fractions for 24 mins at 30 second time periods. Target plates were analysed using a 4700 Proteomics Analyser MALDI ToF ToF instrument (Applied Biosystems, Warrington, UK).

2.6.3. Conventional Peptide LC-ESI-MS of digested samples

LC-ESI-MS was performed using initially the QStar Pulsar i (Applied Biosystems, Warrington) coupled to an UltiMate® 1 HPLC system (Dionex, Camberley) and later the HCT Ultra (Bruker Daltonics, Bremen) coupled to an UltiMate® 3000 HPLC system (Dionex, Camberley). A PepMap 75um C18, 3µm particle size, 100Å pore size column (Dionex, UK) was employed for separations in all cases. Electrospray was facilitated with the use of PicoTips (20µm ID, New Objective). The flow rate was maintained at 300nl/min. Samples were resuspended in plate by the addition of 20 µl of buffer (2% v/v acetonitrile, 97.9% v/v ddH_2O and 0.1% v/v formic acid) by the autosampler prior to injection of 10µl. Reversed phase buffers consisted of 99.5% v/v ddH_2O, 0.5% v/v FA, and 99.5% v/v acetonitrile, 0.5% v/v FA for QStar separations, and 2% v/v acetonitrile, 0.1% v/v FA, 97.9% v/v ddH_2O and 80% v/v acetonitrile, 0.08% v/v FA, 19.92% v/v ddH_2O.

Gradients performed on the UltiMate® 1 ran from 5% v/v to v/v 40% acetonitrile with a 20 minute gradient, 4 minute wash at 90% v/v acetontirile and 10 minute equilibration at 5% v/v step for a total run-time of 36 minutes. Gradients performed on the UltiMate® 3000 ran from 2% v/v to 41% v/v acetonitrile with a 20 minute gradient, 4 minute wash step at 72% v/v acetonitrile, and 10 minute equilibration at 2% v/v acetonitrile, for a total run time of 36 minutes. Column temperature was maintained at 37ºC using a column oven for all UltiMate® 3000 separations.

2.6.4. Fast Peptide LC/MS of Digested Samples

Electrospray ionisation mass spectrometry (LC-ESI-MS/MS) was performed using an HCT-Ultra (Bruker Daltonics, Bremen). An UltiMate® 3000 (Dionex, UK) equipped with a 200 µm x 5cm monolithic PepSwift column was used for separation. Samples were resuspended in plate by the addition of 10 µl of buffer (2% v/v acetonitrile, 97.9% v/v ddH_2O and 0.1% v/v formic acid) by the autosampler prior to injection of 5 µl. A 'direct injection' implementation was employed, where the sample loop was maintained in-line with the column to improve the coverage of hydrophilic peptides. Reversed phase buffers consisted of 2% v/v acetonitrile, 0.1% v/v FA, 97.9% v/v ddH_2O and 80% v/v acetonitrile, 0.08% FA v/v, 19.92% v/v ddH_2O. Gradients ran from 2% v/v to 41% v/v acetonitrile in 10 minutes, with a 4 minute wash step at 72% v/v acetonitrile, and 6 minute equilibration at 2% v/v acetonitrile, giving a total run time of 20 minutes at a flow rate of 3 µl/min. Column temperature was maintained at 60ºC using a column oven.

Electrospray was performed using a Proxeon 30 µm x 50 mm stainless steel nano-electrospray emitter. Data was analysed using the MASCOT software (Version 2.1, Matrix Science, London) to search the complete *Leishmania infantum* genome database.

2.7. Parallel Monolith LC/MS

Final implementation of the parallel system was performed on an UltiMate® 3000 (Dionex, UK) equipped with two 200µm monolithic PepSwift columns was used for separation. Analysis was initially performed using a QTrap 2000 linear ion trap MS (Applied Biosystems) and the QStar Pulsar I. Later analyses (those used for biological samples) were analysed using electrospray ionisation mass spectrometry (LC-ESI-MS/MS), performed using an HCT-Ultra (Bruker Daltonics, Bremen). Lyophilised, digested samples were resuspended in plate by the addition of 10 µl of buffer (2% acetonitrile, 97.9% ddH_2O and 0.1% formic acid) by the autosampler prior to injection of

5µl. Columns were run in parallel such that while an injection was being performed during an isocratic aqueous flow on one column, a previously loaded sample would be separated by gradient flow on the other column, thus eliminating the loading, equilibration and wash times for the column. Reversed phase buffers consisted of 2% acetonitrile, 0.1% FA, 97.9% ddH_2O and 80% acetonitrile, 0.08% FA, 19.92% ddH_2O. Gradients ran from 2% to 41% acetonitrile with a 10 minute gradient (later a 7 minute gradient was implemented), and 1 minute wash step at 72% acetonitrile, for a total run time of 15 minutes at a flow rate of 3µl/minute (the additional time was required for the high organic wash to pass beyond the detector). Column temperature was maintained at 60ºC using a column oven. Development of the instrumentation is described in 5.1.4.1.

Electrospray was performed using a Proxeon 30µm x 50mm stainless steel nano-electrospray emitter and the MicroSprayer (Bruker Daltonics, Coventry). Data was analysed using the MASCOT software (Version 2.2, Matrix Science, London) to search the complete *Leishmania infantum* genome database for *L. donovani* and *L. mexicana* samples or the NCBInr database with the taxonomy restricted to Homo sapiens for PC12 and K562 lysates.

3. Chapter 3: Two Dimensional Liquid Chromatography of Intact Proteins Using Ion Exchange and Reversed Phase Columns

3.1. Aims:

There are many drawbacks to both two dimensional gel electrophoresis and multidimensional chromatography of peptides. Polystyrene divinylbenzene stationary phases have potential to provide a high-resolution, multidimensional separation technology combining many of the advantages of both traditional methods with few of the drawbacks. The aims of this project were:

1. To optimize sample preparation methods that effectively solubilise proteins, but remain compatible with the separation technology.
2. To optimize the method for flow rate, sample size and fraction number, as well as collection and digestion conditions for intact protein separation.
3. To test the developed methodology on complex samples, assessing proteome coverage and separation quality.

3.2. Introduction

Two dimensional liquid chromatography was developed as a separation technique for intact proteins for a number of reasons: While 2D-electrophoresis (see section 1.3.1) is a powerful tool for protein separation, it is difficult to automate, and has limited utility for the analysis of many classes of protein. Multidimensional liquid chromatography of peptides, exemplified by shotgun peptide analysis (see section 1.3.2), is highly automated, but requires the proteolytic digestion of the sample, which greatly increases sample complexity and therefore the required resolution of the subsequent separations. Furthermore, much of the information on differential post-translational modification and isoform expression is lost. The use of polymeric pellicular and monolithic PS-DVB columns allow rapid, improved-resolution, automated separation of intact proteins from complex samples.

ProPac™ ion exchange and ProSwift monolithic reversed phase columns for protein separation were commercially released in 2005. An intuitive next step was their combination to produce a high resolution multidimensional protein separation technique, similar to MuDPIT, directly applicable to intact proteins. To implement this technique, sample preparation, buffer, and fraction collection parameters were required to be optimised. After optimization, rigorous assessment of the technique was performed.

As a model system for protein separation, *Bortetella bronchiseptica* (closely related to the bacterium that causes whooping cough) was chosen as the initial source of protein samples. The three primary reasons for this choice were that:

- It has been the subject of a differential gel electrophoresis (DiGE) experiment, whereby lysates are separated on a two-dimensional gel and proteins are identified via mass spectrometry. It therefore provides an immediate means of comparison.
- It is a simple prokaryote whose genome contains 4185 open reading frames (ORFs). This level of complexity represents a challenging system for full proteome resolution by 2-dimensional liquid chromatography (2DLC).
- Clonal populations of bacteria should provide reproducible lysates in comparison to fractions derived from more complex samples.

Leishmania donovani was chosen as a second model system for several reasons: First, it is more complex with roughly 8000 open reading frames, leading to the expression of a similar number of proteins, together with eukaryotic post-translational modifications. Secondly several mutants exist, including glucose transporter knockout and drug resistant lines that are valuable as targets for quantitative studies (discussed in section 6).

3.2.1. Sample Preparation

To test the functionality of the 2D intact protein LC technique, it was considered important to use samples prepared using a standard methodology, to ensure compatibility with existing methodologies. In order to provide a lysate possessing a broad protein complement but free of significant contaminants, proteins must be precipitated from a crude lysate in order to exclude membranes, sugars and nucleic acids and resuspended in a separation compatible buffer. This is not necessarily trivial, since proteins become aggregated after precipitation, and are not readily resuspended. Detergents, particularly SDS, and high concentrations of urea (6-8 M) are normally used to resolubilise precipitated

proteins. Although urea (a non-polar compound) should not pose a problem for the purposes of ion exchange chromatography, ionic detergents such as SDS are predicted to strongly associate with the stationary phase, resulting in contamination of protein-containing fractions. Moreover, at high concentrations of detergent, proteins become suspended in micelles, which pass through the column without interacting with the stationary phase.

3.2.1.1. Optimisation of Proteolytic Digestion

Protein identification in a proteomics context is usually performed by enzymatically digesting a protein sample, followed by analysis of the resulting peptides by mass spectrometry. The mass peaks obtained form a unique 'fingerprint', by which the parent protein can be identified.

The agent used to cleave proteins in most proteomics work is trypsin, whose optimal conditions require aqueous solvent at pH 8. Enzymatic digestion of proteins is improved when the proteins are unfolded, as this allows access to interior cleavage sites. However, the buffer most commonly used for trypsin digestion (25 mM ammonium bicarbonate) is non-denaturing.

3.2.2. Methods and Materials

3.2.2.1. Sample Preparation

General reagents were obtained as described in section 2.1.

3.2.2.1.1. Methods Development Sample Preparation

For general methods development, a protein mix consisting of nine proteins of varying physicochemical properties and masses was prepared. The protein mix consisted of BSA, cytochrome c, alpha lactoglobulin, beta lactalbumin, carbonic anhydrase, trypsin inhibitor, ovalbumin, myoglobin, lysozyme and alpha casein. A stock solution was prepared from these proteins with equal concentrations of 41 μM per protein. 2 μL of this solution was diluted to 200 μL with ion exchange buffer for ion exchange protein separation. A less concentrated 1 μM 9 protein mix solution was prepared from the concentrated stock described above for reversed phase protein separation methods development, and 10 μL of this solution was diluted to 60 μL or 200 μL (depending on loop size) for each reversed phase separation.

3.2.2.1.2. Complex Sample Preparation

B. parapertussis and *B. bronchiseptica* lyophilates were prepared by Martin Lynch and lysed by the author as described in section 2.2.1. *Leishmania donovani* cells were prepared by Christina Nuala, and lysed by the author as described in section 2.2.2.

3.2.2.2. Sample Separation

Bacterial samples were separated by ion exchange using a BioLC™ GP50 (Dionex, UK) gradient pump, providing a flow rate of 200 µL/min using coupled ProPac™ 2 mm x 50 mm SCX and 2 mm x 250 mm SAX columns as described in section 2.3.1. Reversed phase separations were performed on the ion exchange fractions using ProSwift PS-DVB monolithic columns (Dionex, UK) as described in section 2.4.

Eukaryotic samples were either separated by standard, off-line ion exchange methodology using an UltiMate 3000™ system and coupled ProPac™ 2mm x 50mm SCX and 2mm x 250mm SAX columns as described in 2.3.2, followed by reversed phase separation as described in 2.4 or were separated in both dimensions using the on-line protein separation methodology described in 2.5.

3.2.2.3. MS Analysis of Separated Samples

Fractions were collected into 384-well microtitre plates (AB-1056, ABGENE, UK) and subjected to proteolytic digestion for analysis. Each well was digested with 5µL of 20 µg/mL Promega modified bovine trypsin per fraction. Trypsin was dissolved in 25 mM ammonium bicarbonate containing 10% acetonitrile. Digestion proceeded for 16 hours at 25ºC. Digested *Bordetella spp*. samples were analysed by LC/MALDI as described in 2.6.2. For *L. donovani* samples, fast peptide separations using PepSwift PS-DVB monolithic columns (as described in 2.6.3) were applied to each digested fraction. Data was processed using the MASCOT software (Matrix Science). MASCOT version 1.9 was applied to the *Bordetella spp*. samples, and MASCOT version 2.1 was applied to the *L. donovani* samples.

3.2.3. Results and Discussion

3.2.3.1. Plumbing and Method Development

The basic strategy implemented involves the separation of the sample by ion exchange in the first dimension with fraction collection followed by second dimension separation. Fractions are then collected from the second dimension separation and analyzed by digestion followed by rapid peptide analysis. Alternatively, it is possible to couple first and second dimension separations directly on-line to avoid the necessity for fraction collection from the first dimension. The number of fractions collected will depend on the researcher's desire to limit analysis time or take advantage of the highest resolution possible. Based on the peak widths (roughly 15 seconds half-height for ion exchange and 10 seconds half height for reversed phase) generally obtained in each dimension, 64 ion exchange fractions, each separated into 48 reversed phase fractions would be necessary to obtain the optimum resolution from the technique, giving a total of 3078 fractions for further analysis. Even using rapid monolithic column based peptide separation, analyzing all of these fractions would take many days, making this somewhat impractical on a routine basis. However, to achieve effective separation, while minimizing analysis time, the number of reversed phase fractions collected are commonly restricted to twenty-four, collected at 1.25 minute intervals, for each of sixteen 1-minute ion exchange fractions. This was later decreased to 50 second intervals, again with 24 fractions collected, once the elution profile of proteins from complex mixtures was ascertained.

Data from the many off-line two-dimensional separation of proteins described in this chapter (e.g. 3.2.3.11.1 and 3.2.3.11.2) shows that is an effective method for complexity reduction of complex samples. However, when using an intermediate fraction collection step it is difficult to guarantee the reinjection of the entire collected fraction, and additionally sample loss due to the binding of protein to the plastic in the microplates may also be an issue. Finally, frequent reconfiguration of the system to alternatively run ion exchange and reversed phase columns resulted in reliability problems. Therefore, an on-line method, similar to that employed with automated peptide two dimensional separations was implemented (see Figure 17). In addition to ameliorating the issues described above, on-line 2D separation allows more precise control over the gradient and the process is significantly more automated.

An early example of an on-line 2D system is described by Apffel [164] with the use of large bore (7.5 mm in the first dimension and 4 mm in the second dimension) columns for multidimensional separation. The paper describes a variety of stationary phases, including derivitised silica-based and polyether based size-exclusion columns in the first dimension and either a C18, PS-DVB ion exchange (much like the ProPac™ columns used in this

study) or a silica column derivitised with aminopropyl groups. Although the sample types focus on the separation of small molecules, and are therefore not directly relevant to the protein work described herein, the system plumbing shows some unique properties, especially the use of a second loop to collect each fraction from the first dimension, allowing the elimination of fraction collection apparatus. The use of two reversed-phase columns, however, in the present work, allows a significant increase in speed of separation. However, Apffel's work has benefits in that only a single pump is used, rather than two, and that due to fraction collection into a loop, any species not trapped on the second column will be directed into the detector, rather than to waste.

A system closer in format to the one detailed here, described by Opiteck [165], uses a size exclusion column, like Apffel's work, but couples it to two C18 reversed phase columns for improved separation. The plumbing for the system is almost identical to the online system described in this work, with the benefit that only two pumps are used. Resolution is exceptional, but its application to tryptic digests limits the comparisons that can be made. A follow-up paper [45] applied the same principles and valving to separate intact proteins using the same SEC first dimension, followed by reversed phase PS-DVB monolithic separations. Resolution is poorer than the work described in this thesis, especially in the initial size-exclusion dimension, although the fact that proteins will be eluted in order of size, in much the same way as an SDS PAGE gel provides useful additional information.

Once plumbing schematics were completed, analysis of protein samples could be performed. This involved optimization of sample preparation with the requirement of compatibility with the separation technology.

3.2.3.2. Protein Resuspension

Performing lysis directly into an ion-exchange compatible buffer (10mM phosphate), free of detergent, produced chromatograms noticeably less complex than those from precipitated protein lysates, probably because only reasonably soluble proteins would be retained in a detergent-free buffer. Two options are available for increasing the amount and range of protein obtained by lysis into a detergent and chaotrope-containing extraction buffer: either direct analysis of the crude lysate by LC or precipitation, resuspension and separation of clean lysate by LC. The latter has the advantage that contaminants such as nucleotides, lipids and sugars are excluded by the process of precipitation, leaving a mostly pure protein pellet, and therefore most sample preparation optimization was performed with this in mind.

Resuspension of protein pellets may be difficult, requiring chaotropes and strong detergents. It is therefore imperative that these additives do not interfere with subsequent separation. Two chaotropes are commonly used in protein chemistry: urea and guanidinum hydrochloride. Guanidinium hydrochloride was excluded due to its ionic nature. The use of urea, however, is applicable, as it is not retained by ion exchange columns (see Figure 18).

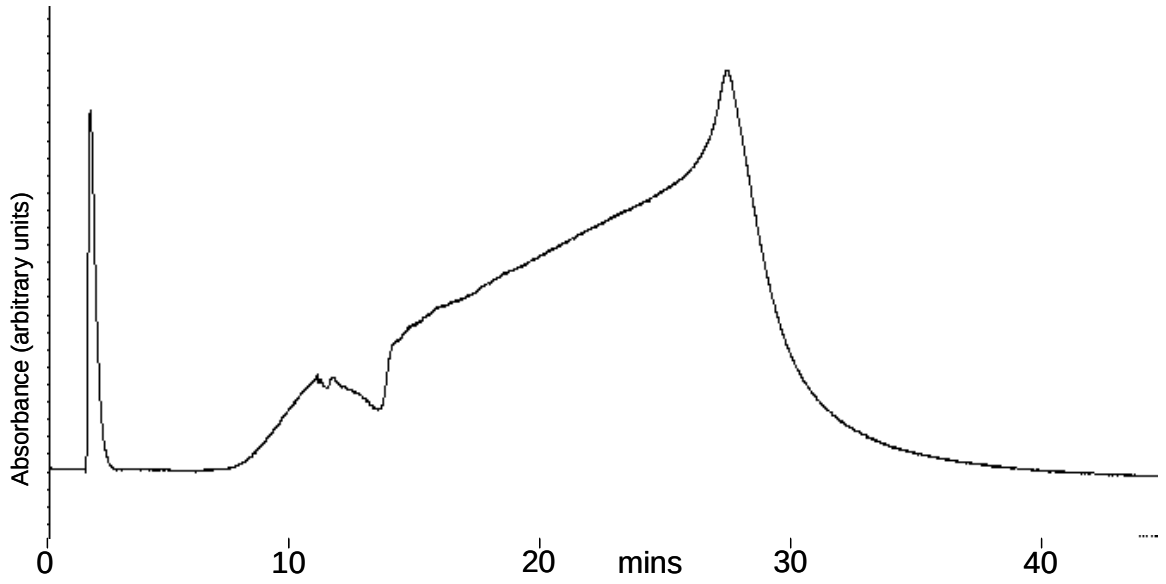


Figure 18: 2 mm x 50 mm cation coupled to 2 mm x 250 mm anion exchange chromatogram of 0.6 M urea. The sharp peak in the 218 nm UV spectrum at the start of the chromatogram demonstrates that urea is not retained by the column. The slope observed in the chromatogram likely to be due to contaminants in the NaCl used to produce the ionic gradient.

The selection of detergent is complicated by the superior performance of most ionic detergents for solubilisation: the detergent of choice for most resuspensions is sodium dodecyl sulphate (SDS); unfortunately this anionic detergent is retained by ion exchange columns until midway through the gradient at a point where many proteins elute (see Figure 19).

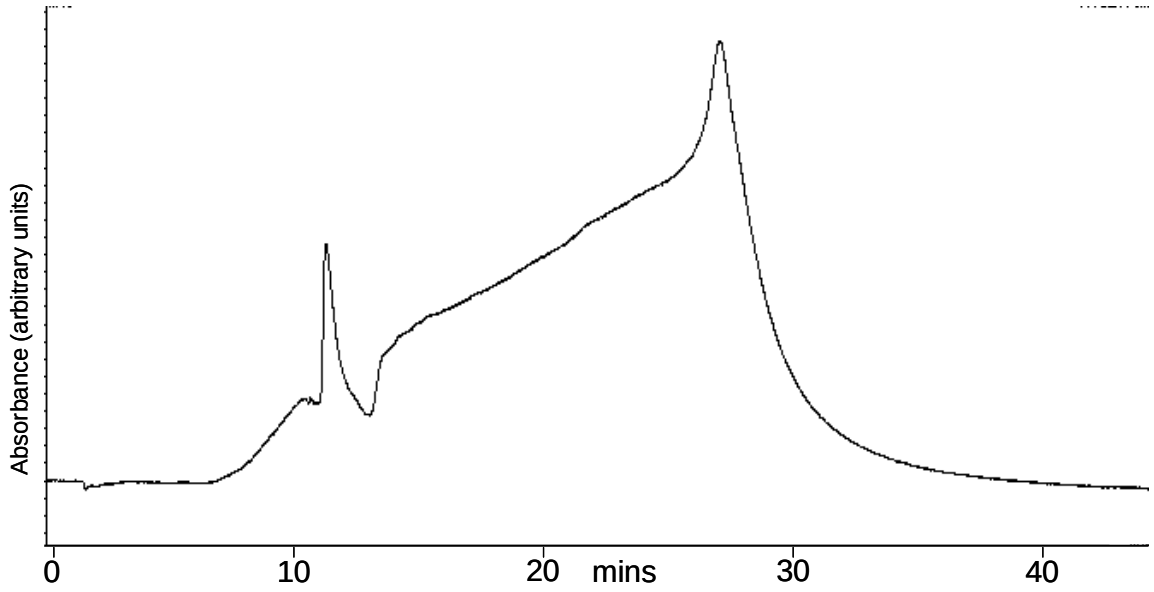


Figure 19: UV chromatogram of the separation of 20 µl of 0.1% w/v SDS by ion exchange using coupled 2 mm x 50 mm and 2 mm x 250 mm ProPac™ columns. Note the sharp peak at a retention time of 12 minutes resulting from the detergent. This is roughly the middle of a typical protein separation and will lead to contamination of collected fractions with concentrated detergent, rendering downstream analysis by mass spectrometry difficult.

It is therefore advisable to employ non-ionic, or zwitterionic detergents as resolubilisation agents to avoid contamination of protein-containing fractions with mass-spectrometrically incompatible detergent. The isoelectric focusing dimension of 2-dimensional gel electrophoresis is similarly restricted to zwitterionic detergents, and CHAPS is commonly used in buffers for this purpose. There was some concern the zwitterionic structure of CHAPS might result in retention on one or both of the cation and anion exchange columns used. Figure 20 shows, however, that it is not retained by either of the ion exchange columns at a buffer pH of 8.0.

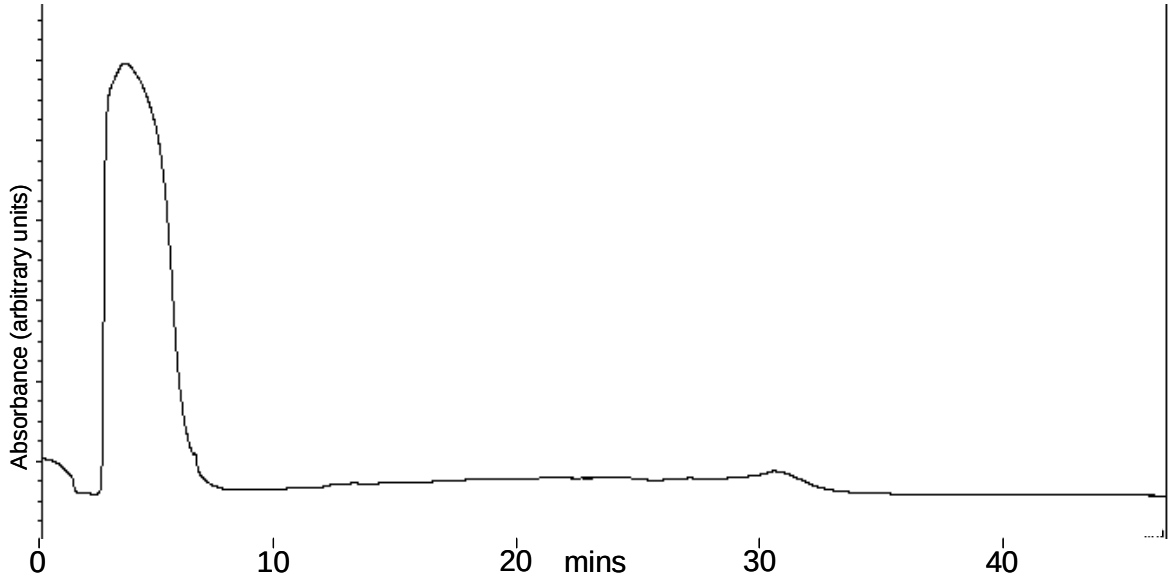


Figure 20: 214nm UV absorbance chromatogram of the ion exchange separation of 20µl of 0.1% CHAPS. The sharp peak at the start of the gradient corresponds to washthrough of the detergent.

A recent paper by Cass et al. [166] describing the use of the MuDPIT technique on schistosome eggs obtained from perfusion of infected mice [167], uses a standardized protocol for lysis of tough samples described in [24]. In this protocol, three sample fractions were produced by suspending cells in pH 7.6 phosphate buffer, containing a number of protease inhibitors and 0.25 M NaCl. Ribolysis (grinding with acid etched glass beads in a rotating machine) was then performed, supernatant was collected and adjusted to pH 8.5, followed by addition of 8 M urea, reduction, carbamidomethylation, and digestion with both Lys-C and trypsin. Finally, buffer exchange was carried out on the peptides using C18 solid phase extraction (SPE) cartridges. The remaining pellets, after removal of the supernatant, were washed either one or three times with PBS, followed by digestion with CNBr dissolved in FA, pH adjustment to pH8.5 and lyophilization to 200ul, followed by the protocol used for the soluble fraction.

This is a very thorough lysis protocol that results in three separate samples to analyse. Orthogonal digestions are used as an intrinsic step in the protocol – especially the formic acid/cyanogen bromide digest, commonly used in membrane protein experiments, which is used not only to digest the very hydrophobic peptides from membrane proteins but also to solubilise them. Save for the ribolysis, this protocol is perfectly applicable to any sample lysis procedure, but the increased time of analysis due to the multiple fractions, and the necessity of removing the integral digestion step will limit its use for the generation of intact protein samples.

As an example of lysis conditions used for two dimensional electrophoresis Luu [168] describes preparation of trypanosomes (closely related to the *Leishmania* samples used in this study). Lysis was performed by suspending cells in DIGE lysis buffer, which consists of 6 M urea, 2 M thiourea, 10 mM tris and 4% w/v CHAPS. The combination of 2 M thiourea and 6 M urea may be more effective at solubilising membrane proteins than 8 M urea alone. The preparation of lysates for 2D PAGE has been considerably standardized over the past 30 years, and the non-ionic constituents of lysis buffer are highly compatible, both with isoelectric focusing used in in the first dimension and the ion exchange first dimension used for intact protein LC.

An exemplar of lysis conditions used for SDS PAGE separation of carp muscle proteins followed by digestion and LC-MS/MS of the resulting peptides is described by McLean

[169]. Interestingly, the lysis performed in this instance was mechanical homogenization of tissue in ice-cold 20 mM phosphate buffer, pH 7.0, followed by clarification of the supernatant by centrifugation and analysis by SDS-PAGE. The great advantage of SDS-PAGE as a separation methodology is that SDS is exceptionally good for resuspending proteins, and retaining them in solution, and therefore performing lysis in a buffer free of SDS, or, in fact, any detergent at all, is likely to retain only very soluble proteins. Additionally, no purification of the samples is performed, which would be likely to result in difficulties of analysis using liquid chromatography, as non-protein contaminants will occur in collected fractions. However, small molecules should pass through a gel without incident.

3.2.3.3. Digestion Conditions

Initial analyses of bacterial lysates suggested that experimental conditions were non-optimal, with only 104 proteins identified with confidence from 55 MALDI spots derived by directly spotting peptide from 384 well plates used for sample collection. Some spots (most notably 40-52) contained no observable proteins. Identification of the complex mix of proteins in each *B. bronchiseptica* RP fraction was difficult for three reasons: efficiency of digestion was poor, resuspension of centrifugally vacuum dried fractions was inefficient, and the amount of recovered material was low. The addition of organic solvent to a protein-containing solution has a denaturing effect, and consequently, 10% v/v acetonitrile was incorporated into the proteolysis buffer in subsequent digestions. Following this modification to the digestion buffer significant improvements in identification rates and scores were observed. Later optimizations of the protocol involved using a Tecan Genesis 150 workstation to resuspend each sample with the trypsin solution, rather than a repeating pipette. This proved to be not only faster and more reproducible, but also provided a more consistent digestion.

As a source of comparisons in the literature with the method of digestion described, the most appropriate are those digestion strategies applied to sample preparation for separation using the PF2D system (Beckman Coulter). The PF2D system is a chromatofocussing system (described further in 4.1.2), coupled to a non-porous silica (NPS) C4 reversed phase separation system. It is used for two dimensional LC separation of proteins and is therefore the most similar methodology to that described in this chapter. Chahal et al. [170] describe a methodology for analysis of tumour-associated membrane antigens using the system. Following separation, samples were concentrated by lyophilisation to ~100 µl followed by reduction and alkylation and digestion for 4-6 hours. Ying et al. [171] describe

the use of the PF2D system for the analysis of human fetal liver samples. Their digestion strategy, in common with Chahal and MacDonald's is imperfectly described as simply a solution digest applied to the liquid sample. Zhao et al. [172] use the PF2D system for analysis of pre-malignant breast cells. After separation, fractions are dried down to ~20µl volume, diluted to 40 µl with 100 mM ammonium bicarbonate and 0.5 µl of trypsin is added to the sample, followed by incubation at 37ºC for 20 hours.

In general, therefore, researchers tend not to apply special attention to the digestion procedure following intact protein LC, except to alter the pH of the sample. A simple improvement would be the addition of organic solvent to the digestion buffer, which is known to considerably improve digestion due to the disruption of the structure of the protein. In the situations described, no lyophilisation is performed on the samples, which makes them more difficult to manipulate, and more susceptible to freeze-thaw shearing, however, this may improve the recovery of proteins upon analysis, due to the lack of aggregation upon drying [173].

3.2.3.4. Critical Micelle Concentration

The concentration of non-ionic detergents is normally increased to compensate for their poor resolubilisation characteristics in comparison with ionic detergents. Unlike gel-based isoelectric focusing, chromatography columns are sensitive to the formation of lipid micelles – spheres of detergent with the polar surface facing the solution and a hydrophobic interior. Solubilised proteins are sequestered in micelles and their elution profiles are significantly affected during chromatography due to their lack of interaction with the stationary phase. The detergent concentration at which micelles form is known as the critical micelle concentration (CMC).

CHAPS has a CMC of ~0.4%, and it was found that resuspending in 2% w/v CHAPS and diluting the sample 1:10 or more in ion exchange buffer prior to separation was normally sufficient to provide good separation. However, high salt concentrations decrease the critical micelle concentration of detergents [174] and in some cases, this procedure was insufficient, presumeably due to insufficient dilution in low-salt buffer or carryover of detergents from the precipitation step. In later separations, the 2% w/v CHAPS in the resuspension buffer was replaced with 0.4% w/v CHAPS, which provided indistinguishable resuspension ability, but, after dilution with low-salt buffer, resulted in a negligible total concentration of detergent. Due to this improvement, the 10% v/v

acetonitrile was excluded from the ion exchange buffers as it was unnecessary and affects binding to the reversed phase column in those cases where the gradient starts at a lower organic concentration than that normally used for proteins (<15%).

The solubility enhancements due to detergent have been extensively studied. For example, Watarai [175] describes the complementarity of proteomic analyses based on sample solubilisations with Triton X-100, CHAPS and SDS. Interestingly, the latter preparation was performed by initial lysis into 10µl of 1% SDS buffer, followed by dilution with a large volume (350ul) of standard 4% w/v CHAPS buffer, which works effectively, and may be the source of further improvement to the intact protein LC protocol.

3.2.3.5. Reduction of Initial Sample Amount

Proteomic samples are often rare or difficult to produce, and an effective analysis technique must be applicable to small amounts of sample. The amounts required for traditional techniques such as shotgun 2DLC (≤100 µg) and 2D PAGE (400 µg) are relatively small, and although the capacity of ProPac™ columns is compatible with up to 10 mg of protein, it must also be capable of separations in the range of 100-400 µg to be a competitive method. The main limitation was the requirement of a maximum 60 µl sample loop volume for the reversed phase monolith, to prevent excessive gradient mixing delays due to the direct injection implementation. The use of a high capacity PS-DVB stationary phase trap would eliminate this requirement but these are not currently available. 2 mm ProPac™ columns were run at 200 µl with a collection rate of 1 fraction per minute for a standard high resolution ion exchange separation. With a 200 µl fraction size and 60 µl loop size the total load volume for reversed phase separation is limited to roughly one third of each fraction. In order to obtain sufficient material for MS analysis this limit must be compensated for by tripling the protein load.

To allow the use of a 200 µL loop, a separation program that used the 500 µm RP monolith was written that allowed 10 minutes of isocratic aqueous flow to clear the sample from the loop and retain it on the column. Subsequently, the loop is switched out of line with the pump and the gradient proceeds normally. This modification enabled reduction of sample loss significantly, and resulted in a reasonable minimum protein load of 200 µg. Most published examples of liquid-phase intact protein separation technologies used substantially more starting material to achieve protein identifications. For example, two dimensional LC of intact proteins as described in this thesis compares favourably to the

PF2D system, which is commonly described as requiring an initial sample amount of 1-5 mg protein. A typical experiment is described in [176], where 2.5 mgs was used as the initial sample amount. Alternative methods, such as acid labile surfactant (ALS) - PAGE/RP use even larger quantities of protein. An example is described in [177] where one gram of *Saccharomyces cerevisiae* cells was used as the initial source of protein. Although these methods are ideal when dealing with samples that are easily produced in culture on inexpensive medium, there is difficulty in producing these large amounts of protein from clinical samples. Coupled with the increasing requirement for experimental replicates, the ability to perform multidimensional separations with a decreased requirement for large amounts of starting material becomes important.

3.2.3.6. Separation Quality Assessment

Once sample loading and buffer composition were optimized, the technique was applied to multiple standard mix samples for assessment of separation quality and to more complex samples for true proteomic analysis using a variety of identification methodologies. Assessment of separation quality was determined by a number of metrics: reproducibility, chromatographic peak width, analysis of fractions by gel electrophoresis and staining, and finally analysis by tryptic digestion followed by peptide mass analysis with mass spectrometry. Both types of column (pellicular PS-DVB ion exchange and monolithic PS-DVB reversed phase) were assessed in terms of their relative standard distributions (RSDs) of retention time and peak widths. SAX columns display an RSD of 1.6% on the most intense peak derived from 6 consecutive runs over a period of 5 hours, while coupled SAX and SCX columns show an RSD of 0.991% on the most intense peak derived from 7 consecutive runs over a period of 6 hours. A column oven was not available for these experiments, and improved RSDs may result from ensuring temperature stability during each run. Peak widths ranged from 0.18 min to 0.45 min at half height. The latter broad peaks widths are the result of shouldered peaks which demonstrate partial separation (see Figure 21). Monolithic reversed phase columns display RSDs from 0.161% to 0.282% with half-height peak widths of an average of 0.4 mins (see Figure 22). Interestingly, increasing the gradient time for the SAX column resulted in significant deterioration of peak shape and signal to noise, thus the standard 20 minute gradient appears the most efficient.

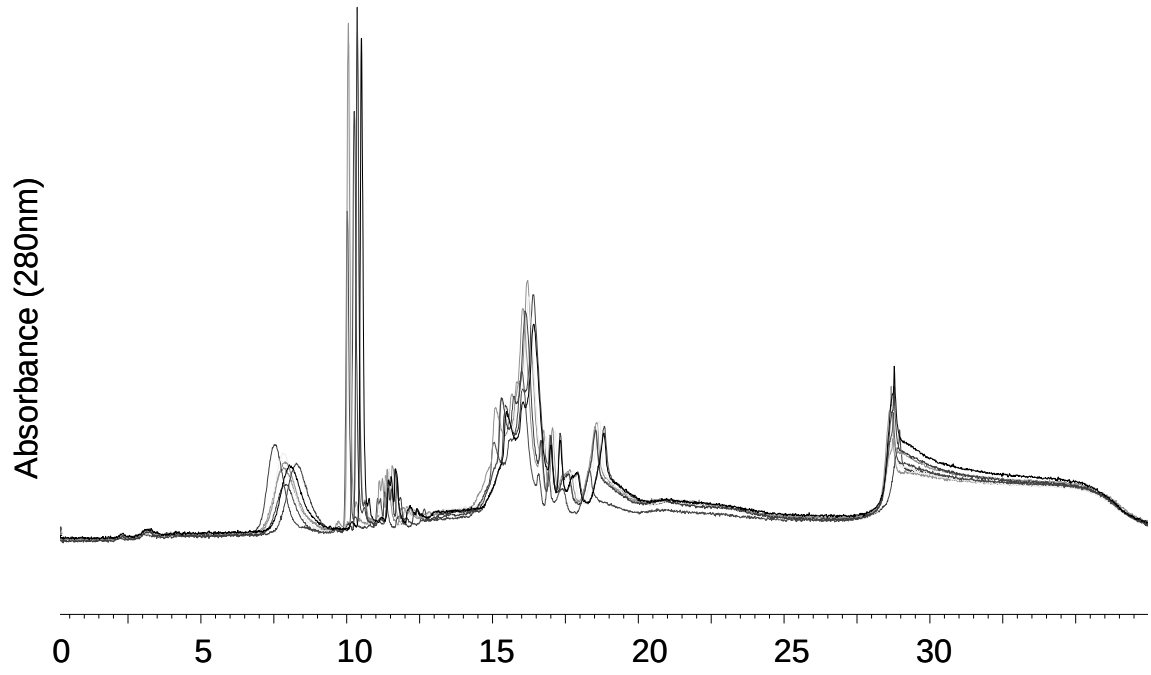


Figure 21: Overlays of successive runs on coupled SAX and SCX ion exchange columns. The sample used was a mixture of 9 proteins spanning a wide range of pIs and molecular weights. Peak widths are generally narrow, except for the shouldered peak at 15 minutes which is a consequence of isoform separation of ovalbumin, one of the constituent proteins.

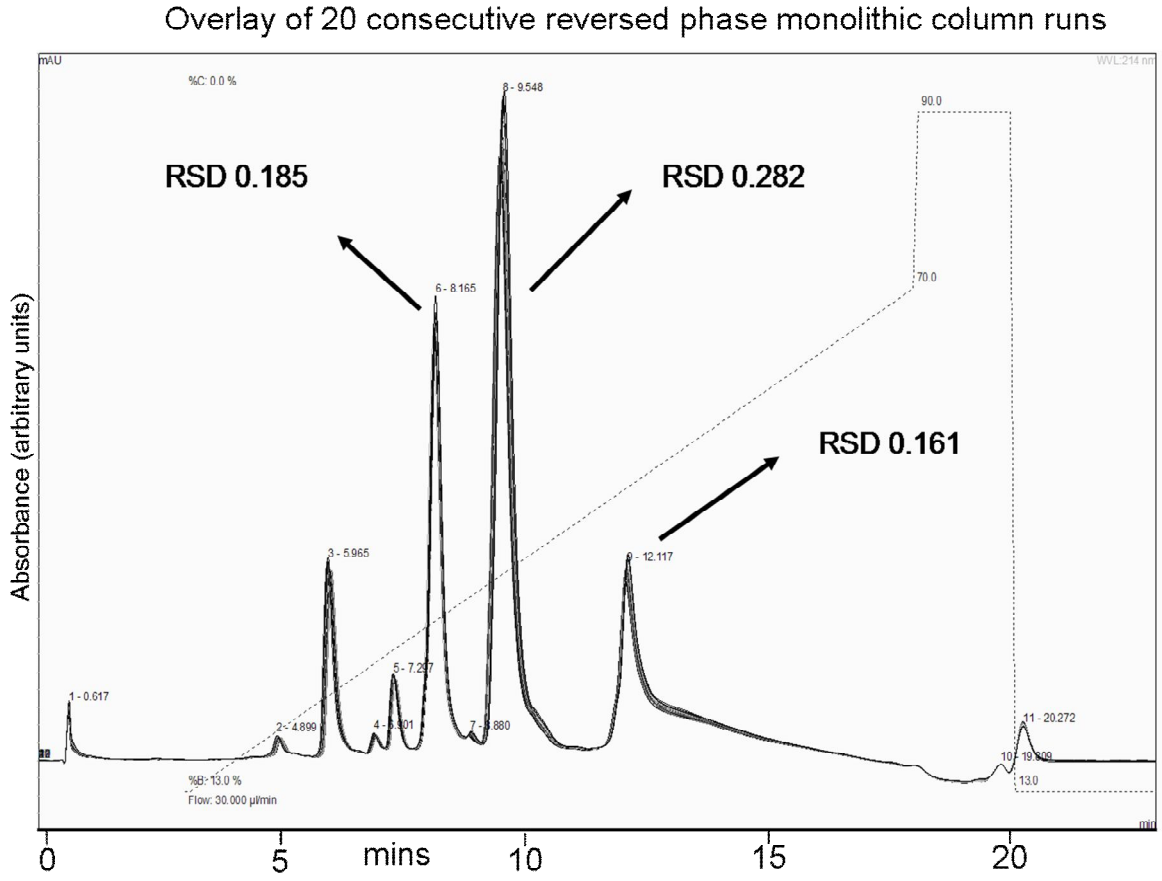


Figure 22: Overlays of successive runs on a 500µm monolithic PS-DVB reversed phase column. Sample used was a mixture of lysozyme, cytochrome c, myoglobin, trypsin inibitor, BSA, ribonuclease A and ovalbumin. Peak widths are narrow (~20 seconds) and reproducibility is excellent, at less than 0.5%. Chromatogram produced in collaboration with Cook, 2007.

3.2.3.7. Ion Exchange Separation

A 2 mm x 50 mm SCX column coupled to a 2 mm x 250 mm SAX column was the column combination used for all initial ion exchange separations. For small sample amounts a 1 mm x 250 mm anion exchange column was used as an alternative. Column capacity for a single protein was assessed (see Figure 23) using the coupled 2 mm x 250 mm ProPac™ and 2 mm x 50 mm ProPac™ columns, demonstrating that for each species, a maximal load of the order of 10 µg should be observed, and that increasing this to 100 µg resulted in substantial peak broadening.

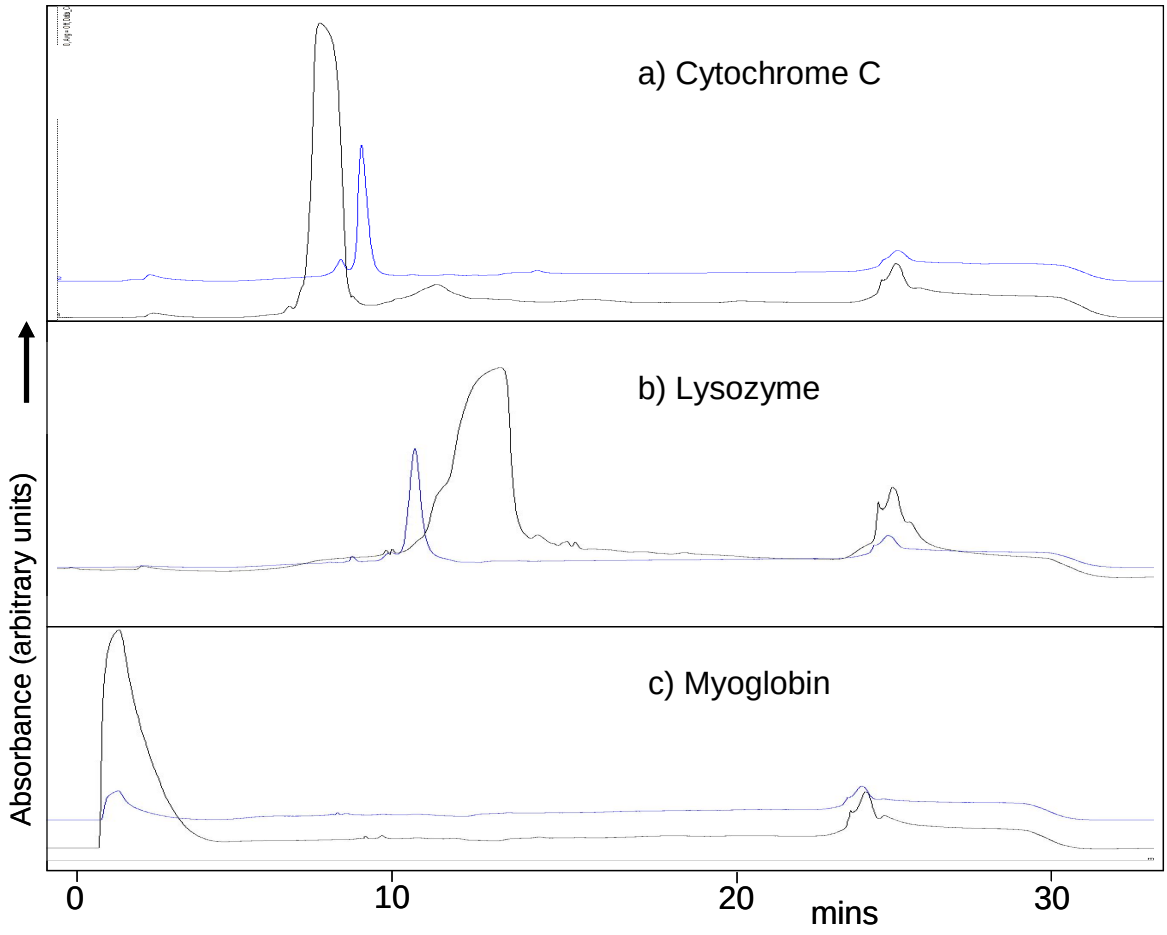


Figure 23: Three images of test proteins separated using coupled ProPac™ 2 mm x 250 mm anion exchange and 2 mm x 50 mm cation exchange columns. Chromatogram (a) shows a separation of cytochrome c, (b) shows a separation of lysozyme, and (c) shows a separation of myoglobin. In each case, the blue trace corresponds to a 10 µg protein load, and the black trace corresponds to a 100 µg protein load. Peak broadening occurs when 100 µg of a single protein is loaded on the column, whereas a sharp peak is maintained with a load of 10 ug.

To provide a more detailed analysis of separation than that provided by UV chromatographic traces of standard proteins, a separation of 3.6 mg (a comparatively high load) of *Bordetella bronchiseptica* lysate was performed by coupled SAX and SCX ion

exchange (see Figure 24) followed by fraction collection and monolithic reversed phase separation. UV absorbance at 214 nm and 280 nm shows a very complex separation, and subsequent silver stained gels of ion exchange fractions show that individual undigested fractions remain both extremely complex and that each fraction is very different from the next, with the exception of fractions 12-15, which show identical bands but decreasing concentrations of protein due to the long-retention washthrough peak (see Figure 25).

Digestion was performed on collected fractions and peptide intensity maps were generated across the ion exchange separation and the reversed phase separation were generated using MALDI-ToF mass spectrometry.

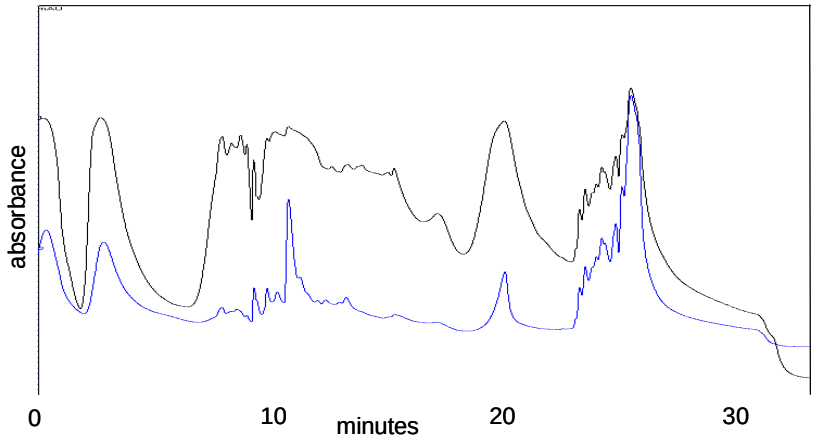


Figure 24: Ion exchange chromatogram of a 3600µg *Bordetella* BHO mutant lysate. Note the flattening of peaks around 1000-1100 absorbance units, indicating that saturation of the detector is occurring. The black trace corresponds to the 214nm absorbance wavelength, the blue trace to 280nm. Despite the overloading, the 280nm absorbance chromatogram shows that even under heavy load conditions, performance of the columns is still good.

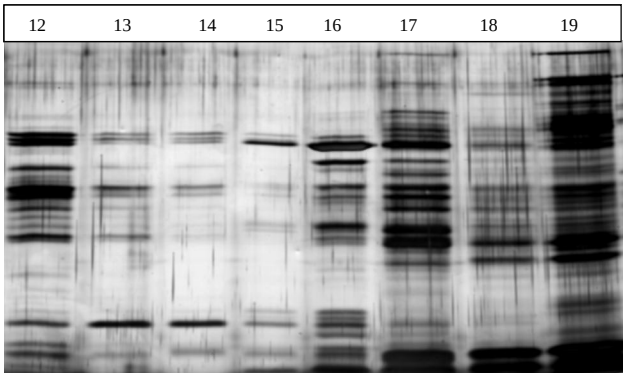


Figure 25: Silver stain of fractions derived from the ion exchange in Figure 24. The majority of proteins are visible in one or two fractions, although some appear to be present throughout. It is clear, despite high resolution separation, that each ion exchange fraction remains extremely complex. The numbers correspond to the fraction number collected from the above ion exchange separation. Fractions were collected at a rate of one per minute, and the area covered by the gel is from the centre of the complex region. Despite this, the first four fractions shown on the gel resemble gradually eluting washthrough proteins.

By obtaining a mass spectrum from each fraction, it is possible to derive a map of separation quality, based not on chromatogram peaks, which may contain multiple species, but on mass spectrometry of individual proteins, following their post-separation digestion.

Once fractions have been digested and analysed in both ion exchange and reversed phase dimensions, unequivocal mass spectrometric data shows the generally good separation provided by the technique. Figure 26 shows a section from a 'pseudogel', generated by graphing ion exchange fraction number against mass, with greater intensity bands denoting more abundant peaks from the mass spectrum. Peaks have not been de-isotoped, and bands circled in red and linked belong to the same peptide. Spectra were determined from LC-MS/MS chromatograms obtained using the QStar Pulsar i electrospray quadrupole time-of-flight instrument. Total ion chromatograms were summed to produce a single spectrum from each fraction. Most peptides are identified in only one or two fractions, and a degree of peak-splitting is inevitable during fractionation which accounts for many of the two fraction peaks. A few peaks persist in a large number of fractions, suggesting either overloading or isoform separation.

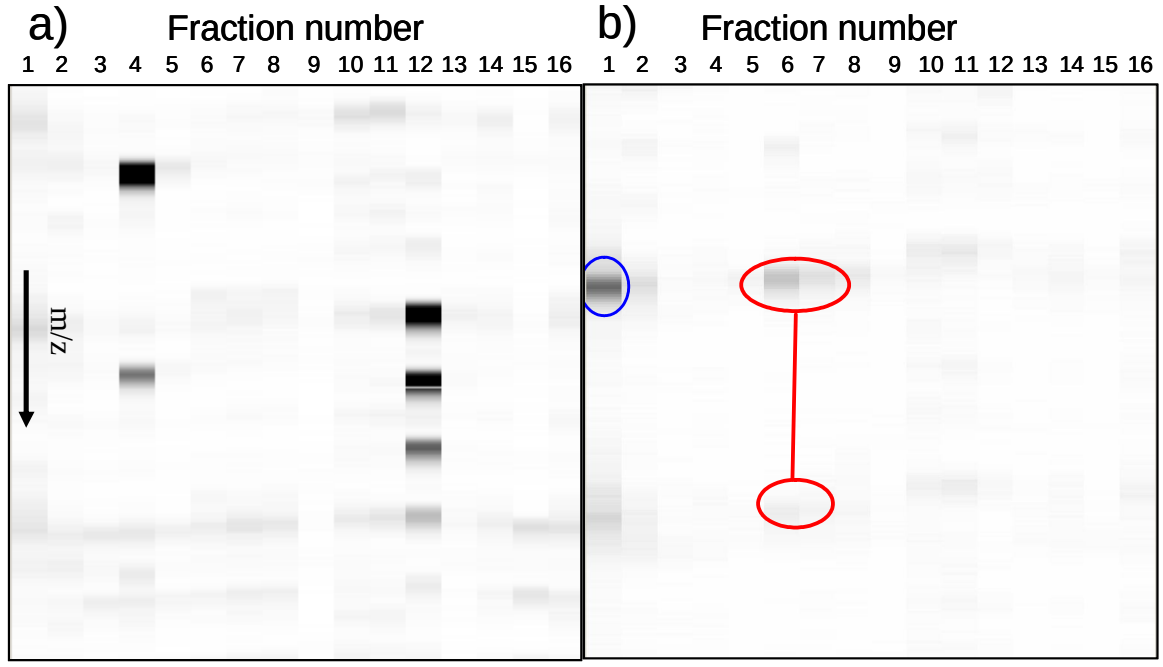


Figure 26: Two 'pseudogels' demonstrating ion exchange separation quality. To generate the pseudogels, MS data was summed from fractions run by LC/MS, and each resulting spectrum was displayed in ascending mass from top to bottom. Darker 'bands' correspond to more intense peaks. Images shown are close-ups of particular sections of interest from the spectra. Monoisotopic and subsequent isotopic peaks from peptides of interest are circled in red. Image a) displays the 564.5 amu peak from an autotransporter, b) a 576.5 amu singly charged peak.

3.2.3.8. Separation of Isoforms by Ion Exchange Chromatography

One of the major benefits of this technique over peptide 2DLC is the separation of protein isoforms. To demonstrate isoform separation, 40 µg undigested ovalbumin from *Gallus gallus* was separated using ion exchange chromatography on a 2 mm x 250 mm ProPac™ SAX column and compared to an identical alkaline phosphatase treated sample (Figure 27).

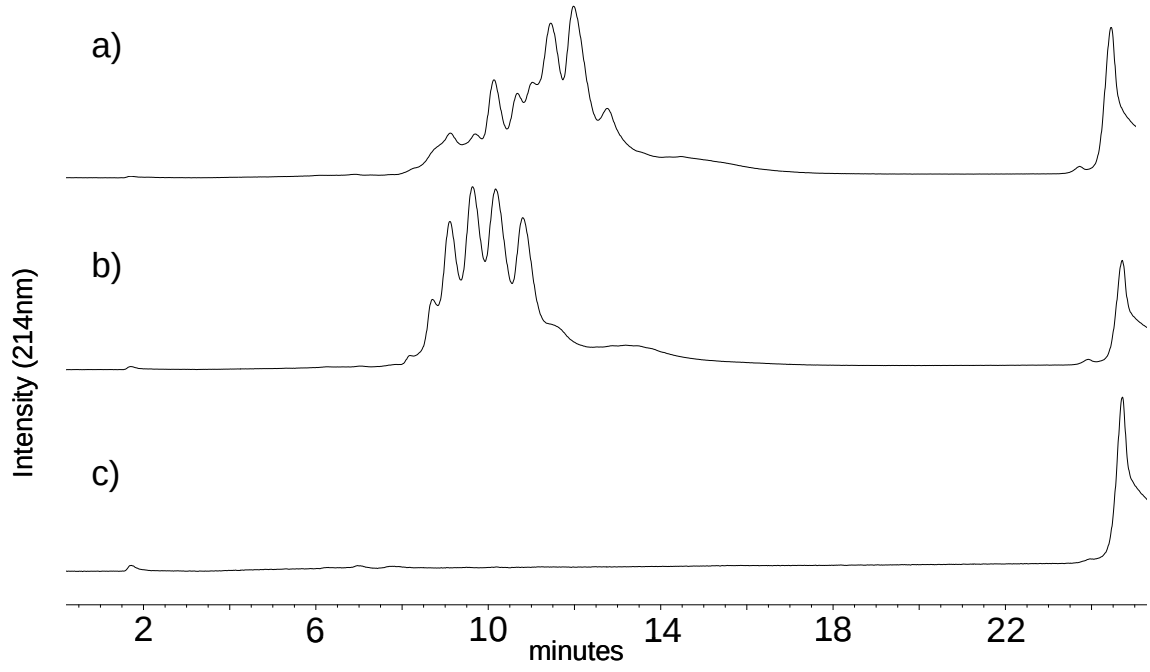


Figure 27: Ovalbumin separation by derivatised PS-DVB ion exchange column. A) is untreated ovalbumin, B) is ovalbumin treated with alkaline phosphatase, and C) is a control showing an equivalent amount of alkaline phosphatase to that used in the treated sample, to show that none of the peaks result from the enzyme. Note the reduction in the number and change in intensity of the peaks in B, as compared to A.

A reduction was observed in the complexity of the chromatogram, the remaining complexity likely being due to differential glycosylation [178]. Further, fractions from an ion exchange separation of 10µg of unmodified ovalbumin using a 1 mm x 250 mm ProPac™ SAX column were collected and subjected to tryptic digestion, followed by analysis by MALDI-MS and MS/MS. A known phosphopeptide with a mass of 2088.9 Da (subsequently confirmed by de-novo sequencing by Laetitia Mouls) was tracked in multiple, high resolution fractions, increasing and decreasing in intensity, matching the UV trace (see Figure 28). The corresponding unmodified peptide (mass 2008.9 Da) mirrored the change in intensity of the modified peptide.

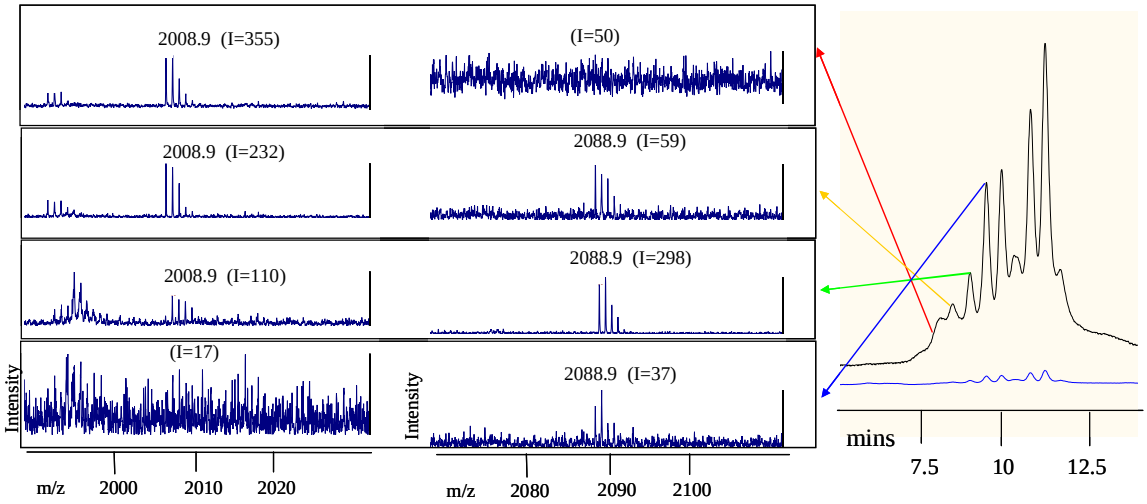


Figure 28: Isoform separation of ovalbumin. The right panel shows the peaks derived from a 40µg separation of Ovalbumin from *Gallus gallus*. The left panel is split into two columns. In the left column, an unphosphorylated peptide is tracked from serial fractions corresponding to 4 successive peaks. In the right column, the equivalent phosphopeptide is tracked in a similar manner. The intensity of the unphosphorylated peak decreases as the phosphopeptide increases over the 4 fractions, corresponding to a 2 minute chromatographic time period.

3.2.3.9. Reversed Phase Separation

The reversed phase second dimension separation is performed with 500 µm x 5 cm polystyrene divinylbenzene monolithic columns. The capacity of these columns is roughly 3 pmol protein without peak broadening and is therefore compatible with the estimated 4 picomoles of average (40 kDa) protein per one minute fraction from a standard 200 µg protein load. Additionally, for samples with limited complexity, a single dimension separation may be sufficient for a proteomic experiment – analogous to a liquid phase gelC/LC experiment. For example, IPs may contain up to 200 proteins and are often heavily contaminated with immunoglobulins. This is an area where intact protein LC should be particularly useful, as the immunoglobulins should be confined to a limited number of fractions, allowing unimpeded analysis of more biologically relevant proteins. An example chromatogram of an IP separated by reversed phase monolith separation is shown in Figure 29.

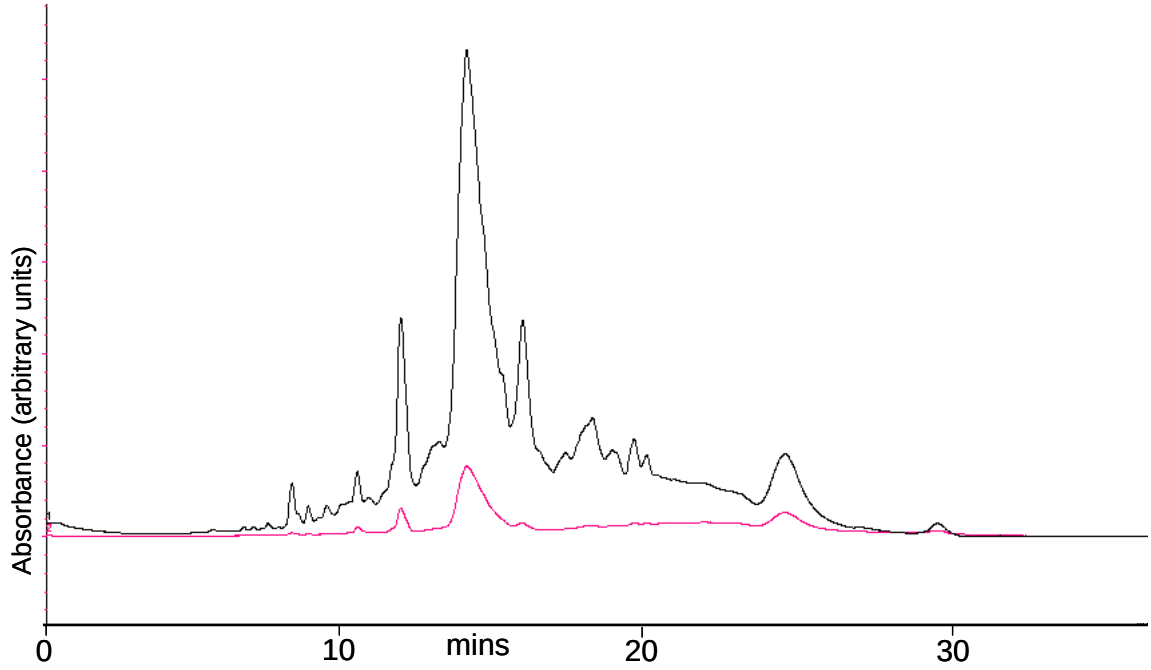


Figure 29: Separation of an immunoprecipitate by 500µm x 5cm monolith only. The very strong peak in the centre of the run corresponds to immunoglobulin G. The black line refers to UV absorbance at 214 nm, the red line absorbance at 280 nm.

The separation quality of the reversed phase monolithic columns was assessed in the same way as the ion exchange columns. Figure 30 shows a 16 fraction PS-DVB monolithic reversed phase separation of an ion exchange fraction. After digestion of separated proteins, peptides were tracked across the plate and assessed for presence and intensity in each well. Like the ion exchange separation, most peptides are present in one or two fractions, but some are present in several, probably a result of peak tailing of highly modified and/or abundant proteins.

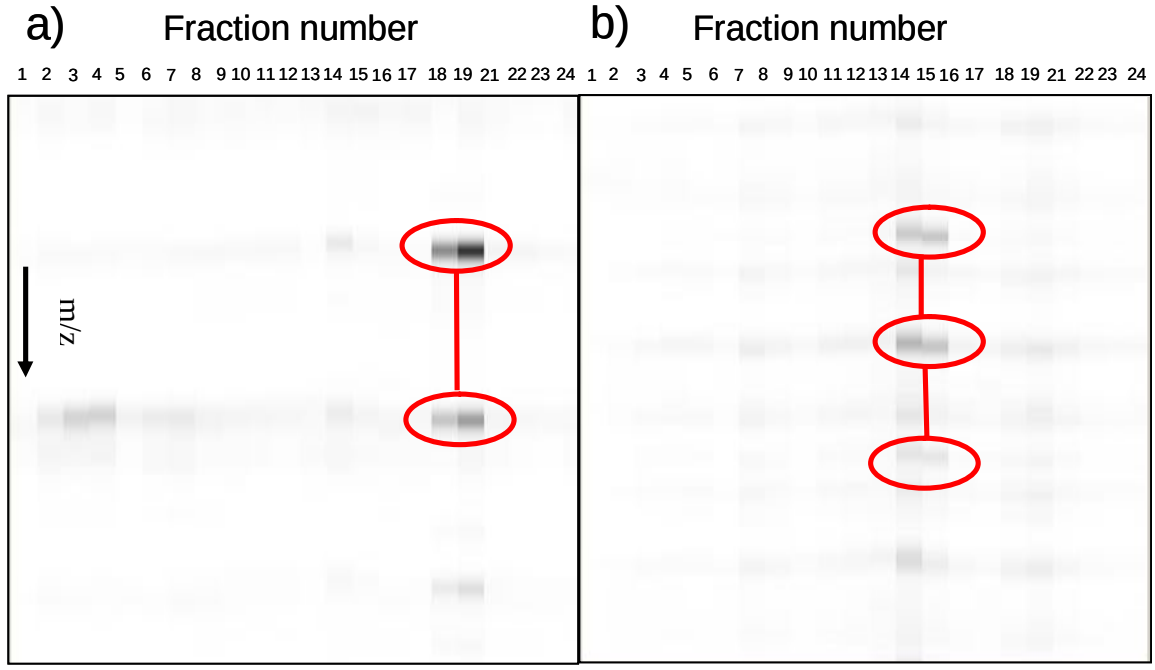


Figure 30: Pseudogels demonstrating reversed phase protein separation ability. Please refer to Figure 26 for a description of the graphic. The image on the left highlights the 458.75amu peak from the Trigger Factor protein. The right image highlights the 782.91 peak from alkyl hydroperoxide reductase.

3.2.3.10. Overall Separation

Good separation quality has been shown for individual dimensions, showing that the individual performance of the columns for complex mixture separation is acceptable. As a two dimensional separation technique, it is meaningful to map protein elution profiles in two dimensions. Both stackplots and heatmaps (see section 6.4.1.1) were used to represent UV data from complete two dimensional separations. Figure 31 represents the 24 reversed phase separations performed on fractions taken from the ion exchange separation of 3.6mg *Bordetella bronchiseptica* lysate.

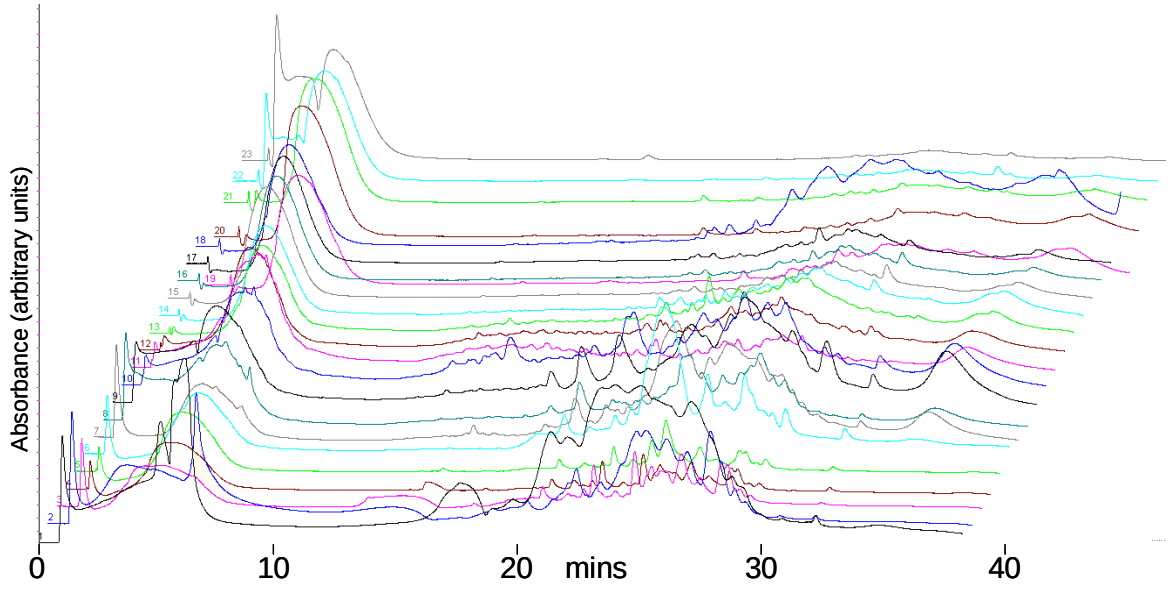


Figure 31: UV data collected from the separation of 24 fractions from the 3600 µg *B. parapertusis* BHO mutant ion exchange. Although broad peaks occur in some traces, the majority are sharp, suggesting that broad peaks are derived from the joining of multiple protein peaks.

Once identifications had been completed by rapid peptide analysis using LC-MS/MS, a few peptides were selected for a more in-depth analysis. Peptides were chosen from identified proteins and their maximal ion abundance within a mass window of ± 0.5 Da and a retention time window of ± 0.3 min was mapped throughout the collected plate (see Figure 32). There is some inevitable smearing of very abundant proteins, but most proteins are confined to one or two fractions in both dimensions.

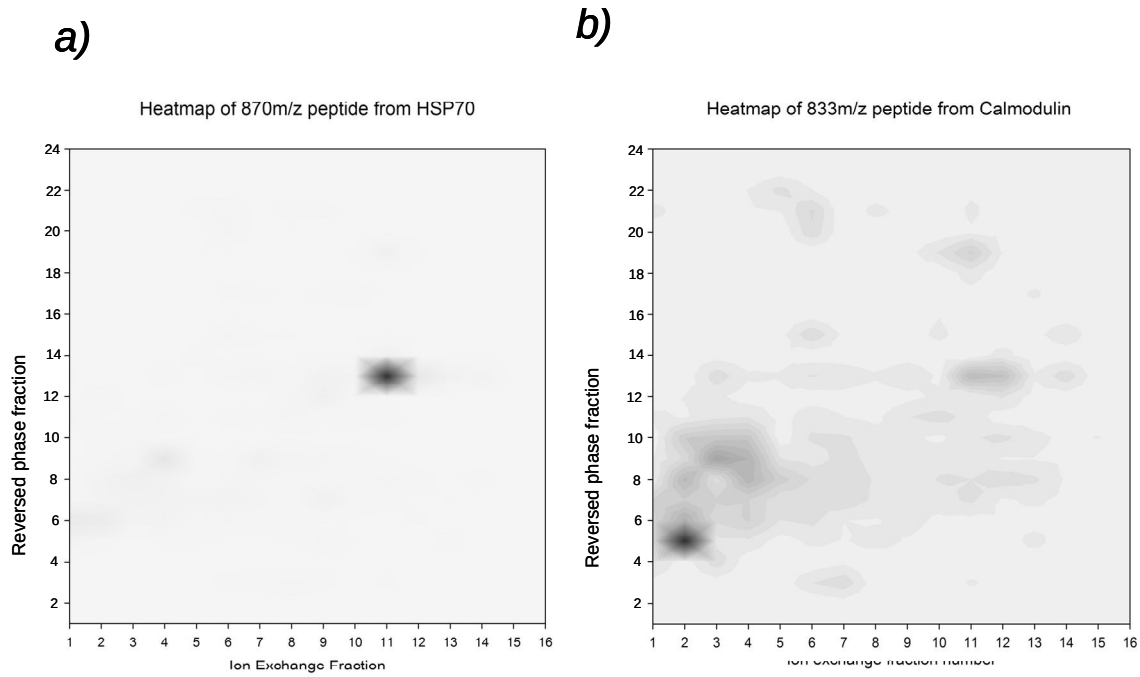


Figure 32: Heatmaps of peptide abundance throughout a 16 (ion exchange) x 24 (reversed phase) fraction intact protein separation. Darker spots indicate higher abundance. The 833 m/z peptide from calmodulin is highly intense in fraction 2, 5, but appears in multiple, less intense, localized spots in different areas, suggesting multiple peptides with a similar mass (within the 0.5 m/z window used for selectivity). The 870 m/z peptide from HSP70, however, is localized to a single, fraction.

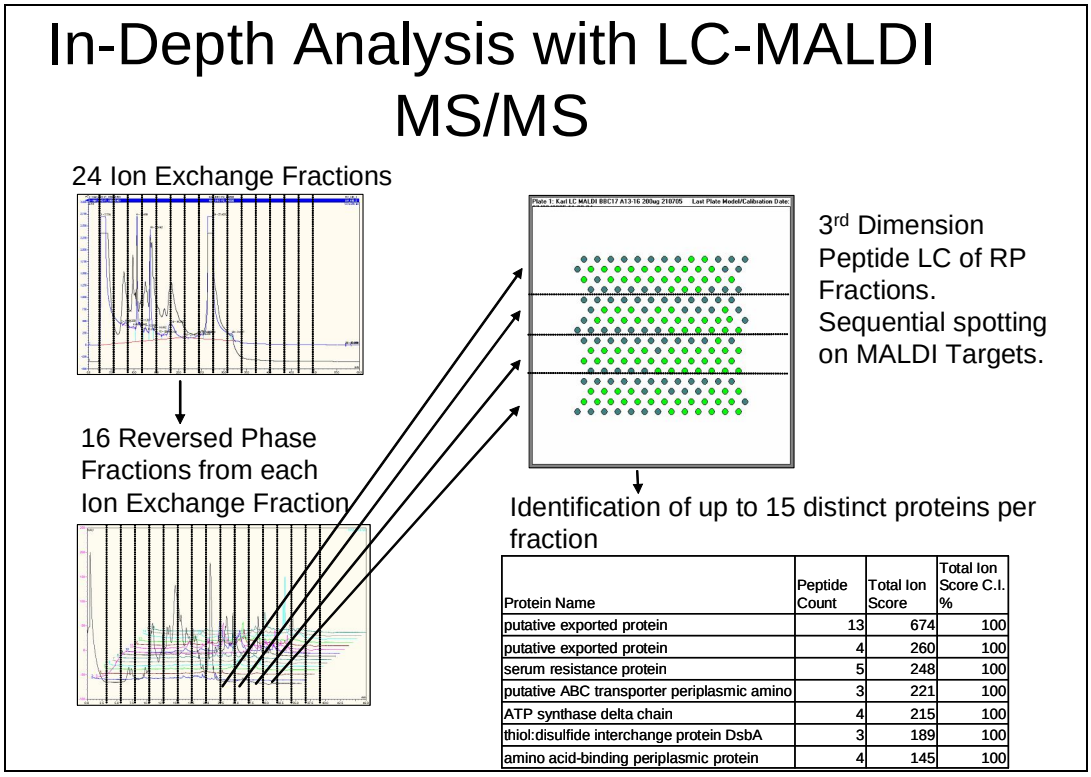
3.2.3.11. Data Analysis

A variety of methods were tested for the analysis of separated proteins, including direct MALDI, LC/MALDI, traditional LC-MS/MS, and fast peptide LC-MS/MS.

In general the range of proteins identified is good, with a variety of pIs and molecular weights. Analysis of the identified proteins for their codon adaptation indices (described in section 3.2.3.11.3) indicates that 2D intact protein LC is of equivalent dynamic range to other global proteomics techniques such as 2D PAGE or shotgun 2DLC.

3.2.3.11.1. Analysis of a Bacterial Lysate by 2D Intact Protein LC and LC/MALDI

200 µg of a *Bordetella bronchiseptica* lysate were separated and proteins were collected in a 384 well plate. After digestion, each well was further separated in the peptide dimension by conventional C18-based reversed phase chromatography and spotted out with CHCA for MALDI analysis. A schematic of the LC/MALDI process used for the peptide separations is shown in Figure 33.



Protein Name	Peptide Count	Total Ion Score	Total Ion Score C.I. %
putative exported protein	13	674	100
putative exported protein	4	260	100
serum resistance protein	5	248	100
putative ABC transporter periplasmic amino	3	221	100
ATP synthase delta chain	4	215	100
thiol:disulfide interchange protein DsbA	3	189	100
amino acid-binding periplasmic protein	4	145	100

Figure 33: Schematic of the LC-MALDI process for peptide analysis from 2D-protein separations. 24 ion exchange fractions were collected and then reinjected for second dimension separation using a 500µm monolithic column. 16 fractions covering the entirety of the separation were collected in a 384-well plate. Each of the fractions was subjected to tryptic digestion and reversed phase separation using a standard C18 column (PepMap, Dionex, Camberley UK). 48 thirty second fractions were collected on MALDI target plates, with 4 LC runs per plate. Analysis was performed using MALDI MS/MS, followed by identification using MASCOT 1.9 (Matrix Science, London). The inset table shows an example of the proteins identified from a specific LC run, where peptide count refers to the number of peptides observed from a specific protein, total ion score is the combined MOWSE scores of all the peptides observed, and Total Ion Score C. I. is the confidence interval of the result.

A very complex first dimension was observed (see Figure 34), which was further resolved by the second dimension. Figure 35 shows the reversed phase separation derived from ion exchange fraction 12. The fraction was separated using a gradient from 15-70% acetonitrile in acidified water (0.08% v/v TFA) over 20 minutes. Individual proteins can be observed as sharp peaks, but the spectrum remains highly complex. However the limited number of fractions collectable requires that a third dimension separation, that of peptides, is employed to fully analyse the data. The analysis was performed using LC-MALDI ToF/ToF mass spectrometry as described in Methods 2.6.2. All spectra were submitted to the GPS software version 3.0 (Applied Biosystems) for identification, and subsequently, 249 proteins were identified with greater than 95% significance.

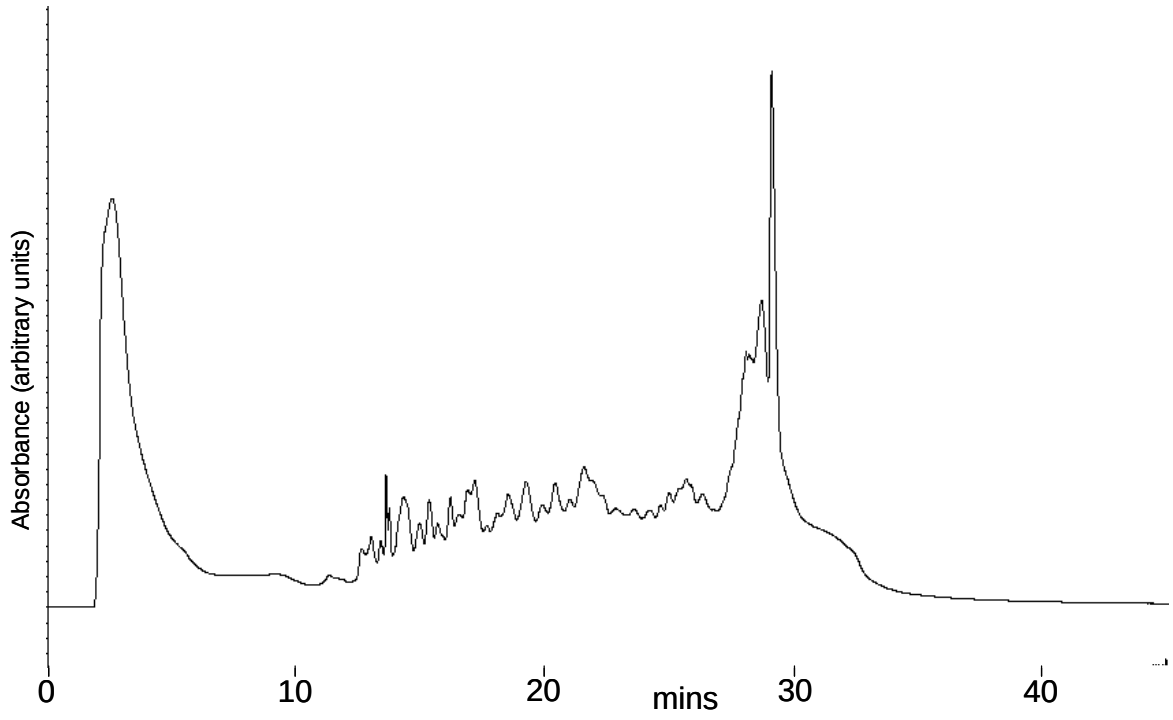


Figure 34: Anion exchange separation of 200 µg of *Bordetella parapertussis* lysate. Note the highly complex nature of the chromatogram. Although many peaks look rather broad, further inspection shows them to be the result of multiple separate peaks.

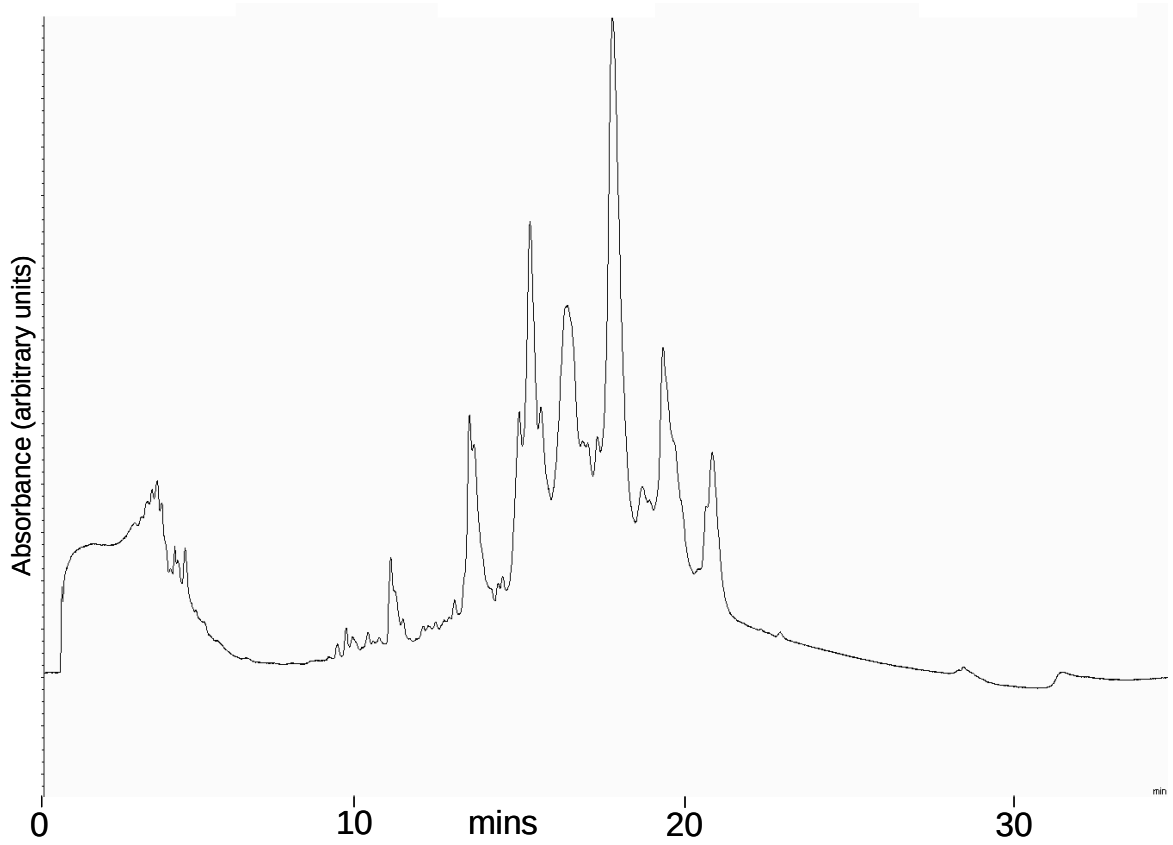


Figure 35: Second-dimension reversed-phase chromatogram of fraction 12 of the ion exchange chromatogram in Figure 34. Still extremely complex, in many cases individual peaks can be observed, with others merging to form broader peaks.

For comparison, a shotgun 2DLC (proteolytic digestion, followed by ion exchange and reversed phase separation) and GeLC (1D SDS PAGE separation of proteins, followed by digestion of gel chips and LC-MS/MS of peptides) of an identical sample (100 µg, which corresponds to the 200 µg 2D protein LC with a 50% load per LC MALDI run) was performed, resulting (in the former case) in 265 protein IDs above 95% significance and, in the latter case, in 187 IDs above 95% significance. Only 58 proteins were common to all three analyses, thus demonstrating that 2DLC of intact proteins is a complementary technique to existing methods (see Figure 36).

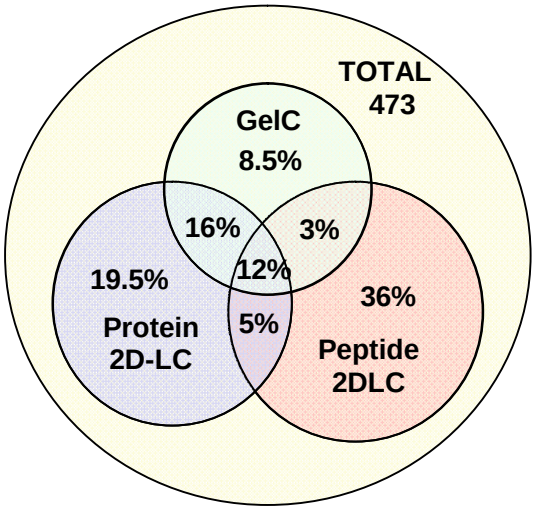


Figure 36: Comparison of two dimensional LC of *Bordetella parapertussis* proteins with GelC and shotgun 2DLC by number of identifications above 95% confidence threshold. A total of 473 proteins were identified using all three methods. 12% of these were complementary to all methods, and the MudPIT technique identified the highest number of unique identifications with 36%. The two intact protein separation techniques (SDS-PAGE separation followed by gel pixellation, digestion and peptide LCMS (GelC), and protein 2D-LC) were the most similar, sharing 28% of identifications.

Although a comparable identification rate to traditional protocols was acceptable, a number of aspects of the LC-MALDI-ToF/ToF methodology were considered deleterious to rapid, high quality identifications. Each of the 384 reversed phase fractions were separated by C18 Pep-Map column using a reversed phase gradient and spotted on a target plate. At a rate of one hour per fraction, roughly 16 days were required to spot all of the fractions on MALDI targets. 24 fractions were separated by LC and spotted across six MALDI target plates over a period of 16 hours. This provides a significant amount of time for contamination from dust and peptide breakdown to occur on the exposed plate. Additionally, the matrix mixed with the sample during spotting degrades when exposed to light for long periods of time.

Four fractions from the intact protein separation were digested and analysed by LC directly coupled to electrospray tandem MS using a QqToF instrument, showing an average of over 50 proteins identified with a significance of greater than 95% in each reversed phase fraction. 22% of these were shared between two fractions, 7.5% between three fractions, and 6% were found in all four fractions (for an example of one fraction, see Table 2).

Table 2: shows Mascot search results for fraction 6 of ion exchange fraction 12. Only hits with a score of greater than 100 are shown. The large number of very highly scoring hits demonstrates the high complexity of each sample.

Protein Name	MW	pI	Sequence Coverage	Peptides Identified	MASCOT Score	Gene Name
alkyl hydroperoxide reductase	20160	5.11	62%	9	308	BP3552
50s ribosomal protein l9	16373	8.92	35%	6	293	BP2793
putative exported protein (Pseudogene)	26212	9.44	26%	6	286	BP3045
superoxide dismutase	21274	6.05	23%	5	234	BP2761
transcription elongation factor	20650	4.93	46%	7	232	BP1117
ATP-dependent protease, ATPase subunit	96275	5.37	5%	5	203	BP1198
elongation factor Tu	42889	5.34	26%	9	195	BP0007
50S ribosomal protein L6	19160	9.59	22%	4	136	BP3631
50S ribosomal protein L19	13906	10.64	27%	3	130	BP0975
aspartate 1-decarboxylase precursor	13541	5.91	19%	2	124	BP1816
putative dnaK suppressor protein	17563	5.15	26%	3	114	BP3085
2-oxoglutarate dehydrogenase complex, E2	41799	5.36	5%	2	103	BP1125

NusG, a protein identified only in the second fraction, is a transcription anti-termination factor. It was identified with a score of 105, which is highly confident. BP0965, a peroxiredoxin, was identified in every fraction. The score increases from 91, through 138, to 524 and down to 165. Although mass spectrometry is not directly quantitative, this pattern is indicative of the shape of a protein peak, with the highest score being the tip.

Peptide coverage ranges from 17% to 62%. In a paradoxical example, the 'tig' protein (a bacterial 'trigger factor' chaperone) was identified in the fourth fraction with a score of 2033 and the first fraction with a score of 73. There were no identifications in intervening samples, possibly an indication of an unknown post-translational modification or washthrough due to overloading.

Identified proteins from the complete LC-MALDI analysis include very high and low molecular weight species. These species are often excluded from 2-D gels, either being retained in the 1st dimension focusing strip or passing too quickly through the gel in the second size separation step. CspA, a cold shock protein, is the smallest protein identified, with a molecular mass of 7344. It functions as an RNA chaperone, preventing RNA from forming secondary structures at low temperatures (between 10 and 37°C). In contrast, the largest protein found with a significant score was clpB, an ATP dependant chaperone protein with a molecular weight of 96275. Two 2-component response regulators: BP0991 and BP0572, were also identified with scores of 74 and 44 respectively. As the Bho mutant is deficient in a 2-component signalling pathway these identifications may be of biological relevance.

Analysis time, unfortunately, is the major drawback to analysis of peptides by LC/MALDI with conventional stationary phases. Taking one minute ion exchange fractions covering the entire LC run for all three dimensions would be the ideal way to analyse a sample, but if this were done, it is likely that four 384 well plates would be required, for a total of 1536 fractions. Any analysis that required more than a few minutes per sample requires a prohibitive amount of time to analyse, and led directly to the implementation of direct 384 well plate-to-target plate spotting, and parallel monolithic peptide LC.

As a more rapid analysis of protein contents, each fraction of a single 2D *Bordetella* separation was spotted across two MALDI target plates. 450 significant scores were observed, only 97 of which are different proteins. This highlights a limitation of this method of analysis, since the MS used to perform the analysis (the 4700 Proteomics Analyser, Applied Biosystems, Warrington, UK) is capable of identifying an average maximum of 6 proteins from a single spot, and high abundance proteins will often swamp lower abundance proteins. Potential exists, however, for an effective method utilizing much higher first and second dimension fraction resolution, especially for lower complexity samples.

A previous experiment by Vidakovics et al. [179] used two dimensional gel electrophoresis to assess the effect of iron starvation on *Bordetella pertussis*. While 900 distinct spots were observed, only 23 were selected for mass spectrometry analysis (based on their differential regulation) and identified. This study illustrates the power of the superior resolution characteristics of 2D gels: that downstream analysis time can be significantly decreased with the use of gel comparison. However, reproducibility of the gels is given as only 80-90%, and may lead to missed quantitative differences.

3.2.3.11.2. Analysis of a Eukaryotic Parasite by 2D intact LC with Peptide analysis by 3rd Dimension Monolithic LC/MS.

200µg of *Leishmania donovani* lysate were separated and fractions were collected into a 384 plate. After tryptic digestion, proteins were analysed by rapid direct injection peptide LC-MS.

Among the 1784 unique proteins confidently identified were several classes of proteins poorly represented on 2D gels, including membrane proteins, large proteins (>500kDa), small proteins (<8 kDa), very basic proteins (pI > 10) and acidic proteins (pI < 6). Identifications are biased towards low-to average molecular weight proteins (10-70kDa) with average pIs, but a large number of less average proteins are also identified (this generally reflects genome composition, as proteins with the highest expression levels are those at 30 kDa or below [180]).

An LC-based separation technology would be expected to be more tolerant to these types of proteins, especially to size and hydrophobicity, and this is supported by the findings. The ion exchange solvents were buffered to pH 8, therefore extremely basic proteins would be expected to be enriched in the first (washthrough) fraction. This is indeed the case, with significant scoring identifications for histone H2B (pI 11.41), Ribosomal proteins S26, L3 and L17 (pIs 11.73, 11.05 and 10.67 respectively) and a ubiquitin fusion protein (pI 9.97).

Membrane proteins are an important target for many proteomic experiments, and are considered difficult to separate using 2D gel electrophoresis [181, 182]. Analysis of proteins identified from a two dimensional LC separation of proteins from pentamidine resistant *Leishmania* using the TmPred software [183] (which estimates the number of transmembrane domains in a protein based on its sequence) indicate that roughly two thirds of the identified proteins are predicted to possess at least one transmembrane domain, and 17.37% are predicted to possess more than six (see Figure 35). This is a

surprisingly large number, almost double the predicted genomic number (9.05%) of proteins with seven or more transmembrane domains, suggesting that 2DLC of intact proteins is biased towards such proteins.

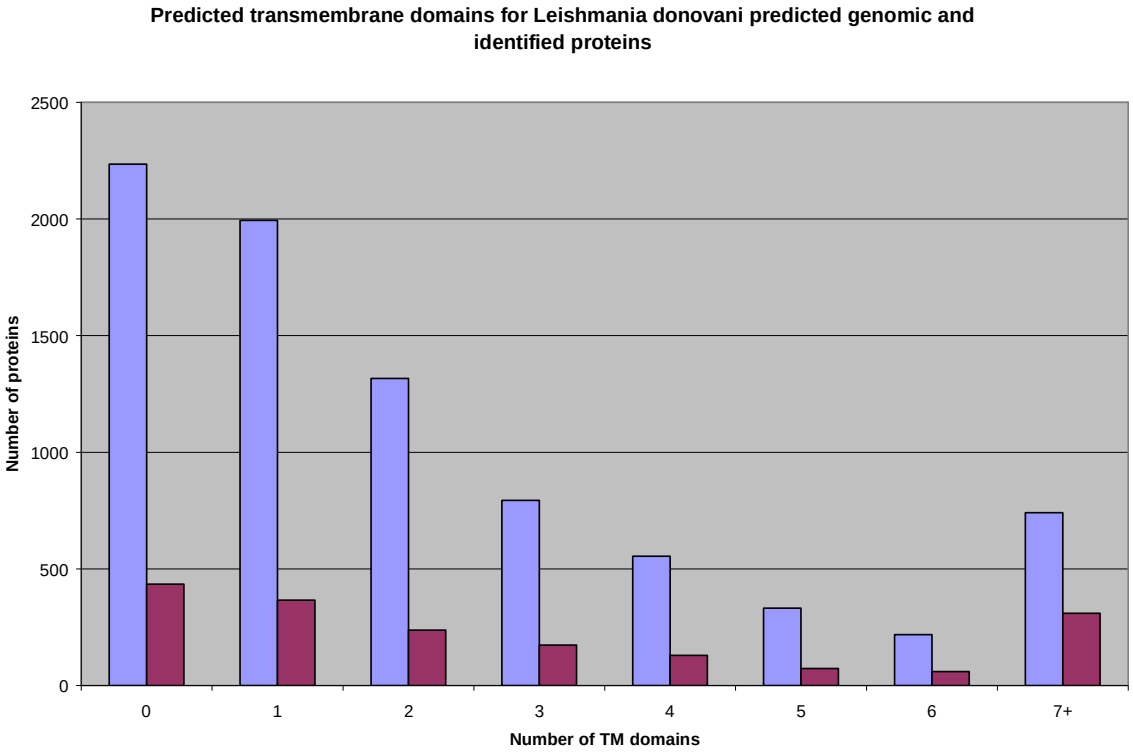


Figure 37: Predicted transmembrane domains in the identifications produced by a 2D intact protein separation by LC of *Leishmania donovani* in comparison (shown in magenta) with the total number of predicted transmembrane domains in the predicted *Leishmani donovani* proteome (shown in blue). A surprisingly large number of proteins possess seven or more transmembrane domains.

A quantitative proteomic analysis of *Leishmania* has been performed by Leifso et al. [184], and involved a comparison between the promastigote (a flagellated form, living in the midgut of the sandfly), and amastigote (which lives inside the cells of a mammalian host). Genomic data demonstrates that most genes (94%) are expressed in both life-cycle stages, but the two organisms are phenotypically, and therefore proteomically different. To analyse the proteomic differences, ICAT reagents and protocol were used to obtain 91 protein identifications of labeled proteins, allowing quantitative differences to be assessed between promastigotes and amastigotes. This is a surprisingly low number of identifications from a technique intended for entire proteome analyses. Although not directly comparable to the non-quantitative results described above, the number of identifications is two orders of magnitude lower in the ICAT experiment. This is likely to be a consequence of the labeling experiment, and subsequent purification of labeled peptides, as the reduction in protein identification confidence due to removal of all non-labelled peptides is substantial.

3.2.3.11.3. Abundance Estimation by Codon Adaptation Index

An important aspect of any proteomic technique is the dynamic range of a separation method. It is trivial to identify the very abundant proteins in any complex mixture, however, their high concentration can mask the signal of lower abundance (and consequently more interesting) proteins. In 1987, Sharp and Li developed the 'codon adaptation index (CAI)' – a means of estimating protein abundance. It was observed that throughout a genome, commonly utilised codons are found in the open reading frames of high abundance proteins. By creating a table of codon usage, and developing statistics about the relative abundance of a particular codon in a set of proteins of known high abundance, it is possible to produce a metric by which the abundance of a given ORF can be estimated. The CAI has been experimentally verified as being a good estimate of protein abundance in the yeast Saccharomyces cerevisiae [10]. In addition to the *Saccharomyces cerevisiae* data, a study of two *Streptomyces* species (*S. coelicolor* and *S. avermilitis*) by Wu et al. [185], and a more recent study by Bridges et al. [186] with Trypanosomes show good correlation of high CAI scoring proteins with experimentally highly abundant proteins verified by two dimensional gel electrophoresis. Of course the CAI does not take into account any environmental effects that may suppress or enhance the abundance of a particular protein, and more rigorous quantitative methods are needed for phenotypic analysis. Several CAI calculators are freely available. However, the best documented example: http://www.evolvingcode.net/codon/cai/cai.php#1 requires you to submit ORFs one at a time via a web interface, which is time consuming to manipulate for large numbers of queries; and the 'cai' software (part of EMBOSS) again only accepts a single input sequence, and requires it to be in trEMBL format, which is not easily generated automatically. Furthermore, few databases are downloadable in this format.

Therefore, as part of this project, a PERL script was written that takes a set of highly abundant ORFs in FASTA format, generates a codon usage table, a relative synonymous codon usage (RSCU) table, and a relative adaptedness (w) table. From these it generates CAI scores for proteins supplied in FASTA format. There is no limit to the number of FASTA sequences thus supplied, and the software can therefore be used to profile entire databases – which is the intended purpose of the tool. It was possible to generate a searchable index of CAIs for organisms of interest, allowing immediate estimates of protein abundance. CAI calculation software requires a group of proteins of very high abundance to serve as a benchmark of codon abundance. These are usually selected arbitrarily based on biochemical properties, e.g. glycolytic enzymes, chaperone proteins and ribosomal proteins. Since a differential gel electrophoresis has been performed in-

house on *Bordetella parapertussis*, an empirically determined set of proteins of known high abundance was used as the reference set of proteins. The CAI calculation produces a score of between 0 and 1 and is pseudo-logarithmic – thus a protein with a CAI of 0.5 will be roughly 10 times as abundant as a protein with a CAI of 0.4. In the case of *B. parapertussis*, CAIs for the genome were within the ranges of 0.89 to 0.22. CAIs of identified proteins from the LC/MALDI analysis of all 384 spots were in the range 0.30 to 0.89, suggesting that although the rarest proteins are not identified, the technique displays good dynamic range. For *L. donovani*, calculated CAIs for the entire genome fell between 0.15 and 0.93. Known high abundance proteins used in the training set fell into the range of 0.51 to 0.87. CAIs of identified proteins fell between 0.26 and 0.93, closely mirroring the results from *Bordetella*. Analysis of the CAIs of the *Bordetella parapertussis* proteins identified using shotgun-2DLC and GELC (summarized in Figure 36) shows that the shotgun-2DLC identified proteins with CAIs ranging from 0.40 to 0.89, and the GELC identified proteins with CAIs ranging from 0.41 to 0.87. Although these results are promising, and suggest that 2DLC of intact proteins detects less abundant proteins, multiple identical replicates are required to determine how well two dimensional LC of proteins matches up to other methods. But this is currently intractable due to the considerable duration of each 2DLC intact protein experiment.

3.2.4. Concluding Remarks

Two dimensional separation of intact proteins using monolithic columns can be performed routinely with excellent separation. The sheer volume of data obtained is challenging to analyse. For highly complex mixtures, such as entire bacterial lysates, additional separation may be required to enhance sensitivity to low abundance species. This technique can be performed with little user intervention, and, unlike ‘shotgun’ approaches, retains physicochemical data based on interaction between the proteins and the stationary phase. A standard sample preparation protocol has been developed, allowing resuspension of precipitated samples while also providing ion exchange compatibility. An effective digestion strategy for the samples has also been developed. Based on both UV chromatograms and preliminary MS data, separation of samples with high detergent content appears to have been substantially improved in the ion-exchange dimension with the addition of organic solvent to the buffers or reduction of overall detergent concentration. Low total protein loads are also possible with larger reversed-phase injection volumes, thus eliminating the need for, but maintaining compatibility with protein loads of greater than one milligram. An automated procedure has been developed

for spotting an entire 2-dimensional separation onto MALDI target plates for analysis by ToF/ToF MS. This technique is ideal for rapid analysis of a complete, high resolution separation, especially where UV based quantitation has been performed to provide candidate wells for rapid analysis. LC-MALDI is an effective method for analysing the complex protein fractions produced by 2D protein separation. It has a number of drawbacks, however. At a rate of roughly 24 samples per day, it takes roughly 16 days to separate an entire 384-well plate. The use of monolith columns for peptide separation would reduce this time significantly. However, depending on the complexity of the subsequent 4700ToF/ToF analysis, MS/MS of the samples can easily take 1 hour per fraction.

There are a number of other current methodologies for separation of proteins by multidimensional liquid chromatography, for example conventional WAX or SAX column followed by a C4 or C5 column, 2-dimensional capillary electrophoresis, size exclusion or anion exchange fractionation followed by conventional reversed phase, and size exclusion or chromatofocusing followed by PS-DVB monolithic separation. Weak anion exchange fractionation followed by conventional RP separation, coupled to both top-down and bottom up approaches has been reported. This technique has been applied several times, especially for top-down proteomics e.g. [47, 187]. It has the advantage that there is a large selection of column capacities, but their generally poor resolution limit their usefulness. Separation in 2-dimensional capillary electrophoresis is based on isoelectric focusing in the first dimension, followed by conventional capillary electrophoresis in the second dimension [42, 43]. This methodology provides extremely fast separations (17 seconds per second dimension analysis) by comparison to two dimensional separation of intact proteins by LC, but resolution suffers with a total spot capacity of approximately 380, and sample loading volumes, and hence total sample loads, are extremely limited. It is therefore ideal where sample is extremely limited, for example single-cell proteomics, but of less value for conventional proteomics experiments with sample volumes in the 20-200 µl range and the necessity for downstream identification of the proteins. Size exclusion chromatography is another method based on an easily verifiable physicochemical property. Coupled to a reversed phase second dimension (either conventional [188] or PS-DVB monoliths [45]), it provides an effective methodology for separating intact proteins. The major drawback to this technique is the poor resolution of the first dimension, providing an overall spot capacity of ~400. The chromatofocusing followed by reversed phase separation commercialized by Beckman Coulter (PF2D) is a popular mechanism for the separation of proteins in two dimensions, and has been applied to diverse biological samples and

questions. For example, McDonald et al. [189] described a comparison between this method and 2 dimensional gel electrophoresis for the purposes of cataloguing mitochondrial membrane proteins. Cussac et al. [39] applied the technique to the discovery of biomarkers in lymphoma cell lines. Finally, Sheng et al. [190] describe the use of the Beckman system for plasma proteome analysis. Although the system is fully automated and capable of producing pseudogel-style maps for crude quantitation, the technique is not without its drawbacks. A large amount of protein is required for separation (3 mg, 2.5 mg and 1-3 mg of lysate respectively) by 2DLC: four to ten times higher than the 200-400 µg required for the equivalent for a 2D gel or the technique described in this paper. Additionally, the variable (and often large) volumes of each chromatofocusing fraction may be problematic for downstream analysis by reversed phase column, unless fractions are dried down prior to second dimension analysis, with the attendant difficulties of resuspension.

The data presented demonstrates that, as a methodology for proteomics, intact protein LC is viable. Conditions for the separation are critical, such that sufficient precipitated protein can be both resuspended and remain compatible with LC, however the conditions are no more stringent than any other two dimensional separation technique. The most significant drawback to the uptake of 2D-intact protein LC as an analytical technique is run-time. Since a third dimension separation must be undertaken for each sample, post fractionation, the technique is rendered significantly longer to perform than a comparable digested protein 2DLC or most DIGE procedures. Two conclusions can be drawn from this: the first is that for lower complexity samples, such as immunoprecipitations, whose average number of proteins per fraction should be less than one, it should be possible to skip the third dimension separation and analyse the plate entirely by direct MALDI-spotting of each fraction on a target plate. Secondly, if it is possible to dramatically increase the throughput of the 3rd dimension separation, far more data can be obtained from the run – additional fractions can be taken, covering the entire ion exchange and reversed phase separations. Parallel monolithic separation is ideal for increasing the throughput of the system, and development of this system is described in 2.7. Online 2D separation of intact proteins provides an additional step of automation for researchers performing a large number of intact protein separations. Reduction in sample loss due to reinjection is also a benefit. Finally, this system allows the possibility of performing entirely automated top-down analysis of complex samples. With a variety of useful applications, the technique of intact protein separation using LC was felt to be of general interest to the proteomic community,

and consequently the description of the methodology and the supporting data has been submitted for publication.

4. Chapter 4 – Alternative First Dimension Separation Technologies for Intact Protein LC

4.1.1. Aims:

Ion exchange chromatography is by no means the only methodology capable of providing an effectively orthogonal separation to reversed phase chromatography for proteins. Two other methods with possibilities for specialised analyses are chromatofocussing and high-pH reversed phase separation. The aims of this section of the project were:

1). To evaluate separation quality of high-pH reversed phase separation.

2). To evaluate separation quality of the pISEP reagents for chromatofocussing especially gradient formation.

3). To use pISEP experimentally to separate a complex mixture.

4.1.2. Introduction

Ion exchange followed by reversed phase separation of proteins was described in chapter 3 as a general, high resolution methodology for proteomics. The superior resolving characteristics of the chromatography described herein are due to the stationary phases used in the separation, and therefore a modification to the buffer constituents to provide alternative orthogonal separation methodologies should maintain the separation quality and capabilities of the columns. Two additional orthogonal methodologies lend themselves to development as the first dimension of a multidimensional protein separation system for proteomics: high-pH reversed phase separation and chromatofocusing.

4.1.2.1. High-pH Reversed Phase Separation

Reversed phase separation can be enhanced with the use of a suitable ion pair in the buffer and is commonly done to improve the retention of polar solvents. As the sample is in equilibrium with the stationary and mobile phases, with the addition of an ion pairing reagent, a second equilibrium is reached between interaction of a polar sample with the ion pairing reagent, which acts to ‘mask’ functional groups on the polar sample to the hydrophobic stationary phase. Ion pairing reagents may either exhibit high pKas or low pKas, depending on the polarity of the solutes, i.e. cationic solutes are paired with high pKa ion pairs, and anionic solutes are paired with low pKa ion pairs. Ion pairing reagents

are commonly aprotic to provide ion pairing characteristics independent of the buffer conditions, however for MS compatibility, acidic ion pairs, such as TFA and FA are appropriate, and, indeed, used ubiquitously. The TFA and FA used in standard MS-compatible reversed phase separations enhance the hydrophobic interactions of anionic solutes. Substitution of an ion pair with a significantly different pKa changes the retention time of analytes, because the 'masking' effect of the ion pair will be applied to different functional groups [191]. By performing reversed phase chromatography on a sample using one ion pair, collecting fractions, and performing reversed phase chromatography on the fractions with a different ion pair, orthogonal separations can be performed.

It has been reported that predominantly-orthogonal multidimensional separation of peptides can be performed using standard octadecyl silica resins [192, 193] as well as PS-DVB monolithic resins [194] by applying an alkaline ion pairing reagent (such as ammonium formate or hydroxylamine) to standard RP buffers in the first dimension. Fractions generated from the first dimension separation are then separated by standard acidic ion pair reversed phase chromatography in the second dimension. The ion pair used in reversed phase separation has a significant effect on the retention time and capacity of the column, and using two diametrically opposed ion pairs, i.e. an alkaline followed by an acidic one provides a sufficiently orthogonal separation to effectively complexity-reduce very heterogeneous mixtures.

4.1.2.2. Chromatofocussing

Chromatofocussing is a separation methodology based on standard ion exchange column chemistries. Rather than employing chromatography based on the charge of sample constituents, with elution performed using strongly ionic solvents, chromatofocussing is based on pH gradients, either ascending or descending. Sample is loaded in a buffer at the starting pH and elution of sample components occurs as the pH of the eluent neutralises the overall charge of a protein – a position corresponding to its pI. Currently, the most widely used chromatofocussing technology is the Beckman ProteomeLab PF2D system (initially described in [195]). The first dimension chromatofocussing separation for this technique is based on the use of carrier ampholytes in the buffer and their application to a high resolution ion exchange column. There are several drawbacks to the system: firstly the ion exchange column used with the technique has a limited lifetime (reported as 4-6 runs under standard conditions [38] which could be due to continual protein deposition in the column), and secondly the pH range is significantly narrower than that of 2D gels (pH 8.5 – 4.0), although other sources provide polybuffers ampholytes with ranges up to 2-11.

Chromatofocussing was first described in 1978 by Sluyterman and Elgersma [196]. The paper describes two different theoretical methods of performing a separation on the basis of pH on an ion exchange column. The first, known as an 'external gradient', is generated by the mixing of two different strong buffers in different proportions to produce a gradient, in the same manner as standard ion exchange chromatography. The second, known as an 'internal gradient', is produced by the buffering effect of the column itself on an isocratically pumped mixture of ampholitic compounds [197]. Sluyterman and Wijdenes [198] implemented the internal method using a weak anion exchange column for the separation of proteins. Ampholite-based chromatofocussing works most effectively at low concentrations of buffer [196, 199] (the latter recommends a 1 mM solution), but gradient formation becomes erratic at a buffer concentration of lower than a 1/10 dilution of the initial solution [200].

An alternative, external chromatofocussing methodology, known as pISEP, has been developed by Tsonev et al. at CryoBioPhysica [201], based on a mixture of strong buffering reagents, which, due to their polarity are not commonly retained on reversed phase columns. Standard descending gradients run from a pH of 9.7 to 2.4. The pISEP technique is based on the technique described by Liu and Anderson [202, 203], which uses a cationic buffer mixture of Tris (pH 8.06), bis-tris propane (pH 6.8), and piperizine (pH 5.68), and an anionic buffer mixture of acetic acid (pH 4.76), lactic acid (3.81) and chloroacetic acids (2.87), with the addition of phosphoric acid buffered to either pH 7.2 (for the cationic buffer mixture) or 2.15 (for the anionic buffer mixture). The external gradient generated by these columns has a wider range and can produce shallower gradients than an ampholyte buffer system. The proprietary pISEP buffers must contain additional components than those described in Liu and Anderson's work as the range of the system is greater (pH 2.4-9.7 as compared to pH 3-8), but the actual composition is not available. Additionally, the column used throughout the work presented in [203] is a weak anion exchange column, but the pISEP technique works equally well on weak and strong anion and cation exchange columns. Ascending gradients can be applied to cation exchange columns by simply reversing the pISEP gradient from low pH to high pH. Generally, in the references described, the analysis of protein mixtures display similar chromatograms to those obtained using the pISEP reagents and detailed in this chapter.

4.1.3. Methods and Materials

Separations were performed using an UltiMate® 3000 system (Dionex, UK). pISEP buffers A and B were obtained from CryoBioPhysica (Maryland, USA). ddH_2O was obtained from a MilliQ (Watford, UK) system. Ammonium formate, hydroxylamine, Tris, HCl, acetonitrile and NaOH were obtained from Sigma (Dorset, UK). Other general reagents were obtained as described in section 2.1.

4.1.3.1. High-pH Reversed Phase Methods

Separations were performed with either the PepSwift (Dionex, UK) 500 µm x 50 mm monolithic PS-DVB or ProSwift (Dionex, UK) 4.6 mm x 50 mm monolithic PS-DVB columns. Aqueous buffer consisted of 2% v/v acetonitrile, 97.9% v/v ddH_2O and either ammonium formate at 25 mM, Tris-HCl pH 9 at 25 mM or hydroxylamine at 0.1% v/v. Organic buffer consisted of 80% acetonitrile v/v, 19.92% v/v ddH_2O and either ammonium formate at 25 mM, tris pH 9 at 25 mM or hydroxylamine at 0.08% v/v. Gradients ran from 2% v/v acetonitrile to 56% v/v acetonitrile in 20 minutes, followed by a step to 72% v/v acetonitrile, wash for 4 minutes, and equilibration at 2% v/v for 10 minutes.

4.1.3.2. pISEP Methods

Columns used were ProPac™ 1 and 2 mm x 250 mm strong anion exchange columns (for descending pH gradients) and ProPac™ 2 mm x 250 or 50 mm strong cation exchange columns (for ascending gradients). Additionally, technique validation experiments were performed using a GE Healthcare MonoQ 5 cm x 5 mm strong anion exchange column. ProPac™ columns were maintained at 37ºC in the flow manager. UV Absorbance was detected using an UltiMate® UV detector.

pISEP buffers were prepared from the proprietary solutions Reagent A and Reagent B. Buffer A (the acidic buffer) consisted of 1 mL Reagent B and 0.2 mL Reagent A in 100 mL ddH_2O, and was adjusted to pH 2.4 with the addition of 5 M HCl. Buffer B (the basic buffer) was prepared by adding 1 mL Reagent B to 100 mL ddH_2O followed by pH adjustment to 10.75 with 0.02 M NaOH.

4.1.3.2.1. Gradients Used for pISEP Chromatofocusing of Proteins

Both ascending (running from low pH to high pH) and descending (running from high pH to low pH) gradients can be produced by the reagents. Ascending gradients are applied to cation exchange columns, and descending gradients are applied to anion exchange columns. Details for the descending and ascending pISEP gradients applied to Mono Q

(GE Healthcare) and ProPac™ (Dionex, UK) columns are described in Table 3 to Table 8. In all cases, sample is loaded prior to a 10 minute equilibration step with the starting buffer.

Descending Gradients

Table 3 describes a descending pH gradient ranging from pH 9.7 to 2.4. Separations were performed on a GE Healthcare MonoQ 5/5 with a flow rate of 1 mL per minute, a 2 mm x 5 cm ProPac™ SAX column with a flow rate of 320 µl per minute, and a 2 mm x 25 cm ProPac™ SAX column with a flow rate of 300 µl per minute. Narrower range pH gradients can be applied to provide improved separation of a narrower range of pIs, analogous to narrow range isoelectric focussing strips. Gradient details for a descending gradient from pH 6.5 to pH 4, at a rate of change of 0.05 pH units per minute, are shown in Table 4.

Table 3: Gradient details for wide range descending pH gradients applied to GE Healthcare MonoQ 5mm x 5cm columns, ProPac™ SAX 2 mm x 5 cm columns and ProPac™ SAX 2 mm x 25 cm columns. The pH gradient ran from pH 9.7 to 2.4 at 0.1 pH units per minute. Flow rate was 1 mL/min for the MonoQ column, 320 µl/min for the ProPac™ 2 mm x 5 cm column and 300 µl/min for the ProPac™ 2 mm x 25 cm column.

Time (mins)	% Acidic Buffer (A)	% Basic Buffer (B)
0	12	88
10	12	88
68	90	10
83	100	0
90	100	0

Table 4: Gradient details for a narrow range descending pH gradient running from pH 6.5 to pH 4 at a rate of 0.05 pH units/min. Separation was applied to a GE Healthcare MonoQ 5 cm x 5 mm column at a flow rate of 1 mL/min.

Time (mins)	% Acidic Buffer (A)	% Basic Buffer (B)
0	55.5	44.5
10	55.5	44.5
54	79.6	20.4
70	86.2	13.8
74	86.2	13.8

Ascending gradients

Table 5 describes the gradient details for a wide range ascending pH gradient applied to a ProPac™ SCX column with dimensions of 2 mm x 50 mm. The gradient ran from pH 2.4 to pH 10.75 at a rate of change of 0.1 pH units per minute. Flow rate applied to the column was 320 µl per minute.

Table 5: Wide range ascending gradient applied to ProPac™ SCX columns of dimension 2 mm x 50 mm. The gradient runs from pH 2.4 (at 100% acidic buffer) to pH 10.75 (at 100% basic buffer). The flow rate applied to the column was 320 µl/min.

Time (mins)	% Acidic Buffer (A)	% Basic Buffer (B)
0	100	0
10	100	0
23.5	90	10
46.5	0	100
50	0	100

To allow automatic re-equilibration, allowing immediate loading of subsequent samples at the end of the run, a step to 100% acidic buffer with a 7 minute hold was employed in separations using 2 mm x 25 cm ProPac™ SCX columns (see Table 6). The gradient applied is a wide range ascending pH gradient running from pH 2.4 to 10.75. Flow rate was 300 µl per minute.

Table 6: Wide range ascending gradient applied to a 2 mm x 25 cm ProPac™ SCX column. The final two steps of the gradient are required to equilibrate the column at acidic conditions required for injection of subsequent samples.

Time (mins)	% Acidic Buffer (A)	% Basic Buffer (B)
0	100	0
10	100	0
25	90	10
91	0	100
98	0	100
99	100	0
105	100	0

Software generated ascending pH gradients

Software for more precise generation of gradients was supplied with the pISEP reagents. Evaluation of both the software generated and simple linear pH gradients under both

ascending and descending conditions was performed. Table 7 describes a software-generated wide-range descending pH gradient applied to a 2 mm x 250 mm ProPac™ SAX column. The pH range applied is the same as the linear gradient described in Table 3: pH 9.7 to pH 2.4 at a rate of pH change of 0.1 units per minute and a flow rate of 300 µl per minute. Table 8 describes a software-derived wide-range ascending pH gradient applied to a 2 mm x 250 mm ProPac™ SCX column. The pH range applied is the same as the linear gradient described in Table 6, pH 2.4 to pH 10.75 at a rate of pH change of 0.1 units per minute and a flow rate of 300 µl per minute.

Table 7: Software derived wide-ranged descending pH gradient applied to a ProPac™ 2mm x 250mm SAX column. The pH gradient ran from pH 9.7 (for 88.7% basic buffer) to pH 2.4 (for 100% acidic buffer). Flow rate was 300µl per minute.

Time (mins)	% Acidic Buffer (A)	% Basic Buffer (B)
0	11.3	88.7
10	11.3	88.7
17.3	23.91	76.09
24.6	35.42	64.58
31.9	44.5	55.5
39.2	52.1	47.9
46.5	59.9	40.1
53.8	68.43	31.57
61.1	76.56	23.44
68.4	82.9	17.1
75.7	88.4	11.6
83	100	0

Table 8: Software generated ascending gradient applied to 2mm x 250mm ProPac™ columns. The pH range runs from pH 2.4 (for 100% acidic buffer) to pH 10.75 (for 100% basic buffer). Rate of change of pH is 0.1 unit per minute, and a flow rate of 300µl per minute was applied to the column.

Time (mins)	% Acidic Buffer (A)	% Basic Buffer (B)
0	100	0
10	100	0
22.5	91.65	8.35
28.7	83.3	16.7
35.05	73.2	26.8
43.4	63.5	36.5

51.75	54.3	45.7
60.1	45.6	54.4
68.45	35.35	64.65
76.8	22.01	77.99
85.15	7.94	92.06
93.5	0	100

4.1.4. Results and Discussion

4.1.4.1. High-pH Reversed Phase Separation as a First Dimension

Initial development of high-pH reversed phase separation of proteins as a first dimension focussed on ammonium formate and hydroxylamine as high-pH ion pairs. Poor chromatography was observed with ammonium formate as an ion pair (see Figure 38), and the volatility of the buffer was such that run-to run reproducibility was poor. The UV absorbance of the hydroxylamine containing buffer at 214 and 280nm saturated the detector (see Figure 39), such that an estimate of separation quality could not be performed. Second dimension analyses of separations using hydroxylamine showed a large washthrough peak at the beginning of the gradient, but apart from a significant amount of absorbance in the washthrough fraction (likely to be the hydroxylamine still remaining in the fractions), showed little protein.

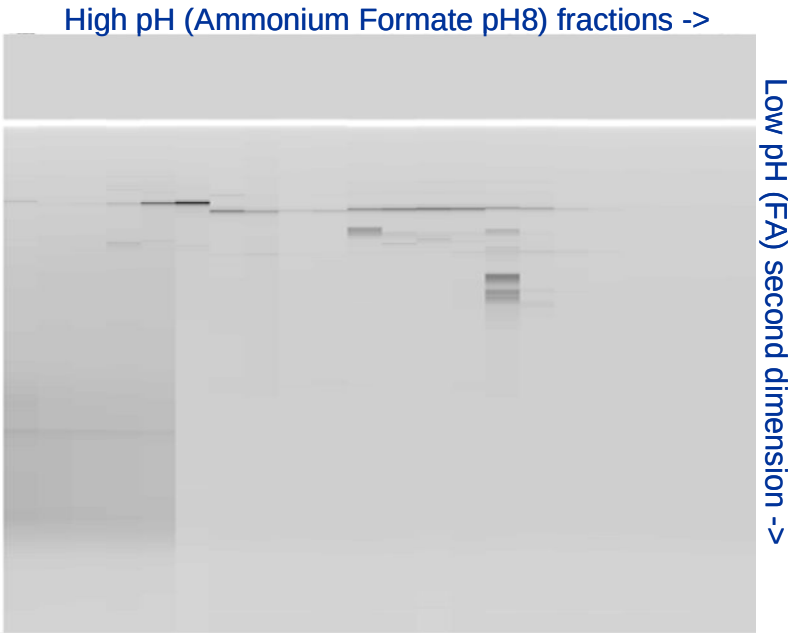


Figure 38: Reversed phase/reversed phase separation using ammonium formate as a high-pH ion pair in the first dimension (displayed horizontally) and formic acid as the low-pH ion pair in the second dimension (displayed vertically) for a separation of a 9 protein mix. Although a number of differential bands are observed between fractions, the separation appears only partially orthogonal, and the orthogonally separated species appear to smear across each ion exchange fraction.

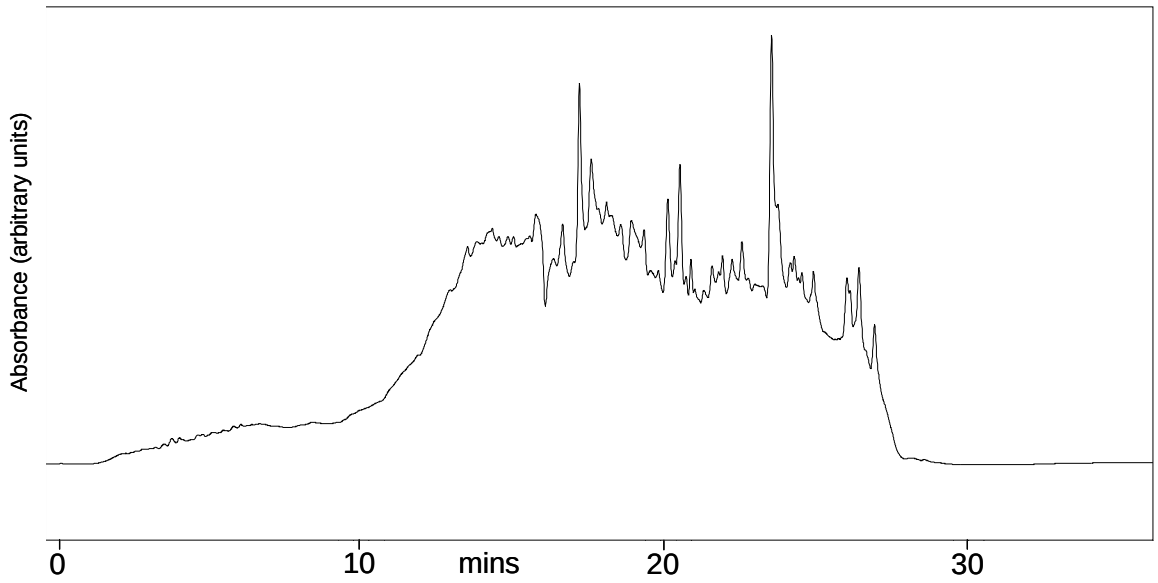


Figure 39: Hydroxylamine ion pair chromatogram with no sample injection at 280nm absorbance. Repeated analysis showed non-identical highly complex traces, again without the requirement of a sample.

The most successful high-pH ion pairing reagent was Tris-HCl, buffered to pH 9. The plot produced can be summarized as streaky in the second dimension (see Figure 40 and Figure 41). This is likely to be due to the use of formic acid as an ion pair, and may be improved significantly with the use of TFA or HFBA. Further optimisation of gradients in the first dimension to expand the area of high complexity between fractions 25 and 45 (corresponding to ~8 to 15 minutes) is likely to produce a significant improvement over the separation shown.

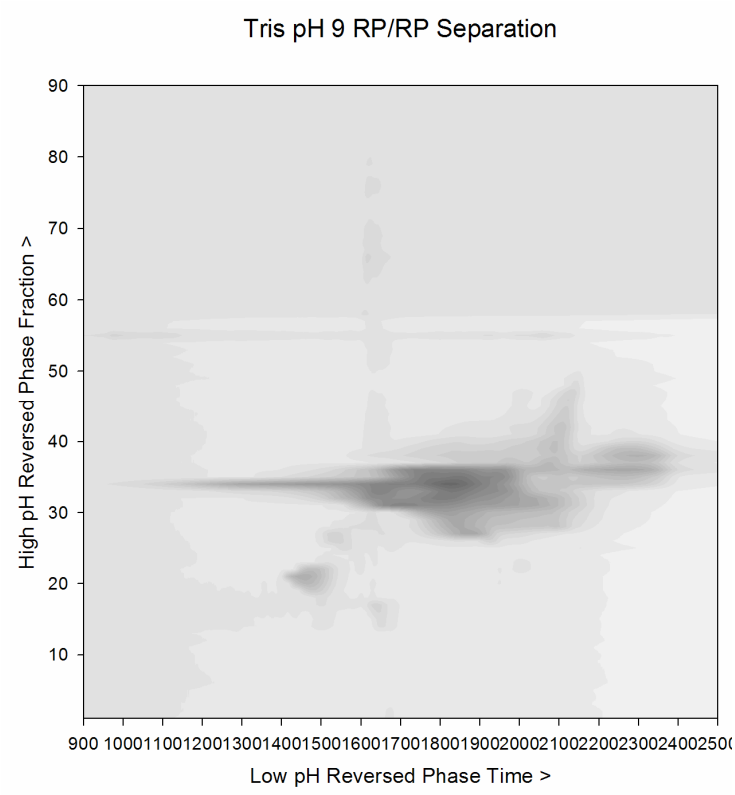


Figure 40: Pseudogel, plotting high pH first dimension reversed phase fraction number against low pH reversed phase time of RP/RP separation using tris-HCl at pH 9 as the first dimension ion pair and formic acid as the second dimension ion pair.

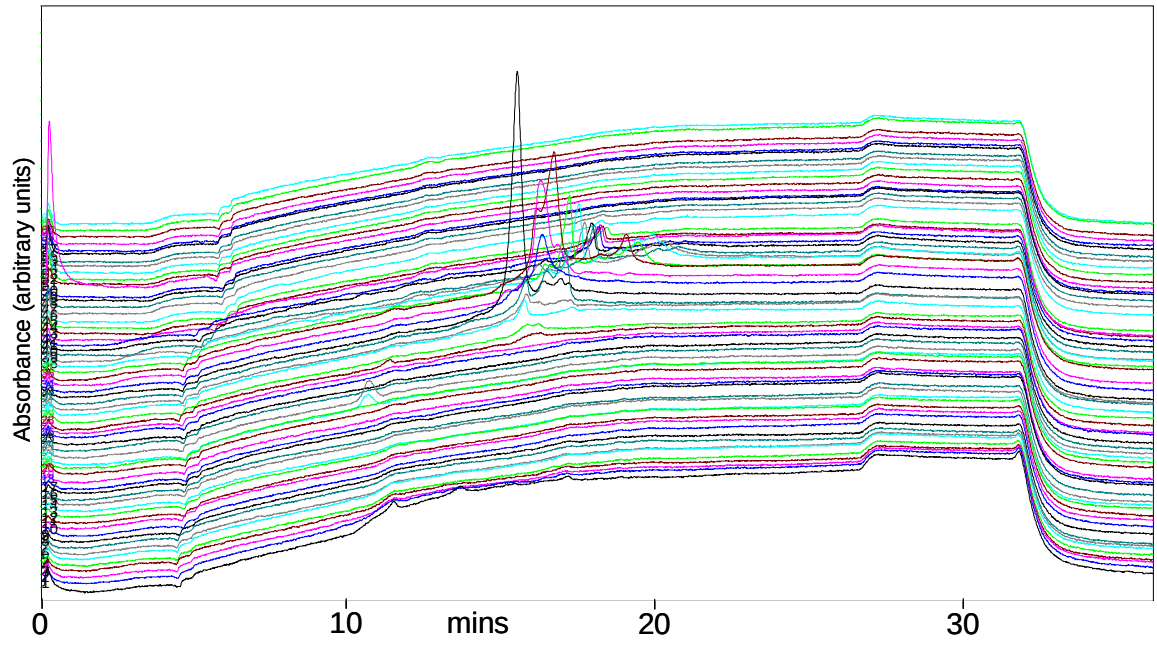


Figure 41: Stackplot showing the same separation as in Figure 40. Note that in the complex area of the chromatogram, separation is very good. The broad peak widths in the low-pH reversed phase dimension are a result of the use of formic acid as an ion pair.

Ammonium formate and hydroxylamine are commonly used ion pairing reagents in general chromatography, but are more rarely used for proteins and peptides (a few examples are described in [204-206] and [194]). Tris is more rarely described in

chromatography. Garcia [207] reviews a variety of alternative ion pairing reagents used in protein and peptide chromatography, but excludes most of the salts on the basis that they are incompatible with MS, which is not an issue for the first dimension separation of a multidimensional technique, because salts are excluded in the second dimension. An early paper on the use of ammonium formate as an ion pair in reversed phase HPLC for peptides [206] describes the use of 0.5 M ammonium formate, pH 6.3 for an additive to the mobile phase for the separation of lipid associated peptides. It is described as providing better separation characteristics than both gel filtration and ion exchange chromatography, to the extent that it can be used as a purification step for the peptides. Toll et al. [194] discuss the use of PS-DVB monolithic columns with hydroxylamine (pH 11.9) as an ion pairing reagent for the separation of peptides. The difficulty of analysis due to saturation of the UV detector is bypassed by the use of direct coupling to ESI-MS, but separation is shown to be orthogonal and of high resolution. It suggests that the application of triethylamine to protein separation is worth further consideration, despite the lack of success of early attempts. A paper describes Tris-HCl as a mobile phase additive [208]. In it an isocratic flow consisting of 20% v/v acetonitrile and 80% v/v 0.5 M tris pH 7.2 for separation of bilirubin and biliverdin is described. This is done presumably to maintain the separation at physiological conditions and prevent precipitation of bile salts.

Time constraints precluded further investigation of this technique, but it remains an intriguing possibility, especially for samples where the lack of charge separation would be seen as a benefit such as in plasma samples where albumin can be restricted to a single area of the separation space rather than spread throughout due to its heterogeneity. The major anticipated difficulty would be a requirement for the careful evaporation of organic solvents from fractions collected in the first dimension, prior to their reinjection into the second column.

4.1.5. pISEP as a First Dimension

Evaluation of the technique initially focussed on replication of results from standards to confirm function of the general method. For this reason a MonoQ 5/5 strong anion exchange column from GE Healthcare was used for comparison with the chromatograms and methods provided in the pISEP documentation (see section 4.1.3.2.1).

4.1.5.1. Reproduction of Reference Chromatograms

We performed an initial ion exchange salt gradient separation of Sigma ovalbumin to confirm the performance of the column we were to use for the evaluation, and this matched well with the reference chromatogram (Figure 42).

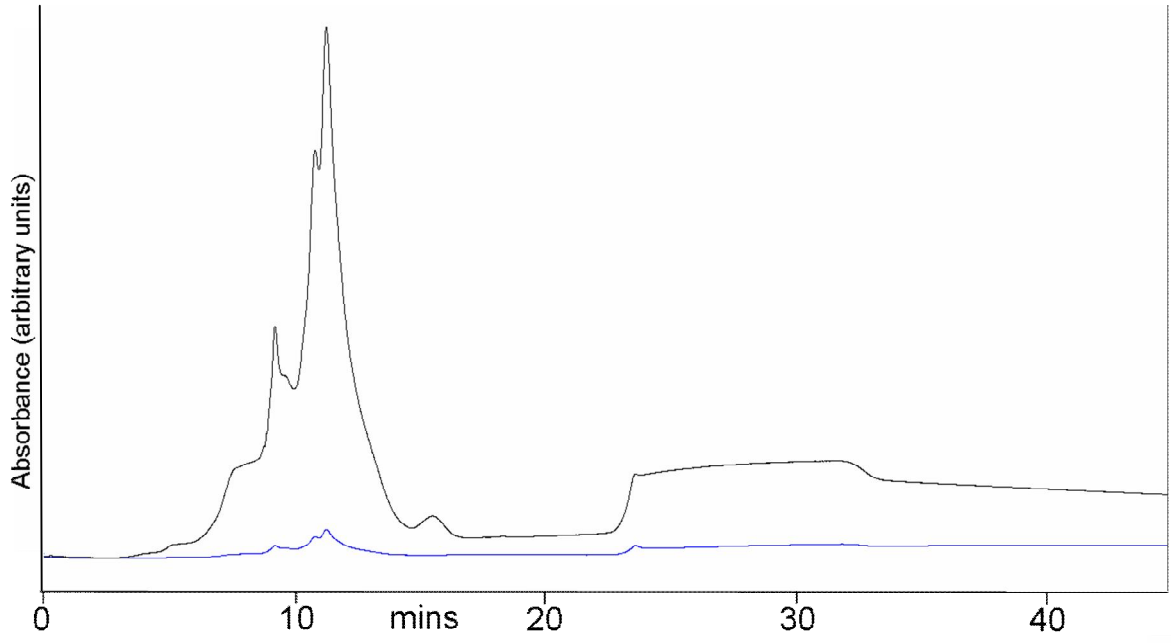


Figure 42: Ion exchange separation of 200µg of ovalbumin on MonoQ 5/5. The gradient used was 0 – 500mM NaCl in 20 minutes, with a step to 900mM NaCl for 10 mins, and equilibration at 0mM NaCl for 15 minutes. Both low and high ionic strength buffers contained 10mM phosphate pH 8. The black line denotes 214nm absorbance, the blue line denotes 280nm absorbance.

Implementation of the buffer system for the pISEP was straightforward, but somewhat time-consuming as the buffers are only stable for a short time. Solvents therefore have to be made freshly and adjusted to the appropriate pH for each run (a maximum of 200 mL of buffer was prepared prior to each analysis).

The first pISEP samples consisted of 200 µg of ovalbumin separated using full-range (pH 9.7-2.4) (Figure 43) and narrow range (pH 6.5-3.65) (Figure 44) descending gradients. Ovalbumin was chosen due to its multiple phosphorylation and glycosylation isoforms, which should provide a limited number of distinct peaks. The pISEP chromatogram produced was roughly the same resolution as the anion exchange chromatogram of the same sample. Additional peaks were present in the reference chromatogram from the pISEP documentation, but this may be explained by the difference in source of ovalbumin (CalBioChem in the case of the reference chromatogram, and Sigma in the case of the chromatogram described here). The system was tested at a wavelength of 215 nm, the

commonly used wavelength for proteins, there is considerable drift in the baseline at this wavelength due to the absorbance of the buffers (see Figure 43). Moving to longer wavelengths, where detection of proteins is less effective, gave less, although still substantial background.

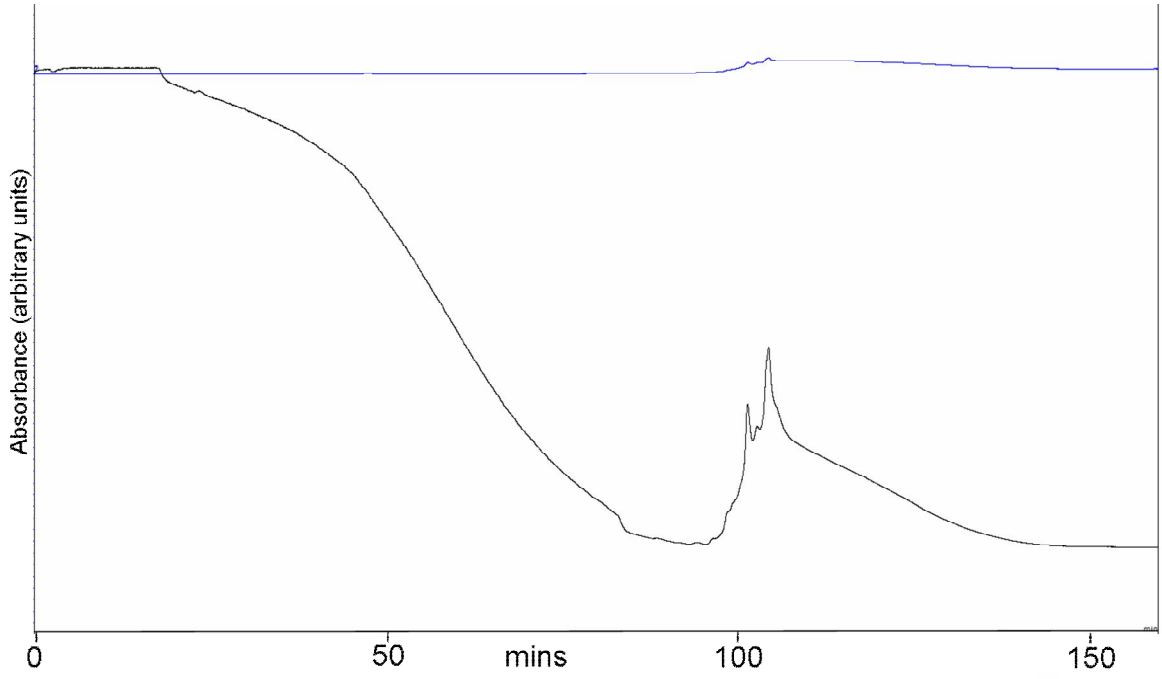


Figure 43: Wide pH range descending pISEP separation from pH 9.7 to pH 2.4 of Sigma chicken egg ovalbumin. UV absorbance data was collected at 215 nm (black) and 280 nm (blue). Note the dip in absorbance during the run: this is a consequence of the pISEP buffers causing a change in absorbance at 215nm over the course of the gradient.

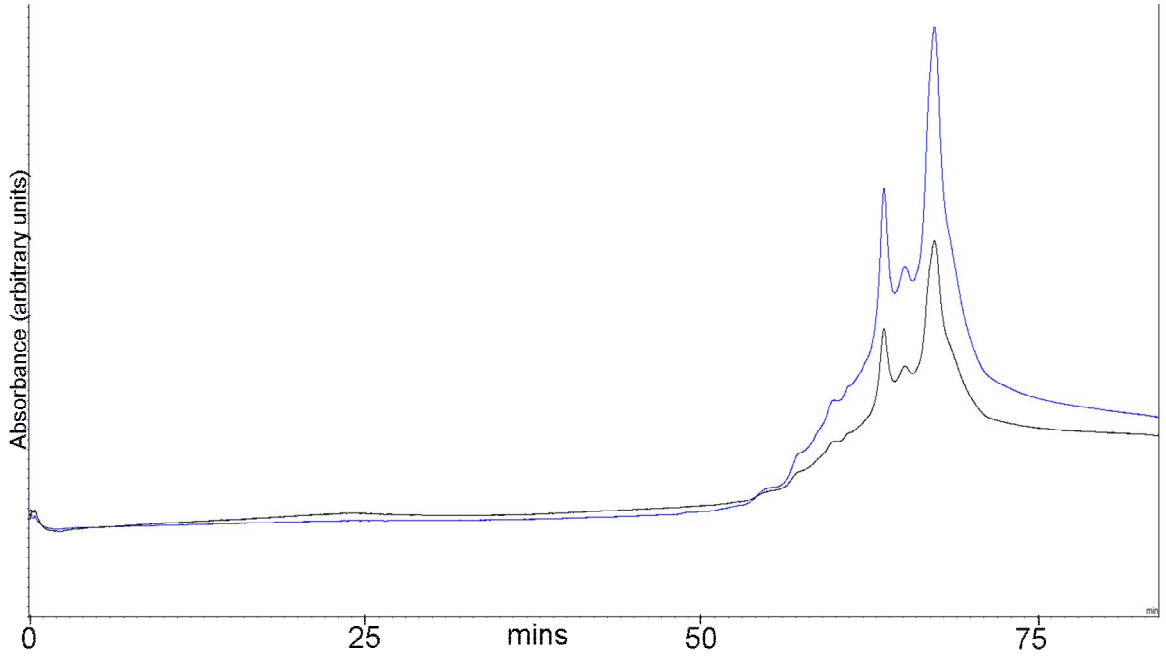


Figure 44: Narrow range (pH 6.5 - 3.65) separation of Sigma ovalbumin. The black line denotes absorbance at 256nm, the blue line denotes absorbance at 280nm. Note the improved separation on the easily generated narrow pH gradient, and lower background at these detection wavelengths, rather than the more commonly used 214 and 280nm.

A chromatogram of Sigma ovalbumin was obtained from CryoBioPhysica that showed a similar, although slightly higher resolution separation (the reference chromatogram was performed on a 4 mm x 50 mm Dionex ProPac™ WAX column, run at a significantly higher flow rate (1 mL/min). The pISEP system was evaluated using two different chemistries of ProPac™ column, with two different column dimensions: 2 mm x 50 mm (SAX and SCX) and 2 mm x 250 mm (SAX and SCX). Separation of a test sample (ovalbumin) was markedly improved by using a ProPac™ column over a MonoQ column (Figure 45).

Based on volumetric calculations of column bed sizes, initial runs were rather time-consuming – however these calculations did not allow for the fact that pellicular resin columns such as ProPacs (because of the lack of pores in the packing material) have a lower dead volume. After discussion with the column manufacturers, separations on the 25 cm ProPac columns were performed using 1 column volume = 300 μL (rather than the initially calculated ~1 mL). The apparent increase in peak resolution observable in the chromatograms from the 2 x 50 mm columns as compared to the 2 x 250 mm columns is probably due to the additional column volumes applied in the short column separations (Figure 46), which effectively increase the length of the gradient.

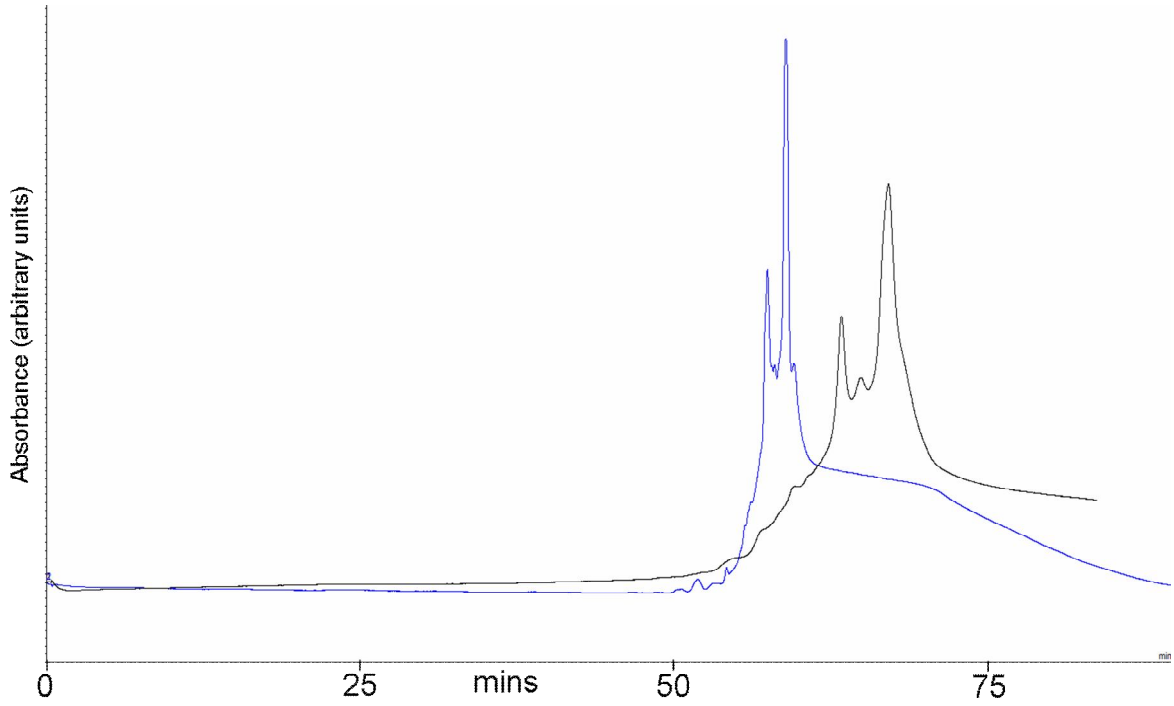


Figure 45: Improved resolution provided by ProPac™ columns. A wide range (pH 9.7 to 2.4) descending pISEP separation of ovalbumin was performed on a MonoQ 5/5 (GE Healthcare, black line) and a 25cm x 2mm ProPac™ SAX column (blue line). Note that the shouldering of the final peak in the MonoQ chromatogram has resolved into two peaks in the ProPac™ chromatogram.

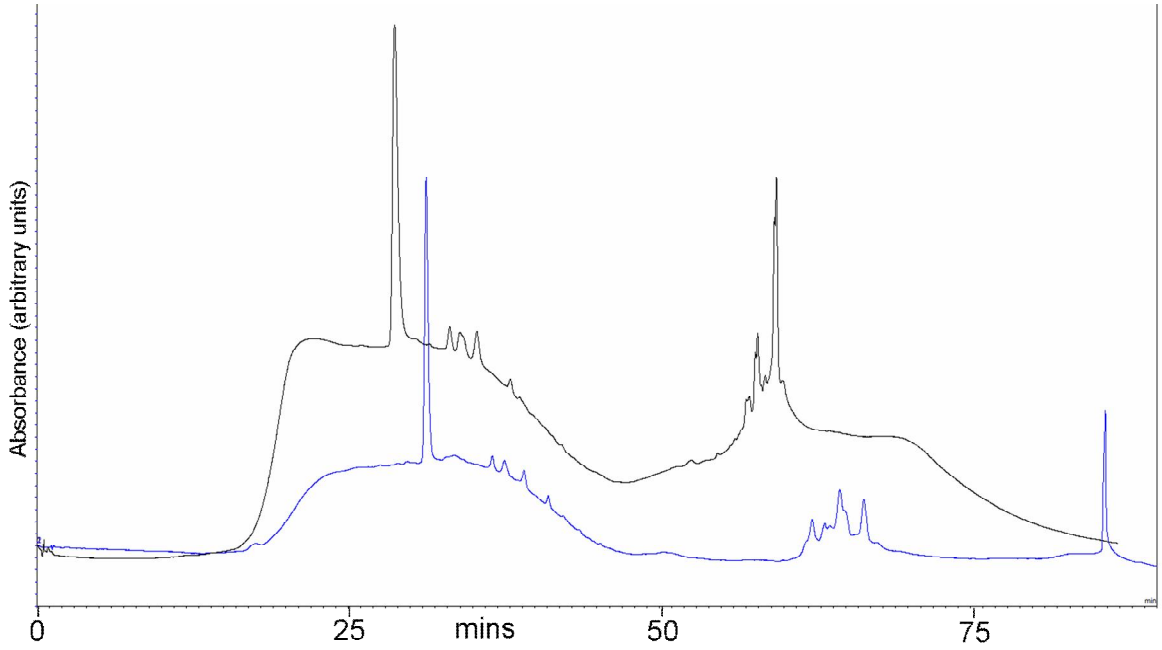


Figure 46: Comparison of 2mm x 50mm ProPac™ SAX column and 2mm x 250mm ProPac™ SAX column using a descending wide range pISEP gradient (pH 9.7-2.4). The fine peak resolution of the 2mm x 50mm separation (black line) seems slightly improved over the 2mm x 250mm separation, but baseline separation is better on the 2mm x 250mm (blue line).

4.1.5.2. Separation of Moderate Complexity Samples

Further analyses evaluated the separation performance for a moderately complex sample (a 9 protein mix, each injection of which contained 8 micromoles of BSA, ovalbumin, myoglobin, β lactoglobulin, α lactalbumin, cytochrome c, α-casein, carbonic anhydrase and lysozyme) using anion exchange with salt gradient (Figure 47), and narrow pH range (Figure 48) and wide pH range (Figure 49) pISEP separations. 1 minute fractions were collected from the wide range 9 protein mix separation and the pH of each fraction was recorded, demonstrating that the pH gradient was well formed (Figure 50).

Separation of the 9 protein mix using the wide range pH gradient was good – several more proteins at the extremes of the pI range were retained on the column than in the narrow gradient or standard anion exchange, and separation was improved in comparison to the ion exchange and narrow pH range gradients.

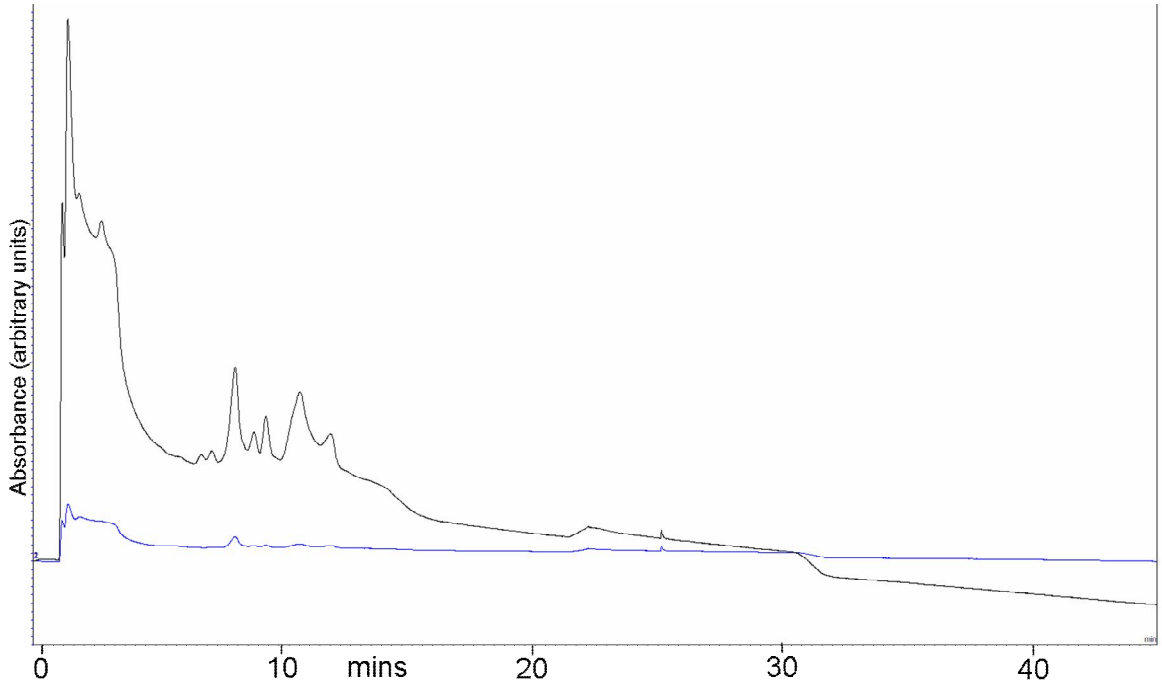


Figure 47: Anion exchange chromatogram of a 9 protein mix separated on a MonoQ 5/5 eluting using a 0-500 mM NaCl gradient. Note the large washthrough at the beginning of the run resulting from one or more of the proteins failing to bind to the column. The black line denotes absorbance at 214 nm, the blue line denotes absorbance at 280 nm.

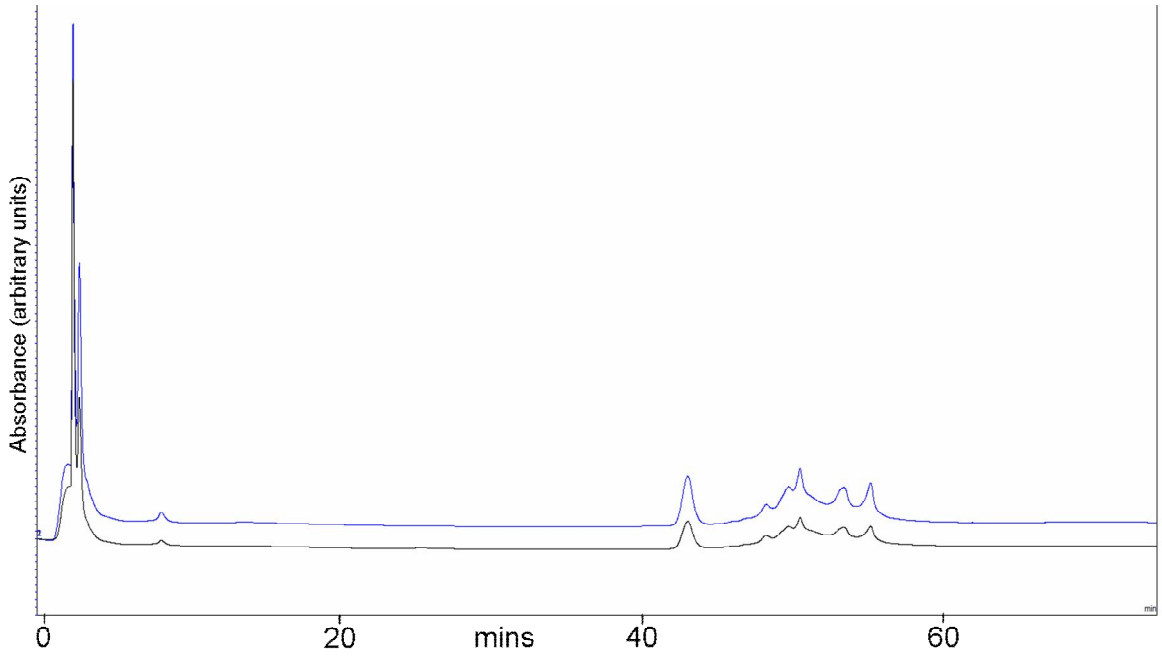


Figure 48: Narrow range (pH 6.5-3.65) descending pISEP separation of a 9 protein mix on a MonoQ 5/5. Note the sharp washthrough peaks at the beginning of the run - a consequence of proteins with pIs greater than 6.5 failing to adhere to the column. The black line denotes absorbance at 256 nm, the blue line denotes absorbance at 280 nm.

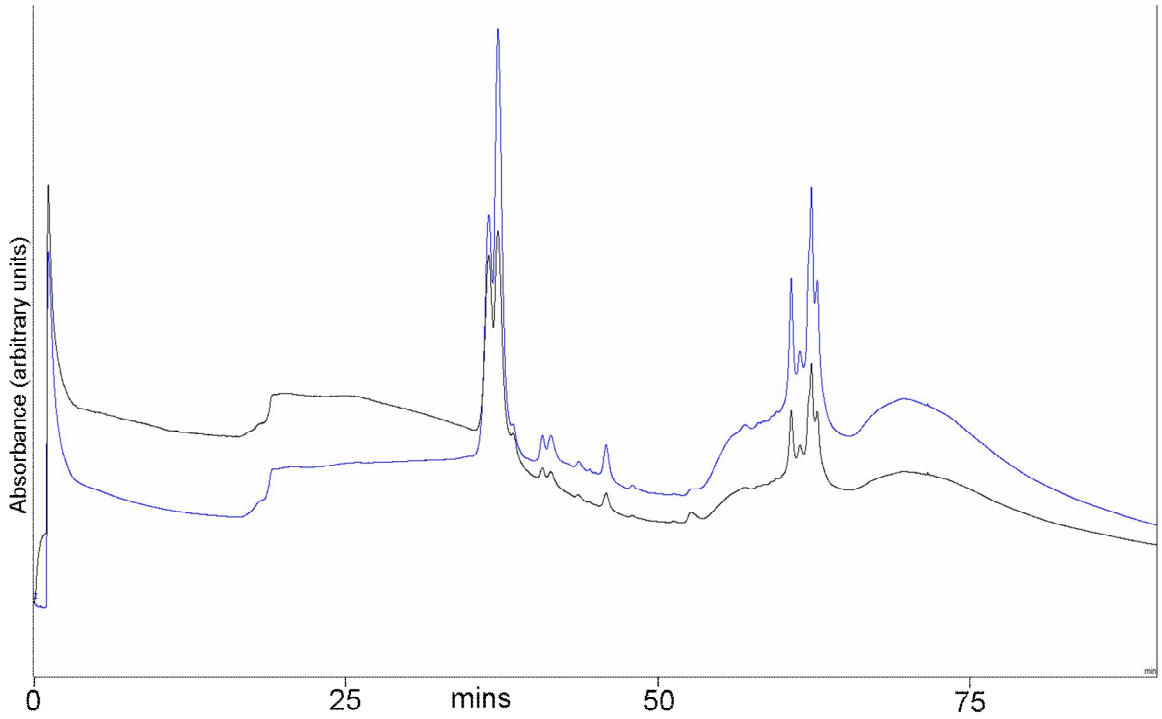


Figure 49: Wide range (pH 9.7-2.4) descending pISEP separation of a 9 protein mix. Note the high resolution of the separation in general, and the small peak at the beginning compared to the narrow pH range, which may be a consequence of a water dip, or may be a small amount of protein washthrough. The black line denotes absorbance at 256 nm, the blue line denotes absorbance at 280 nm.

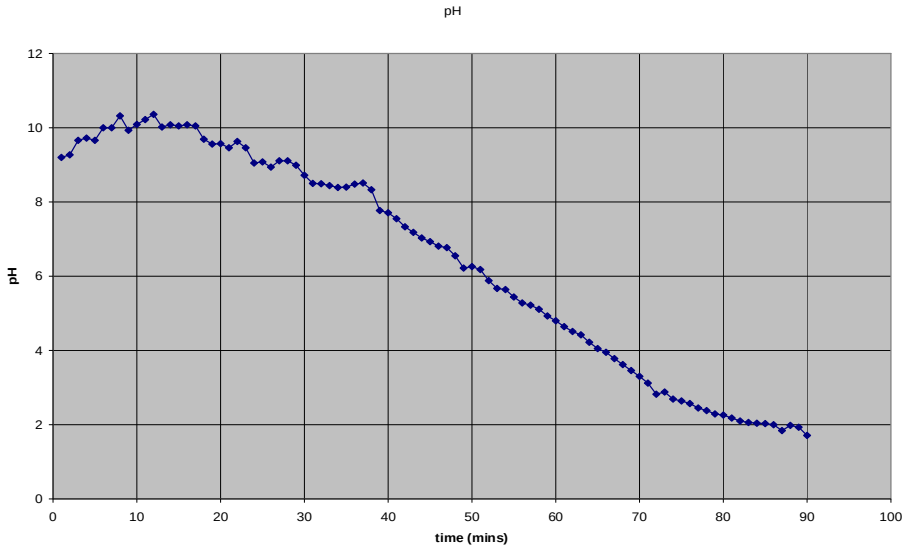


Figure 50: pH gradient from wide range (pH 9.7-2.4) descending pISEP separation. pH change averages around the predicted 1 pH unit per 10 column volumes. Note that the first ten minutes of the separation are isocratic, and pH fluctuations during this period may be due to mixing effects from the injection.

Evaluation of an ascending pH gradient was performed using the 9 protein mix on both 2 mm x 50 mm and 2 mm x 250 mm cation exchange columns. With both columns, peak widths were similar to that for descending gradients, with only a slight loss of resolution (Figure 52). However, the separation of the sample was significantly worse than for the descending gradients, with far fewer species being observed. For comparison, a standard 20 minute gradient cation exchange was performed using the same column (Figure 51) which demonstrated that the column is working effectively. This may be due to selective protein precipitation at the loading pH.

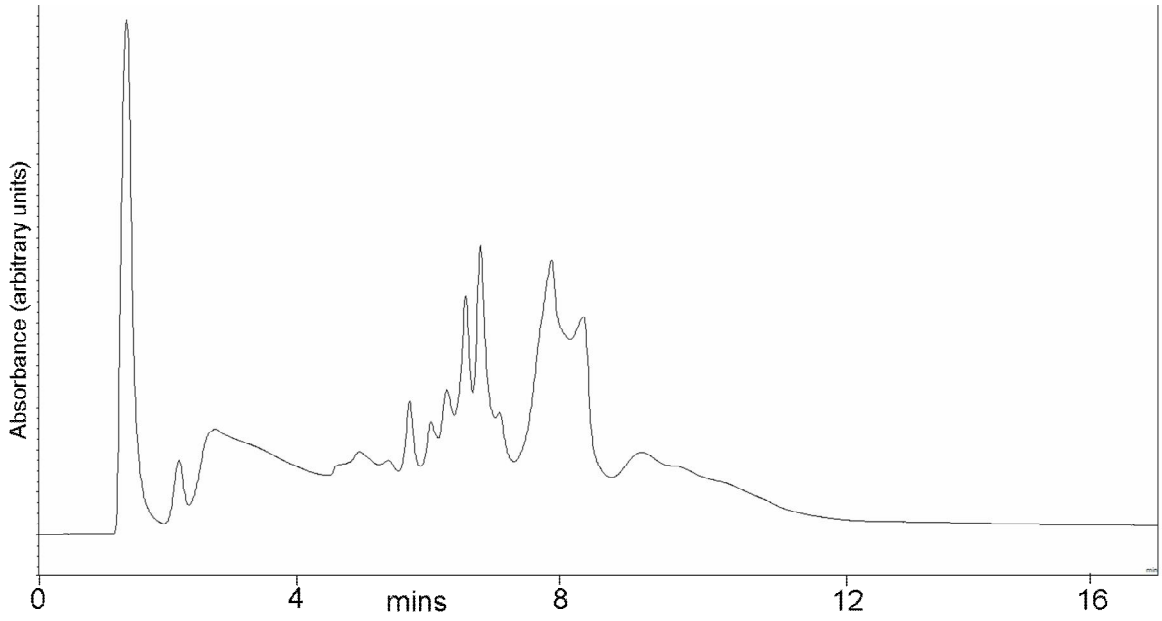


Figure 51: Complex section from a 20 minute gradient cation exchange chromatogram of 9 protein mix using a 20 minute gradient from 0 mM to 500 mM NaCl. This provides an ion exchange chromatogram that may be used for comparison with the ascending pISEP separation shown in Figure 53. The black line shows UV absorbance at 214 nm.

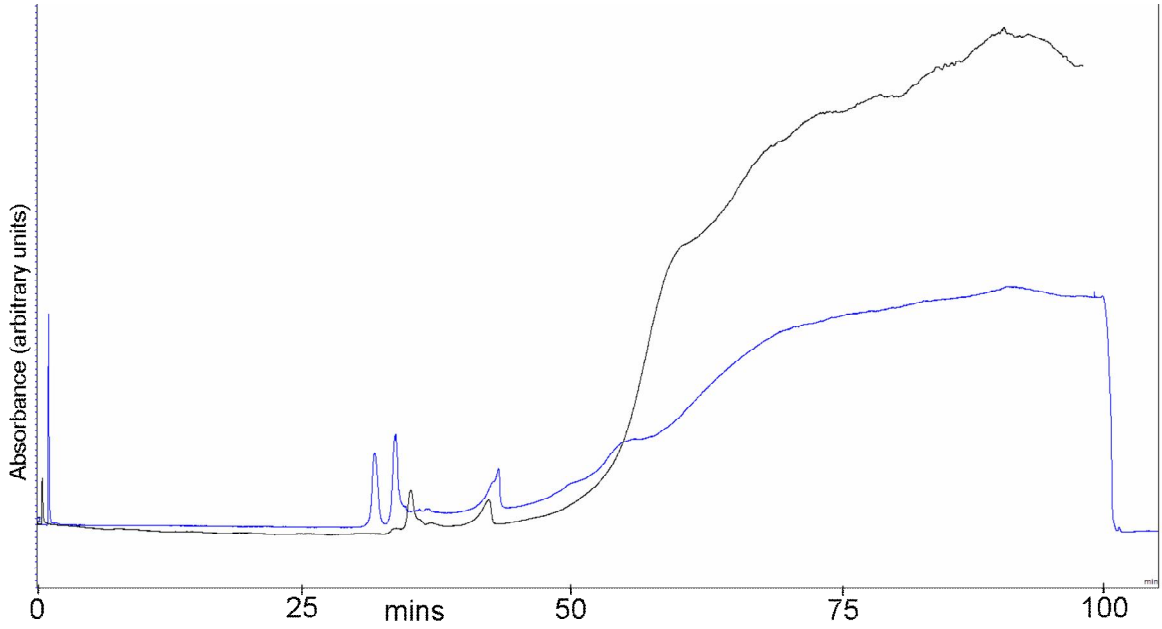


Figure 52: Ascending pISEP gradient applied a 9 protein mix separated on 2mm x 50mm ProPac™ (blue line) and 2mm x 250mm ProPac™ (black) columns. The chromatogram from the 2mm x 250mm ProPac™ column demonstrates slightly lower resolution, and fewer components are visible. Note the step at the end of the 2x50mm run. This is from a re-equilibration step to allow immediate injection of the next sample, and shows a rapid change in UV absorbance, suggesting rapid gradient changes are possible with this system.

4.1.5.3. Comparison of Linear Gradients with Software-Generated Gradients

Some concern was expressed that the gradients used in most of the separations (which were taken from page 4 of 'Instructions pISEP Buffers Lot 1.pdf' document supplied with the buffer kit) were not generating the most effective pH gradients. To compare pH gradient generation by the pISEP software with the reference gradients from the kit, an electrochemical detector with pH probe was connected, post UV detection. Both ascending (Figure 53) and descending (Figure 54) gradients were evaluated both for chromatography and pH gradient formation (Figure 55 and Figure 56).

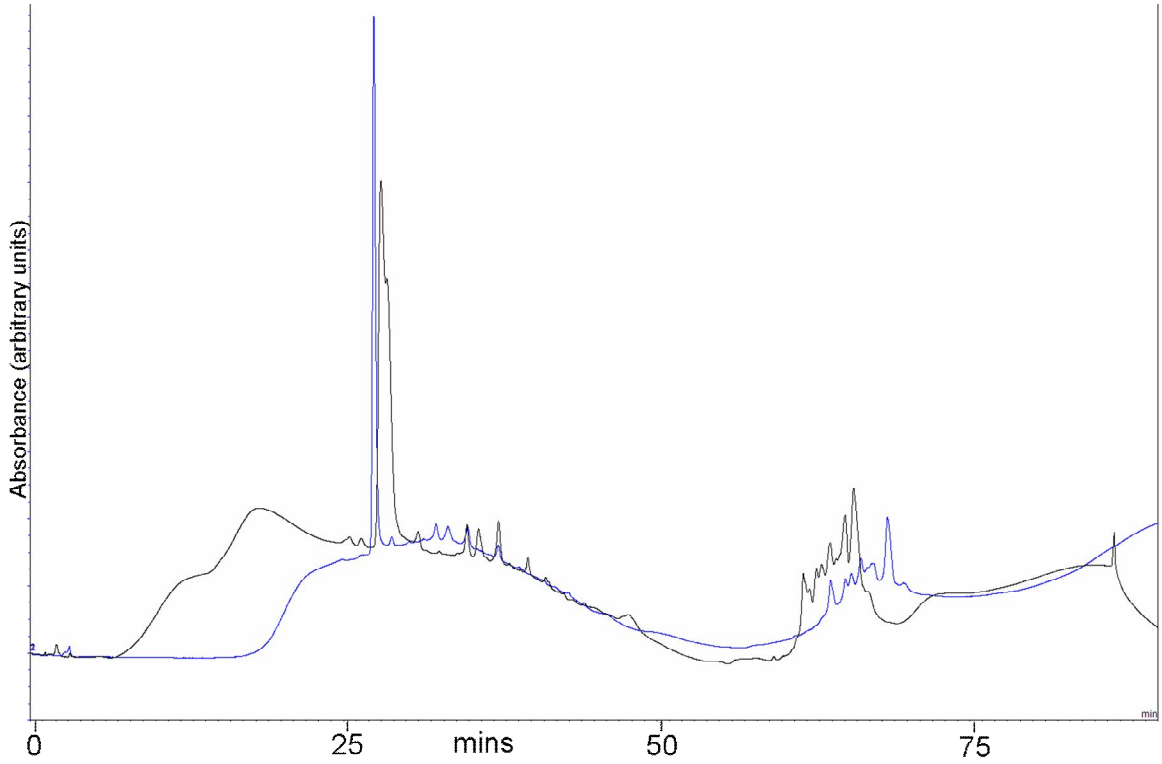


Figure 53: Comparison chromatograms from separations of a 9 protein mixture separated using descending pISEP gradients from pH 9.7 to 2.4 generated by the pISEP software (blue) and taken from the pISEP documentation (black). The slightly less sharp separation seen for the software generated gradient is within normal run-to-run variation.

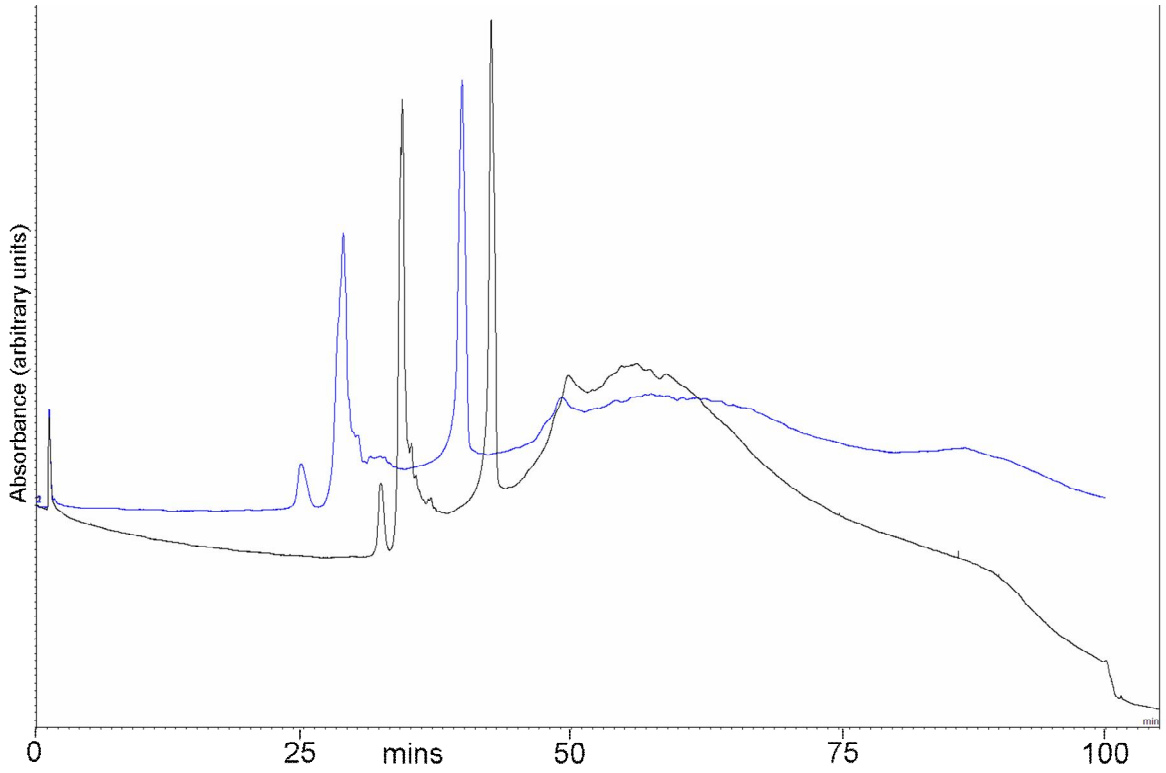


Figure 54: Comparison of documentation-derived (black) and software-derived (blue) linear ascending pISEP gradients from pH 2.4 to 10.75 applied to the separation of a 9 protein mix. Again, peaks seem slightly sharper in the gradient from the documentation.

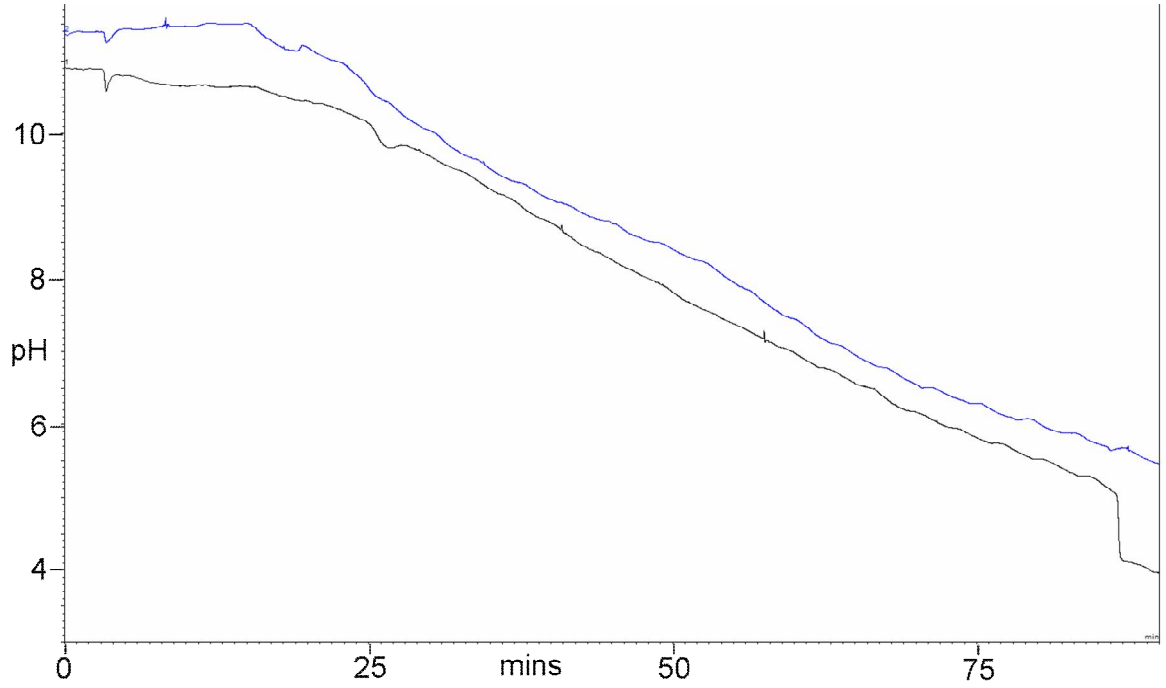


Figure 55: Comparison of pH readings obtained from descending pISEP gradient programs from pH 9.7 to pH 2.4. The gradient program based on the supplied documentation is shown in black and the gradient program generated using the supplied pISEP software is shown in blue. pH measurement was performed using an electrochemical detector and shows that the gradient seems to form well for both programs.

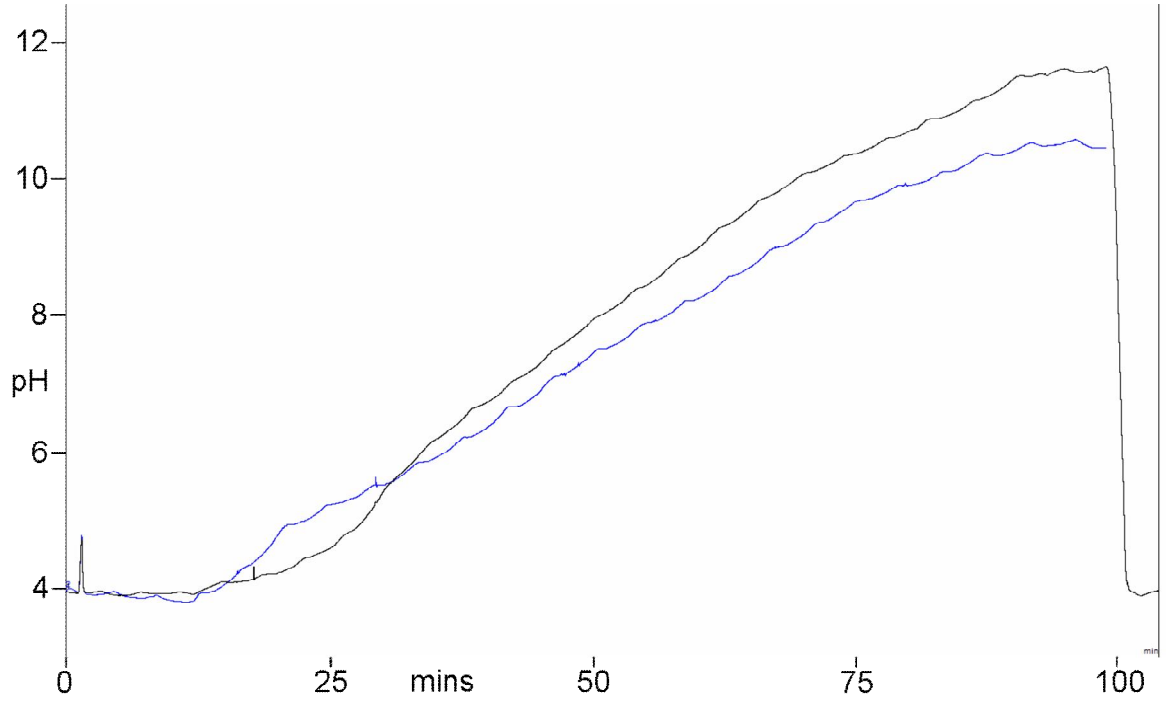


Figure 56: Comparison of pH readings obtained from ascending pISEP pH gradients from pH 2.4 to 10.75. pH readings from the gradient program obtained from the pISEP documentation are shown in black and and readings from the gradient program obtained from the software are shown in blue. Both separations and gradients are roughly comparable. Note the significant drop in pH during the final five minutes of the run in the gradient obtained from the documentation: this is a reequilibration step to prepare the column for another ascending gradient. It demonstrates that the switch from high pH to low pH and equilibration is rapid and effective.

Dissolution of sample in appropriate buffer (basic for descending separations, acidic for ascending separations) was compared with unbuffered sample preparation. The use of basic buffer eliminated any washthrough from the 9 protein mix for descending separations, but dissolution of sample in acidic buffer was still associated with some washthrough for ascending separations (data not shown), and may have resulted in some precipitation of the sample, resulting in less complex chromatograms.

4.1.5.4. Application of pISEP to Complex Lysates

A complex sample (300 µg of *Bordetella parapertussis* lysate suspended in 10 mM pH 8.0 phosphate buffer) was analysed using a wide pH range gradient. A significant washthrough peak is present – this may be due to protein washthrough or due to other components in the sample (e.g. CHAPS carried over from lysis, or the use of the phosphate buffer). Further evaluation in a second dimension separation by reversed phase will be needed to clarify this, although it is likely to be the latter. The separating power of the technique distributed

peaks throughout the pH range of the methodology, and a complex chromatogram was produced.

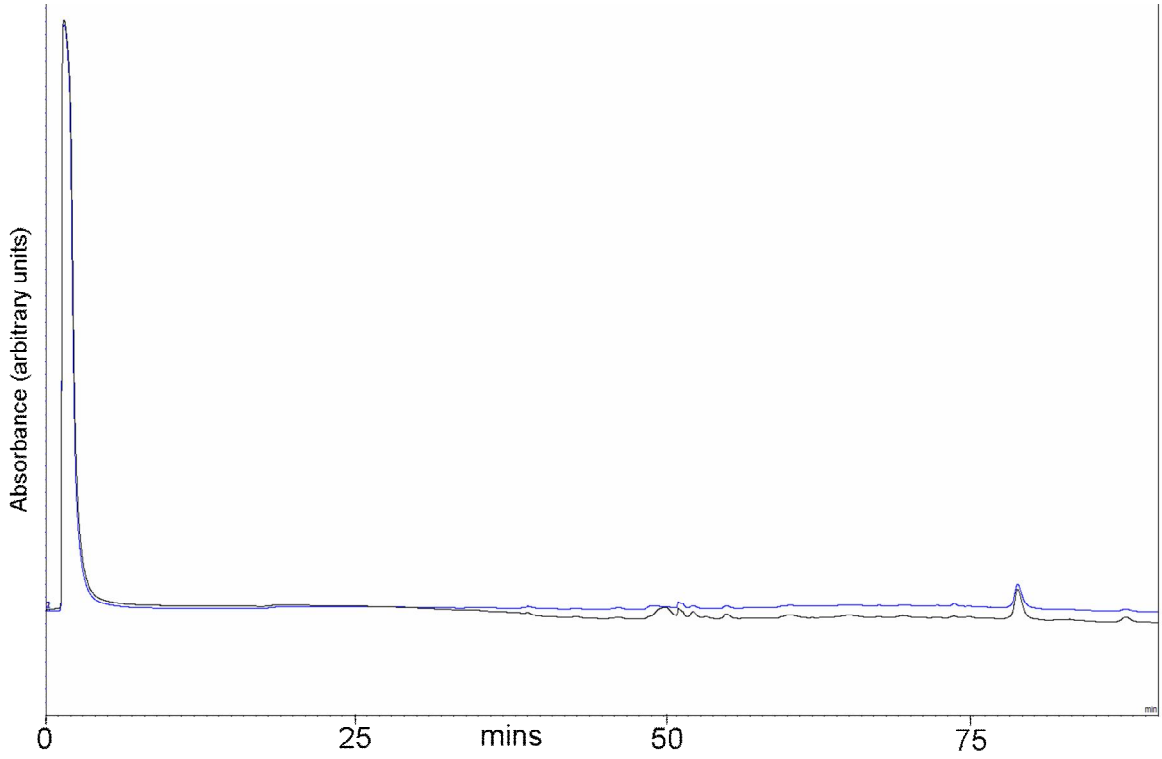


Figure 57: Wide range descending (pH 9.7-2.4) pISEP separation of 300 µg of *Bordetella parapertussis* lysate using a MonoQ 5 cm x 5 mm SAX column. Note the large washthrough peak at the beginning of the run, as well as the complex chromatogram starting at ~40 minutes (see Figure 58). The black line denotes absorbance at 256 nm, the blue line denotes absorbance at 280 nm.

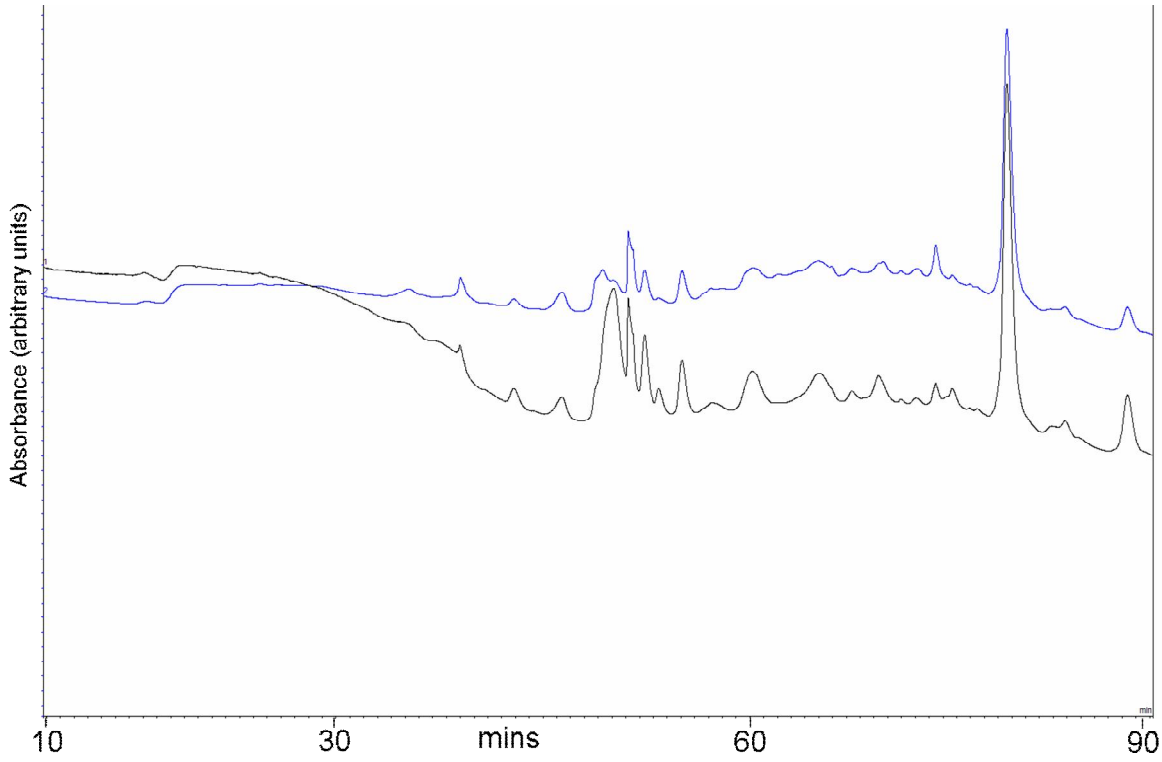


Figure 58: Magnification of the complex region from Figure 57. The black line denotes absorbance at 256 nm, the blue line denotes absorbance at 280 nm.

4.1.5.5. Application of pISEP to Two Dimensional Protein Separations

For use as a proteomics technique, pISEP is ideal for use as the first dimension of a two-dimensional separation method to provide an additional dimension of separation. For this reason, and to evaluate the orthogonality of the technique in comparison to ion exchange chromatography and performance and compatibility in a 2D approach, one minute fractions of wide pH range ascending and descending gradients were collected, and each fraction was then separated by 500 µm PS-DVB Monolith (Dionex) using a standard 20 minute gradient from 15% v/v to 70% v/v acetonitrile. Heatmaps of time vs. absorbance were generated and are shown in Figure 59 and Figure 60. Figure 59 confirms the good separation observed in the first dimension chromatogram for the descending gradient. Several of the proteins appear to have streaked across a number of fractionsin the pISEP dimension, albeit at a low absorbance level; this could be due to non-specific adsorption, the column loadings, or quite possibly a consequence of post-translational modification (in our experience the post-translational modifications of the proteins in our test set are not significantly separated in the second dimension reversed phase runs). In contrast, Figure 60 shows the first dimension chromatogram for the ascending gradient. Some peaks are well isolated, but others are extremely streaky, and in general, there seem to be fewer protein

species eluted in the ascending gradient. As discussed before, this may be a consequence of a pH drop resulting in precipitation of some proteins.

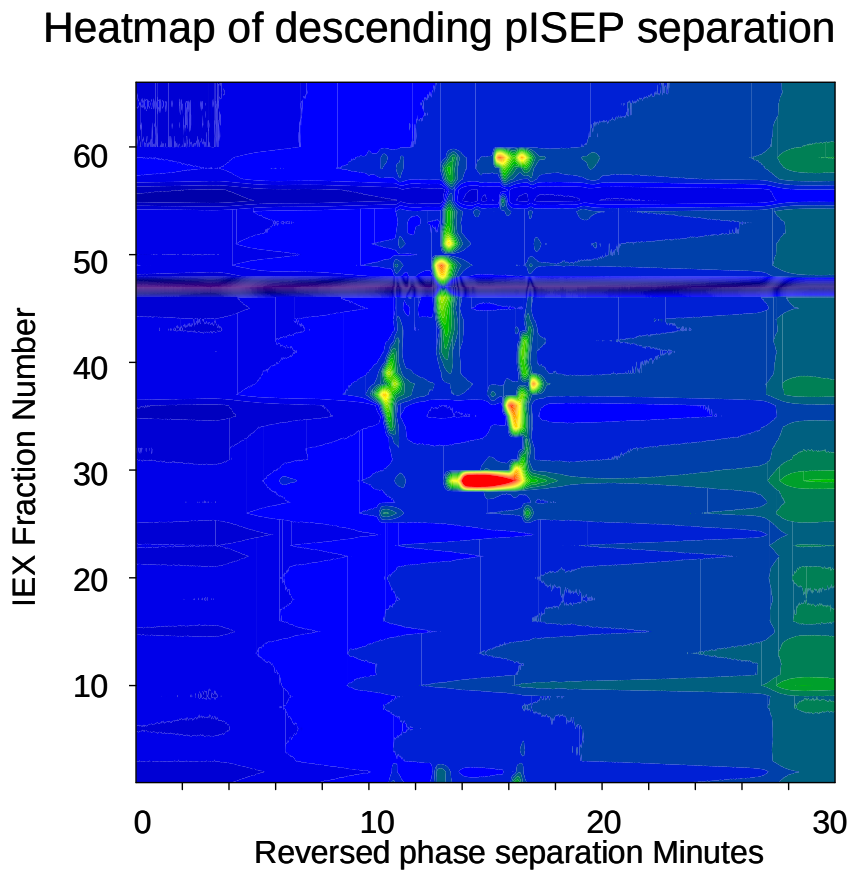


Figure 59: Heatmap of descending gradient pISEP fractions from pH 9.7 to pH 2.4 (y-axis) separated by 500 µm PS-DVB monolithic reversed phase chromatography (x axis). UV intensity is plotted on a colour axis from blue (low intensity) to red (high intensity). Note that species are well separated in both dimensions, and although some streaking is observed in the pH dimension, this could be due to post-translational processing.

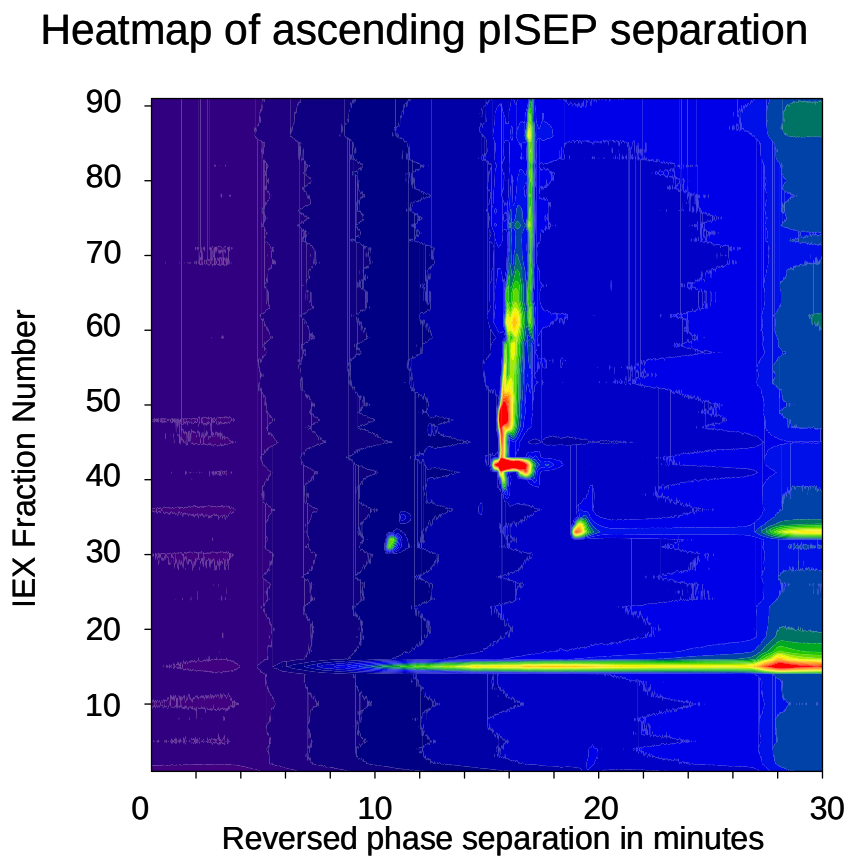


Figure 60: Heatmap of ascending gradient pISEP fractions from pH 2.4 to pH 10.75 (y-axis) separated by 500 x 5cm µm PS-DVB monolith reversed phase chromatography (x axis). UV intensity is plotted on a colour axis from blue (low intensity) to red (high intensity). Significant streaking is observed, but the heatmap suggests that separation is slightly better than that implied by the original chromatogram.

A complete analysis of a complex lysate was performed using two dimensional separation methodology with a descending pISEP gradient as the first dimension and monolithic reversed phase separation as the second dimension. Fractions were digested and analysised using conventional peptide LC/MS as described in section 2.6.3. 1172 non-redundant protein identifications with a significant Mascot score were obtained from 1mg of *Leishmania donovani* lysate. This result is comparable to the two dimensional separation performed using ion exchange followed by reversed phase, using identical lysate, described in section 3.2.3.11.2. Analysis of the results obtained displays significant complimentarity to the ion exchange/reversed phase technique, with only 754 identifications matching in both analyses (the ten highest scoring proteins unique to pISEP are shown in Table 9, and those for IE/RP are shown in Table 10). Scores for the pISEP analysis were generally lower than those for the ion exchange/reversed phase separation, despite the larger initial sample amount (see Table 11). However, some very significantly scoring proteins (scores of 200 and over) were found only in the pISEP fractions, again suggesting the complementarity of the techniques, although based on predicted pIs, there is

no significant difference in the type of proteins identified (see Figure 61). This is not entirely unexpected, as proteins not bound to the column in the ion exchange separation should be collected in the washthrough and observed in those fractions.

Table 9: Top ten identifications only found with the pISEP technique.

fraction	protein name	protein MW	protein PI	protein sequence coverage	number of peptides matching	MOWSE score	accession number
dpep0734	ATPase beta subunit, putative Leishmania major chr 25	56285	5.14	15%	6	308	LmjF25.1180
dpep0734	ATPase beta subunit, putative Leishmania major chr 25	56313	5.14	15%	6	306	LmjF25.1170
dpep0733	ATPase alpha subunit Leishmania major chr 5	62509	9.73	13%	8	276	LmjF05.0500
dpep0732	hexokinase, putative Leishmania major chr 21	51686	8.72	15%	5	211	LmjF21.0240
dpep0732	hexokinase, putative Leishmania major chr 21	51714	8.72	15%	5	211	LmjF21.0250
dpep0734	hypothetical protein, conserved Leishmania major chr 1	39244	6.01	10%	5	172	LmjF01.0480
dpep0731	proteasome alpha 2 subunit, putative Leishmania major chr 21	25081	5.65	16%	3	152	LmjF21.1700
dpep0683	RPA1 replication factor A, 51kDa subunit, putative Leishmania major chr 28	52399	6.19	8%	3	142	LmjF28.1820
dpep0731	proteasome alpha 7 subunit, putative Leishmania major chr 11	27769	5.85	8%	2	142	LmjF11.0240
dpep0707	methionine aminopeptidase 2, putative Leishmania major chr 21	51199	6.51	18%	4	129	LmjF21.0840
dpep0563	macrophage migration inhibitory factor-like protein Leishmania major chr 33	12504	5.89	22%	2	119	LmjF33.1740

Table 10: Top ten identifications found only with the IEX/RP technique.

fraction	protein name	protein MW	protein PI	protein sequence coverage	number of peptides matching	MOWSE score	accession number
dpep0013	delta-1-pyrroline-5-carboxylate dehydrogenase, putative Leishmania major chr 3	61860	8.25	20%	16	515	LmjF03.0200
dpep0275	glutamyl-tRNA synthetase, putative Leishmania major chr 30	67576	8.65	17%	8	307	LmjF30.3240
dpep0374	aspartate aminotransferase, putative Leishmania major chr 35	45960	6.57	16%	9	229	LmjF35.0820
dpep0033	hypothetical protein, conserved Leishmania major chr 28	12978	6.75	53%	7	220	LmjF28.1920
dpep0107	hypothetical protein, conserved Leishmania major chr 19	43382	6.86	10%	4	208	LmjF19.1550
dpep0109	aminopeptidase, putative Leishmania major chr 33	55721	6.67	22%	8	206	LmjF33.2570
dpep0154	peptide methionine sulfoxide reductase-like Leishmania major chr 7	19753	6.79	35%	7	175	LmjF07.1140
dpep0156	aldehyde reductase, putative Leishmania major chr 31	31601	7.06	20%	5	169	LmjF31.2880
dpep0107	LACK1 activated protein kinase c receptor (LACK) Leishmania major chr 28	34378	6.05	12%	3	163	LmjF28.2740
dpep0181	elongation factor, putative Leishmania major chr 25	70283	6.14	6%	4	161	LmjF25.0130
dpep0084	peptidyl-dipeptidase, putative Leishmania major chr 2	76519	5.8	10%	8	161	LmjF02.0740

Table 11: Top ten hits from the ion exchange/reversed phase separation matching proteins identified in the pISEP separation. Most of the matching proteins are the higher scoring pISEP proteins, but a few are identified in the pISEP/RP separation with barely significant scores.

protein name	protein MW	protein PI	protein sequence coverage	number of peptides matching	Score IEX	Score pISEP	accession number
dipeptidyl-peptidase III, putative Leishmania major chr 5	75920	5.3	28%	35	***617***	***43***	LmjF05.0960
chaperonin Hsp60, mitochondrial precursor Leishmania major chr 36	59281	5.33	26%	12	***516***	***233***	LmjF36.2030
chaperonin Hsp60, mitochondrial precursor Leishmania major chr 36	60142	5.38	23%	12	***514***	***336***	LmjF36.2020
eukaryotic initiation factor 4a, putative Leishmania major chr 1	45299	5.83	30%	10	***460***	***155***	LmjF01.0770

aspartate carbamoyltransferase, putative Leishmania major chr 16	35317	6.21	24%	8	***406***	***205***	LmjF16.0540
EF2-1 elongation factor 2 Leishmania major chr 36	94073	5.77	12%	16	***392***	***256***	LmjF36.0180
nucleoside diphosphate kinase b Leishmania major chr 32	16629	7.71	49%	13	***313***	***362***	LmjF32.2950
cysteine synthase, putative Leishmania major chr 36	35427	6.55	46%	11	***312***	***256***	LmjF36.3590
hypothetical protein, conserved Leishmania major chr 25	32293	9.51	27%	10	***302***	***82***	LmjF25.0540
RNA helicase, putative Leishmania major chr 21	49456	5.89	24%	8	***291***	***37***	LmjF21.1552
PGKB phosphoglycerate kinase B, cytosolic Leishmania major chr 20	44872	8.05	27%	9	***276***	***50***	LmjF20.0110

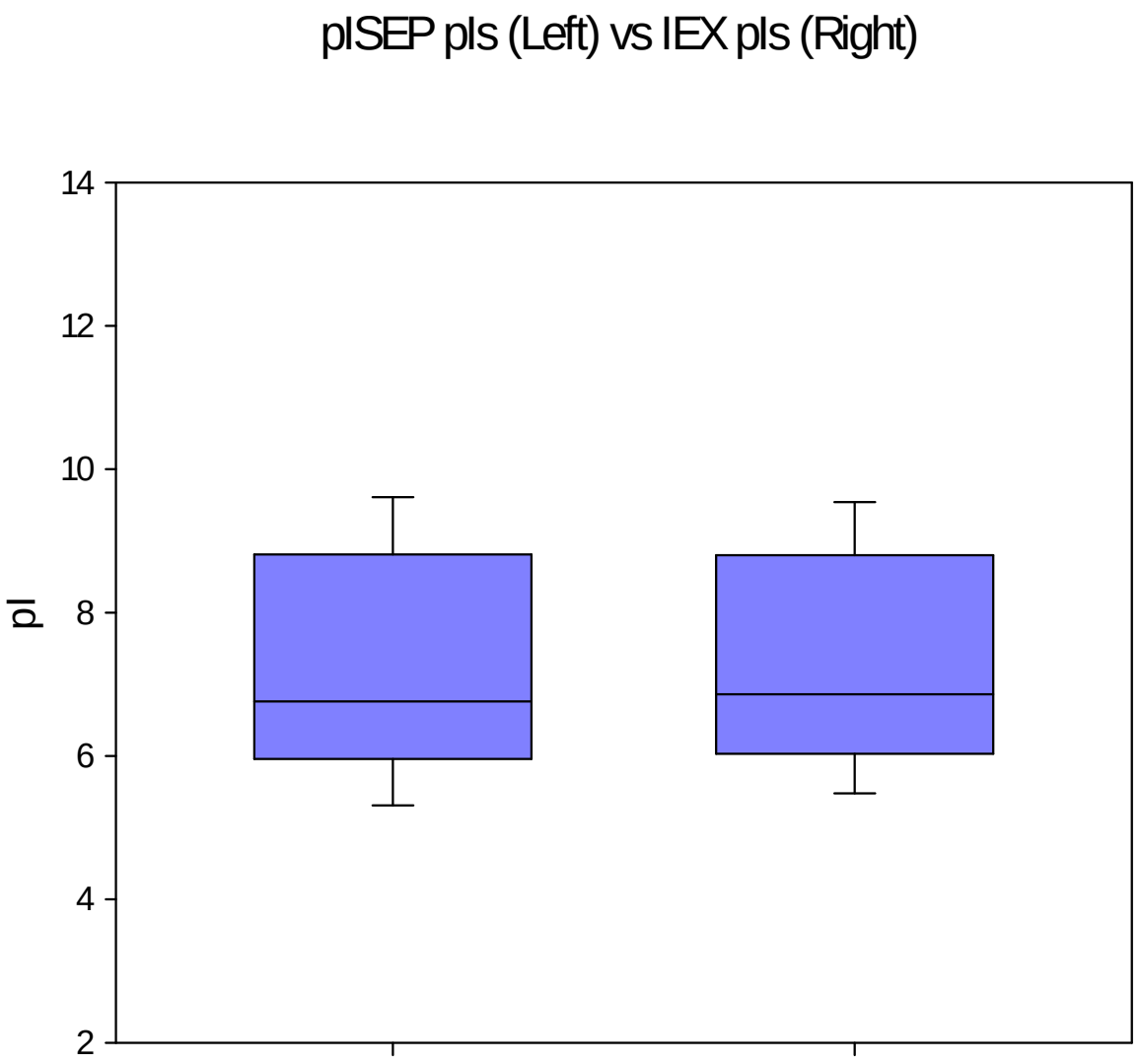


Figure 61: Box plot of the pIs of proteins identified using the pISEP/RP technique (left) and the IEX/RP technique. There is no significant difference between the pI range of both sets of data.

The methodology seems fairly robust. If suitable care is taken in making up the buffers then the gradients appear to form well and give good linearity and separation performance. It should be noted that a UV chromatogram at 214 nm cannot be obtained due to the high absorbance of the buffer at this wavelength and even at 254 nm there is a significant amount of drift in the baseline that makes it difficult to analyse small quantities of protein. The lowest practical loading on the MonoQ 5/5 is estimated to be 5 µg of individual

protein species. For 25 cm ProPac™ columns the lowest practical loading is estimated to be 0.5 µg of individual protein species.

As described in 3.2.4, the closest technique to the method described in this thesis is the PF2D system (Beckman). When pISEP is applied as an alternative first dimension, the similarities increase due to the use of an ampholytic buffer-based chromatofocussing system in the first dimension. All of the arguments for an externally generated gradient over an ampholytic buffer based system apply equally to the PF2D system. The PF2D system implements a pH detection system which controls fraction collection to ameliorate the effects of the poor reproducibility of the gradient, but this results in variable-sized fractions which may cause problems for downstream analysis. In addition, column lifetime is poor, with 4-6 runs being reported as being standard, although this limitation can be overcome to some extent with the use of different separation and re-equilibration conditions [38].

The only chromatofocussing experiment applied to *Leishmania* in the literature is that of de Andrade [209], in which a pH gradient was applied as the final polishing step of a serine protease purification protocol. Although not a proteomics experiment, the experiment does involve the separation of protein species by chromatofocussing. The gradient was generated with the use of polybuffer (GE Healthcare) and applied to a MonoP weak anion exchange column. The pH gradient generated approaches linearity, but tails off at approximately pH 4.3, resulting in a complete gradient from pH 7.3 to pH 4.3 – a fairly narrow range. In comparison with pISEP, the range and linearity is fairly poor. A non-quantitative proteomic experiment with similarities to the study of *Leishmania* described in this chapter was performed on human breast cancer cell (CA1a.cl1) lysates by Yoo [210]. In this instance, 3 mg of protein (average for this type of experiment) was used as the initial material, and separated on a gradient from pH4-7, using polybuffer 74 (GE Healthcare). After chromatofocussing using the PF2D system and second dimension separation using NPS-RP together with direct analysis with ESI-TOF, samples were dried completely and subjected to enzymatic digestion, after which they were analysed by a variety of methods: MALDI-TOF, MALDI-QIT-TOF, and PS-DVB monolithic LC-ESI-MS. The general workflow employed was to split chromatofocussing fractions into three parts, generate intact mass data from the NPS-RP second dimension, perform a digest of the initial chromatofocussing fractions and analyse them by PS-DVB-LC-ESI-MS/MS, and finally confirm any single peptide identification by analysis of digested second dimension fractions using PS-DVB-LC-ESI-MS/MS. Interestingly, to compare intact masses of

observed proteins with single peptide identifications, a mass window of 500Da was chosen as additional confirmation. Although the mass accuracy of ESI-TOF instruments can be as low as 100 ppm, this equates to 1 Da for a 10 kDa protein. Of course PTMs, especially truncation events, will significantly affect the mass of a protein, and should be taken into account in identifications, but a mass window of 500 Da is extremely large.

The complete set of proteins identified is not shown, as only a small number of the fractions were completely analysed, and attention was given to those proteins identified using multiple techniques. The separation quality, although excellent in the second dimension NPS-RP separations, is poor in the chromatofocussing dimension with peak widths of several minutes visible, and clearly shows the utility of external gradient chromatofocussing, as exemplified by pISEP.

4.1.6. Conclusions:

Most of the development work performed on pISEP by CryoBioPhysica used GE Healthcare MonoQ columns. Improvement in the quality of the chromatograms is confirmed by analysis of the technique with the ProPac™ columns used for high resolution protein separations. Separation quality of the 9 protein mix is visibly enhanced by the use of pISEP, and the separation of the *Leishmania donovani* lysate gives results comparable to the ion exchange/reversed phase separation. pISEP does appear to have significant potential for proteomics work. One key feature is that it enables a wider range of pIs of protein to be separated on a single column compared to ion exchange, rather than relying on analysis of the washthrough peak for proteins with pIs of higher than 8.

Ascending gradient separations provide equivalent peak widths to the descending gradient, however lower complexity chromatograms suggest loss of protein by acid precipitation. Second dimension analyses provide additional evidence of good separation of complex mixtures for the descending gradient, and further emphasise the fewer peaks in the ascending gradient.

5. Chapter 5 - Rapid Peptide Separation using Monolithic Columns.

5.1.1. Aims:

It cannot be escaped that the number of analyses required to completely cover a two dimensional separation is very large. This limits the application of the method to those laboratories capable of applying an LC and MS system to the analysis of the samples exclusively for several weeks. To achieve faster analyses, matching those for competing methodologies, rapid peptide separations using PS-DVB monolithic columns were assessed.

The aims of this study were:

- To develop methods for performing rapid analysis of digested peptides using PS-DVB monoliths.
- To optimize spray conditions, and evaluate competing electrospray apparatus.
- To develop a parallel LC system, capable of eliminating injection and reequilibration times from separations.

5.1.2. Introduction

High throughput analysis of complex peptide samples is at the heart of modern proteomics research. Tandem MALDI mass spectrometry of trypsin digested samples is extremely rapid, but for prefractionated samples containing more than one or two proteins, a single MALDI spot does not provide sufficient sample for identification of more than a few proteins. For this reason, liquid chromatography coupled to electrospray tandem MS is the optimal technique for the analysis of most samples [211-213].

Using standard column technology, such as a silica-bead based C18 column, a typical analysis takes half to three quarters of an hour, including a 20 minute gradient, washing, and re-equilibration of the column. In addition, peaks corresponding to eluted peptides are usually characterised by a base width of approximately 30 seconds when using formic acid

as the ion pairing reagent [214]. This can be slightly improved with the use of other ion pairs, such as trifluoroacetic acid, but use of this acid, even in low concentrations, is associated with significant ion suppression, and is therefore deleterious for analysis by mass spectrometry [205].

Several authors (e.g. [215, 216]) have shown application of monolithic PS-DVB columns to fast peptide separations using rapid gradients and requilibration times. PS-DVB monolithic columns have a number of significant advantages over standard columns: they do not have the bead structure of porous resins, consequently, diffusion into pores is significantly lessened [52]. Additionally, the reversed phase separation is a characteristic of the packing material, rather than a result of derivitisation of a silica matrix with aliphatic carbon chains [69]. This direct interaction with surface chemistry leads to a lower resistance to mass transfer. In summary, due to the reduction in diffusion, low mass transfer resistance, and solid construction, monolithic columns are more robust, have faster wash and equilibration times, and provide significantly sharper peaks for peptide separations than conventional C18 columns [215].

Monolithic columns' significant drawback is a reduction in capacity, resulting in a requirement for larger bore widths and higher flow rates to achieve comparable loading to a silica-based column [217]. The bore size of the most commonly used PS-DVB monolithic column in the UK (the PepSwift 200, Dionex, UK) is 200 µm, flowing at three microlitres per minute. The combination of an HPLC equipped with this column and a fast-scanning ion trap is a relatively inexpensive source of high throughput proteomics data. However, due to the flowrate of this column being tenfold greater than a standard nanoflow HPLC, some discussion has occurred over the optimal method of electrospray. It was felt that a rigorous analysis of the most commonly used electrospray emitters would assist those who have, or plan to set up a monolithic electrospray system.

Five electrospray emitters were chosen as test systems for analysis on the HCT ion trap, three of these based on a standard, nebuliser assisted electrospray emitter with the appropriate high flow spray shield, and the remaining two were nanospray systems using the nanospray shield. The standard electrospray emitter was discounted as a source due to the very large internal diameter of the needle, which would negate the sharp peak widths provided by the monoliths due to mixing.

The first source used was the commercially available 'low flow' source from Agilent. This is a 50 µm internal diameter steel needle with a coaxial nebuliser assist. Two further sources consisted of modifications to the standard electrospray source. Each source is a 100 µm internal diameter steel needle with a coaxial nebuliser flow. The sources were modified by passing a 90 µm outer diameter, 20 µm inner diameter (the same internal diameter used in the HPLC system post-column) through the needle and (in the first instance) stripping the acrylamide coating from the last centimetre of tubing or (in the second instance) grinding the end to a point using a mechanical capillary grinder equipped with 2 µm grain abrasive paper. The two nanospray sources consisted of either the Proxeon stainless steel nanospray needle (a 5 cm x 30 µm ID ground ended non-tapered emitter) and a new-objective 360 µm OD, 50 µm ID picotip with a 30 µm ID tapered tip.

To test the efficacy of the five electrospray sources for the analysis, the identification rate of proteins derived from a membrane preparation from *E. coli* 0157 was evaluated. *E. coli* 0157 was chosen as a source of protein because it is a bacterium of medical importance with a number of unique virulence factors. It has a comparatively simple proteome and can be readily cultured in the laboratory.

5.1.3. Methods and Materials

5.1.3.1. Sample Preparation for Electrospray Source Comparison

Bacterial outer membrane proteins (OMP) were prepared in collaboration with Dr. Neil Inglis at the Moredun Institute, Peniciuk, Edinburgh, using a modification of the method described by Gauthier *et al.* [218]. *E. coli* O157:H7 (strain TUV933-0) was incubated statically overnight at 37°C in 10 mL of minimal essential medium (MEM), Eagle's HEPES modification (Sigma M7278), supplemented with L-glutamine at 2 mM before sub-culturing 1 in 50 into 100 mL of fresh, pre-warmed medium. Static incubation at 37°C was continued to mid-log phase (OD_{400}) in an atmosphere of 5% v/v CO_2. Organisms were harvested by centrifugation at 3500 x *g* and washed once in phosphate buffered saline (PBS) before resuspending in 1mL of ice-cold sonication buffer (10 mM Tris-HCl pH 7.0 containing protease inhibitors (Complete Mini™ – Roche)). The cell suspension was then sonicated x 5 for 15 seconds. Unbroken bacterial cells were removed by centrifugation at 16,000 x *g* for 2 minutes. Membrane fragments were collected by centrifugation at 50,000 x *g* for 1 hour at 4°C in a Beckman TL-100 bench-top ultracentrifuge. Pelleted

membranous material was washed once before resuspension in a final volume of 200 µl of sonication buffer containing 0.5% ($^{w}/_{v}$) N-lauroylsarcosine to solubilise the inner membrane. Outer membrane components were collected by further ultracentrifugation at 50,000 x *g* for 60 minutes at 4°C. The resulting pellet was washed once with 100 µl sonication buffer containing 0.5% w/v N-lauroylsarcosine before resuspension in 100 µl final volume of sonication buffer containing 0.5% w/v N-lauroylsacrosine and 0.1% SDS. Protein concentration was measured using a bicinchoninic acid-based assay [219]. 50 µg of *E.coli* OMP was loaded into a single large sample well (75 mm x 15 mm) of a discontinuous Tris-glycine SDS-PAGE mini-gel (3% w/v stacking gel; 10% w/v resolving gel) and fractionated at 200 V (constant voltage) over ~45 mins using a Mini-Protean™ II Dual Slab Cell (80 x 70 x 0.7 mm) (Bio-Rad Laboratories). Resolved proteins were visualised using SimplyBlue™ Safe Stain (Invitrogen). The stained gel (see Figure 62) was cut into 10 vertical lanes of 7.5mm in width. Each lane was then sliced horizontally from top to bottom to yield a series of 25 equal gel slices of 2.5 mm deep. Each of the resulting 250 gel slices was then subjected to standard in-gel de-staining, reduction, alkylation and trypsinolysis procedures. The resulting digest supernatants were combined in rows of 10 to obtain 25 pooled samples representing gel slices 1-25 inc. Pooled digest supernatants were transferred to sealed HPLC sample vials and stored at 4°C until required for ESI-MS/MS analysis. Samples of 4 µl of each of the 25 pooled digest supernatants were applied to the column by direct injection in a series of 5 sequential replicate injections. This process was repeated for each of the 5 individual ESI sources.

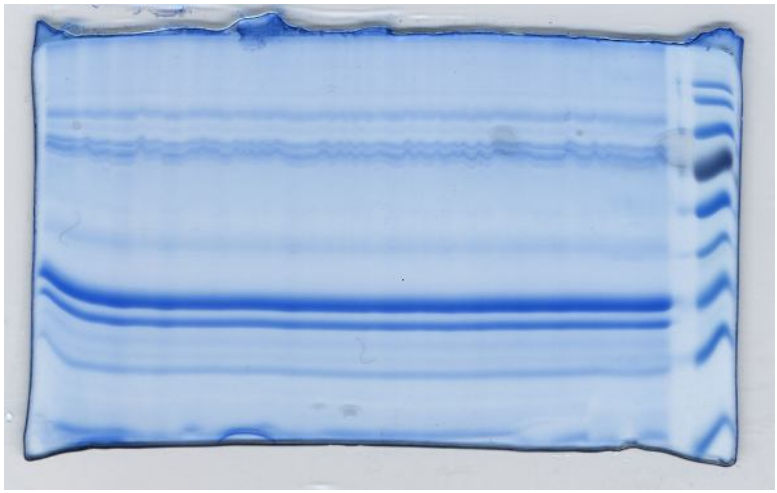


Figure 62: Coomassie stained slab gel of an *E. coli* outer membrane preparation. A well-less system was used to ensure equal loading across the gel, and provide clear bands for excision.

5.1.3.2. Sample Analysis by LC/MS

The UltiMate® 3000 (Dionex) HPLC system was configured in a direct injection configuration with a 5 µl loop in line with the 200 µm x 5 cm ProSwift monolithic column (Dionex). Post-column tubing was limited to 20 µm internal diameter to minimise mixing. The mass spectrometer used was a spherical ion trap (the HCT Ultra PTM discovery system) (Bruker Daltonics). Gradient conditions were 5% to 42% acetonitrile in 10 minutes, followed by 4 minutes wash at 72% acetonitrile, and 6 minutes equilibration at 5% acetonitrile for a total run-time of 20 minutes.

Source voltage was optimised to 4000 for the modified electrospray sources with a nebuliser pressure of 6 psi and a curtain gas pressure of 4 L/min. Scan rate was set to 26000 amu per second for both MS and MS/MS, with 3 averages used per recorded MS spectrum. The 5 most intense precursors were chosen for MS/MS per MS spectrum.

Data was collected and searched using MASCOT version 2.1 (Matrix Science). Only proteins identified above 95% confidence threshold in all 5 replicates were included for data analysis. Although this does not allow for misidentifications of proteins by the software, if a peptide is consistently observed and fragmented in every replicate while using a particular source, while remaining unobserved when using other sources, the observation is the key result, rather than the raw confidence interval of the identification. The 95% confidence threshold therefore serves as a filter for spectral quality.

5.1.3.3. Parallel Monolithic System LC/MS

The parallel monolithic system and analysis of samples are described in section 2.7.

5.1.4. Results and Discussion

Initially assessing the protein identifications at the threshold of 95% confidence, a total of 326 proteins were identified by combining all sources. This was reduced to a total of 136 proteins identified in all 5 sources when combined with the requirement that each protein was observed in 5 replicates (see Figure 63). As a percent of the total, the number of proteins from each source expressed was (see Figure 64):

Low Volume Sprayer = 81 (60%)
Ground Needle ESI Source = 77 (57%)
Proxeon Stainless Steel Needle = 6 (4%)

New Objective Picotip = 87 (87%)

Stripped Needle ESI Source = 121 (89%)

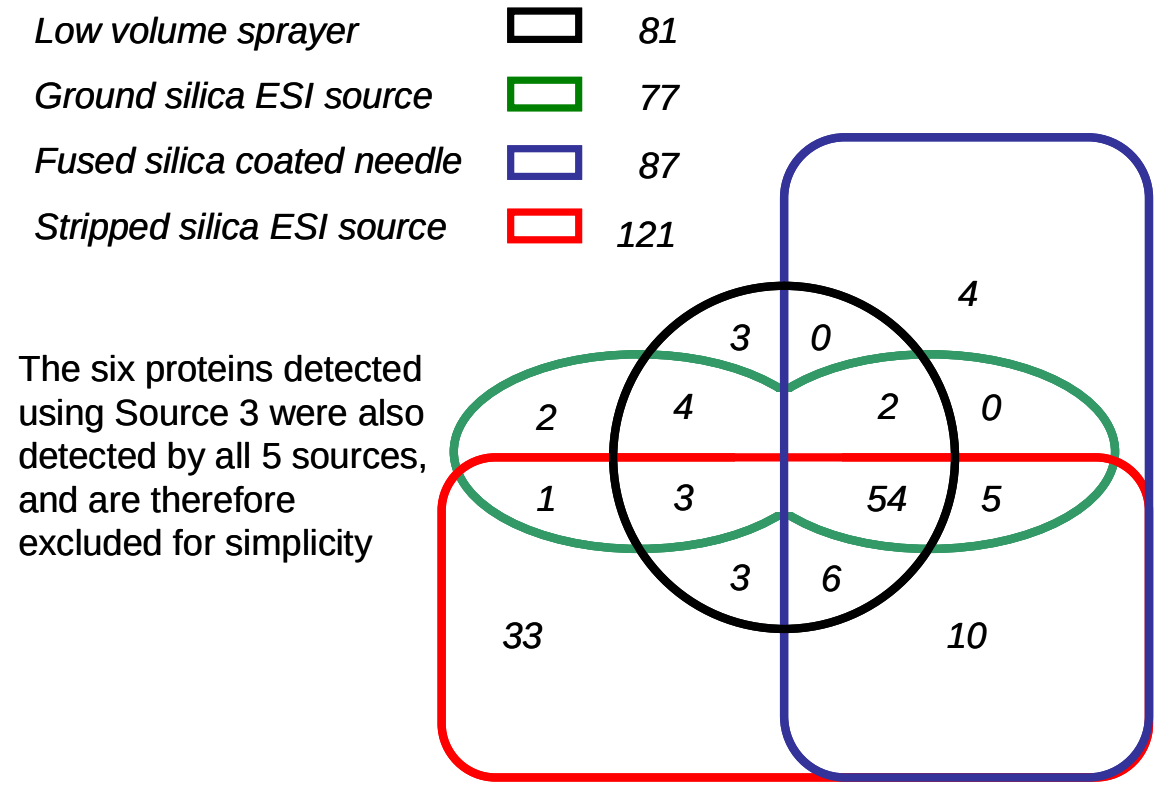


Figure 63: Diagram showing proteins identified in all five replicates per gel band from each source. Most proteins (54) were identified by 4 of the sources, while only 6 were identified by all five (data not shown). The stripped silica ESI source consistently and reproducibly identified more proteins overall than the other 4 sources.

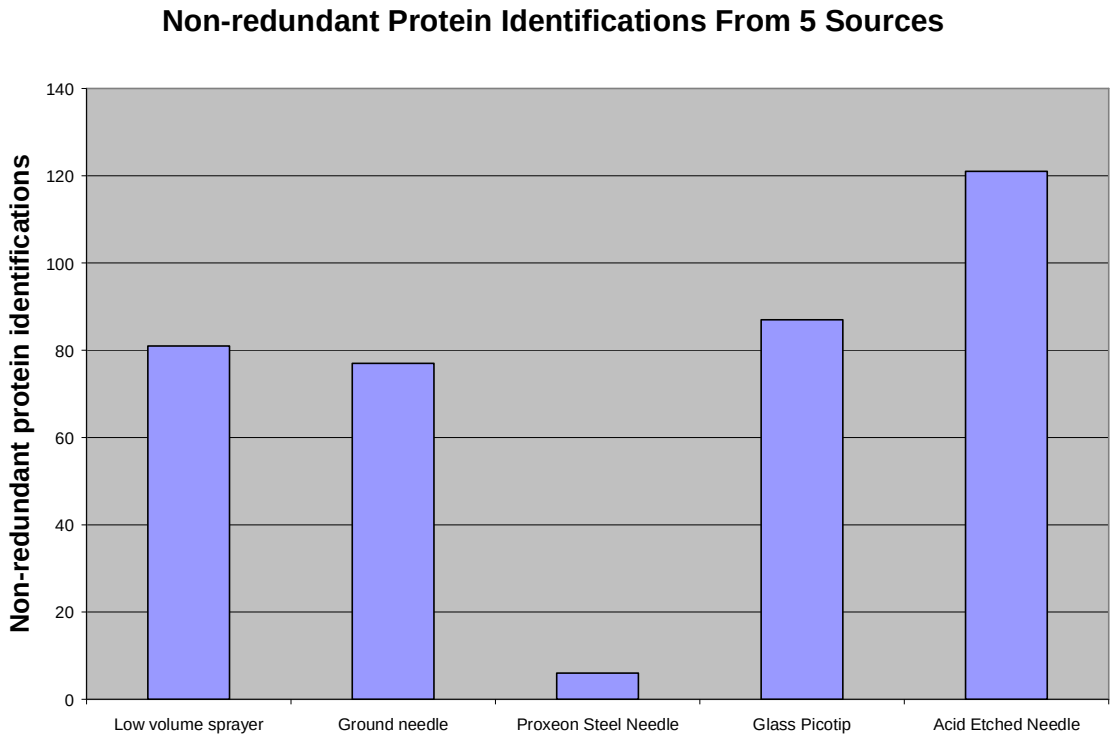


Figure 64: Protein identifications in all 5 replicates from all 5 sources. Although the Proxeon needle identified more proteins at higher scores overall, our strict requirement for 5 replicates resulted in poor performance in the reproducibility assessment.

Initial expectations were that the two 'nanospray' emitters (the Proxeon Steel Needle, and the New Objective Picotip) would be inherently more sensitive, providing a greater number of identifications at higher scores. This is due to the different design of the sprayshield and the less orthogonal angle of the emitter, allowing the mass spectrometer to intake more, potentially lower abundance ions. The data partially sustains this expectation:

Low Volume Sprayer =	704.5
Ground Needle ESI Source =	700
Proxeon Stainless Steel Needle =	929
New Objective Picotip =	787.7
Stripped Needle ESI Source =	758

By taking the average score of the 6 proteins observed reproducibly in all sources, it can be seen that the Picotip and Proxeon needle MASCOT scores are higher, slightly in the case of the Picotip, but significantly in the case of the Proxeon needle. Additionally, the Proxeon needle consistently provided the highest scores with the greatest number of identified peptides and sequence coverage (see Table 12). These identifications,

unfortunately, were inconsistent, and rarely found in more than two or three replicates of the same sample.

It is possible that the higher sensitivity provided by the Proxeon needle and the nanospray shield has resulted in so many available precursors that even with the extremely high scan rate it is impossible to scan a complete complement of precursor ions in chromatographic time, and therefore due to minor differences in ion intensity from run to run the same proteins are not identified in all replicates. We are therefore led to two conclusions depending on whether extreme sensitivity is preferred to overall reproducibility. The Proxeon stainless steel needle provides better raw data with higher scores and greater sequence coverage at the expense of reproducibility. The stripped needle ESI source on the other hand, provides lower scores in general, but is far more reproducible, identifying 121 of the 240 total proteins observed in all 5 replicates.

Table 12: Representative selection of protein identifications from each source, eliminating the requirement of 5 replicates. 'Scr' is the MOWSE score of the protein, '%' is percentage coverage of the protein length, 'no. peps' is the number of peptides observed in the identification. Both consistently high-scoring, and low scoring proteins were included. The proxeon needle shows a consistant, significant increase in score, sequence coverage and number of peptides identified.

Protein Name	**Low Volume Sprayer**			**Ground Needle ESI Source**			**Proxeon Steel Needle**			**New Objective Picotip**			**Stripped Needle ESI Source**		
	Scr	*%*	*No. Peps*	*Scr*	*%*	*No. Peps*	*Score*	*%*	*No. Peps*	*Score*	*%*	*No. Peps*	*Score*	*%*	*No. Peps*
Enolase	**513**	50	*18*	**529**	54	*17*	**790**	57	*21*	**622**	50	*15*	**597**	45	*17*
Membrane-bound ATP Synthase, F1 sector, Beta subunit	**442**	33	*14*	**310**	31	*11*	**832**	37	*17*	**667**	38	*13*	**602**	33	*13*
Protein chain elongation factor EF-Tu	**520**	49	*24*	**551**	58	*25*	**730**	53	*23*	**628**	50	*22*	**684**	49	*27*
Trigger Factor	**294**	32	*11*	**290**	26	*10*	**699**	34	*11*	**490**	24	*8*	**583**	32	*13*
OMP 1a	**683**	60	*23*	**671**	49	*26*	**868**	53	*27*	**685**	35	*19*	**579**	30	*27*
OMP 3a	**776**	54	*41*	**648**	50	*42*	**787**	57	*39*	**678**	48	*22*	**640**	45	*34*
CoA-linked acetaldehyde dehydrogenase	**1329**	47	*53*	**1255**	46	*54*	**1681**	45	*57*	**1417**	49	*51*	**1278**	46	*66*
PTS enzyme 2ab	**371**	51	*12*	**360**	33	*10*	**509**	37	*10*	**424**	27	*8*	**443**	33	*8*

Putative carbon starvation protein	**81**	5	*3*	**94**	5	*4*	**127**	8	*5*	**55**	3	*1*	**55**	4	*2*
Acid sensitivity protein	**85**	10	*3*	**90**	6	*2*	**178**	12	*4*	**118**	7	*2*	**81**	6	*2*

The difficulty of spray from the increased flow rate of monolithic columns has been ameliorated with the appropriate choice of electrospray emitter. In terms of LC-MS/MS workflow, direct injection implementation of monolithic separation is now common for fast scanning ion trap instrumentation. The advantage is that all of the peptides in the sample will be analysed by mass spectrometry, as washthrough from the column is simply eluted through the electrospray needle. This is, of course, also the technique's most obvious drawback, as any contaminants in the sample will not be excluded but will be sprayed into the mass spectrometer along with the sample. Additionally, there is a delay to the gradient commensurate with the time taken for the flow to pass through the loop, and a second delay resulting from the injection program. The latter can take anything up to 5 minutes, if, for example, automated resuspension of the peptides is performed by the autosampler prior to injection. To eliminate both problems, a valve-based switching system using two PS-DVB 'parallel' monolithic columns was implemented.

5.1.4.1. Implementation of Parallel Monolithic System

Initially the UltiMate® Plus Dual gradient system was used as the test bed for the system, coupled to the QStar Pulsar i (Applied Biosystems). The plumbing for the system involved an isocratic pump connected to the injection valve pumping acidified aqueous solvent through the loop, loading the sample on to one of two monolithic traps. Each trap was connectable by valve selection to a corresponding monolithic separation column, and an organic gradient was then applied to both trap and column, eluting peptides via the UV detector and mass spectrometer. During any LC run, one trap would be eluted via its separation column, while the other would be loaded and washed, thus eliminating the injection and wash times. Obviously, due to the direct connection of the gradient pump to the trap and column, the gradient delay due to the loop was also eliminated.

The Ion Spray source was attached and optimised to a voltage of 5200, declustering potential of 50V, nebulising gas pressure of 12 psi and curtain gas pressure of 50 psi. The rather high voltage, declustering potential and curtain gas pressure are a consequence of the lack of heated curtain gas, which made desolvation of the spray very difficult. UV data,

however, was promising. Optimisation of the capillary connections resulted in sharper peaks and alterations to the parallel programs allowed more effective trap equilibration. In concert, these changes produced substantially improved chromatograms. Myoglobin peptides were well separated with peak widths of approximately 3 seconds half-height (see Figure 65 and Figure 66).

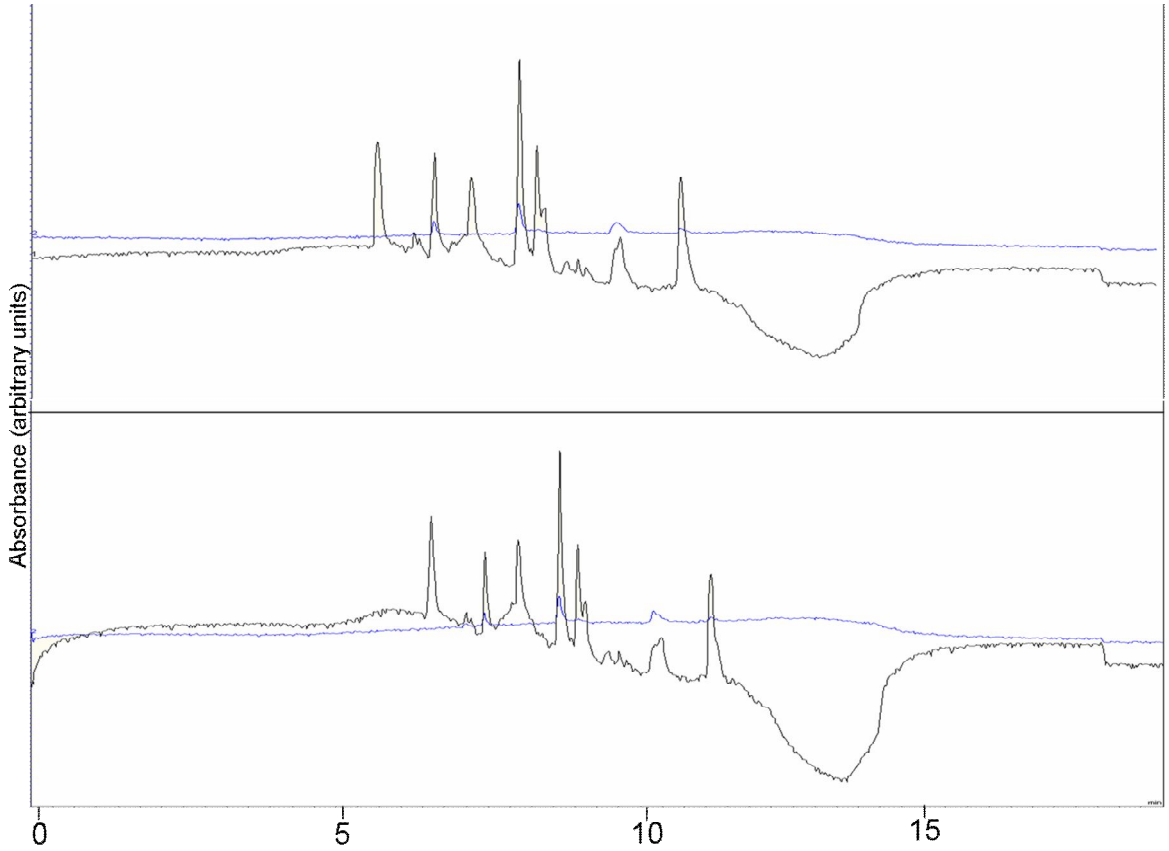


Figure 65: Result from the replacement of capillary lines in the parallel system. Myoglobin digest UV (214 nm shown in black, 280 nm shown in blue) traces performed on column A (top) and B (bottom) are shown. Note that peak width is significantly improved, as is overall separation quality.

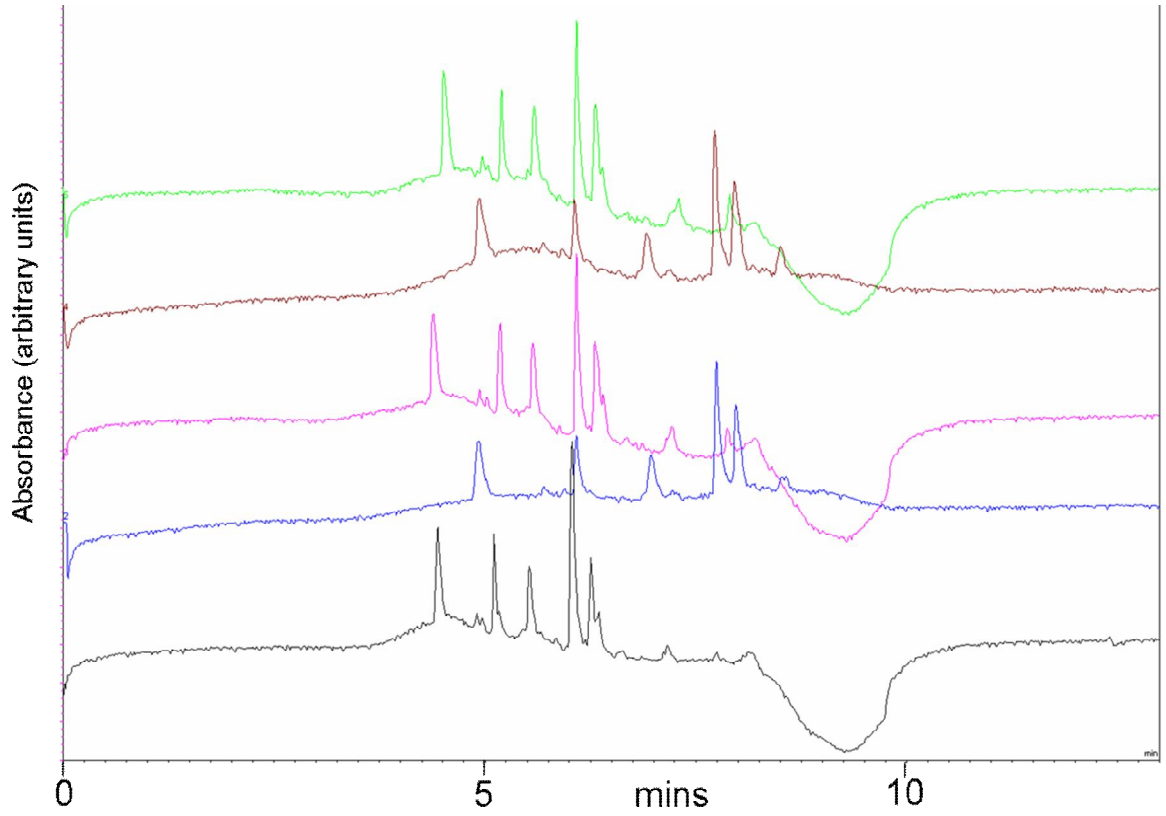


Figure 66: Stacked chromatogram of myoglobin separations. Chromatograms alternate between columns. Note that reproducibility is significantly improved, but the resolution of one column is significantly lower than the other. This is likely to be a result of post-column dead volume, resulting from the high complexity of the plumbing and the high backpressures generated.

Although UV traces show distict peptide peaks, very little is seen using MS analysis. This suggests that either the UV detector is more sensitive to low abundance peptides than both the QTrap and QStar mass spectrometers (which is unlikely), or that physical properties of the ion spray sources used on both machines are inappropriate for the flow rate of the system. In addition, the UltiMate® Plus dual gradient system was unable to provide reproducible chromatograms, as a drift in retention time over a period of hours was observed. This is likely to be due to the high backpressure resulting from the unheated tubing in the system.

Parallel monolithic analysis was later implemented on the more appropriate UltiMate® 3000 system. The primary advantage of this system is the combined column oven and valve compartment, ensuring that, between the injection loop and the transfer line to the detectors, all components of the system are maintained at a constant temperature. Function of the monolith was confirmed using direct injection onto a 200 µm x 5cm monolith column with a 5 µl loop. Rather than using the monolithic trap system, it was decided to implement a direct injection parallel set up, since this was fluidically less complicated and was considered likely to be more reliable. Using a flow rate of ~3 µl per minute with a 5 µl

loop, equilibration and loading of one column can be performed during the application of a 10 minute gradient to a previously loaded sample on the other column. A diagram of the fluidic connections is shown in Figure 67.

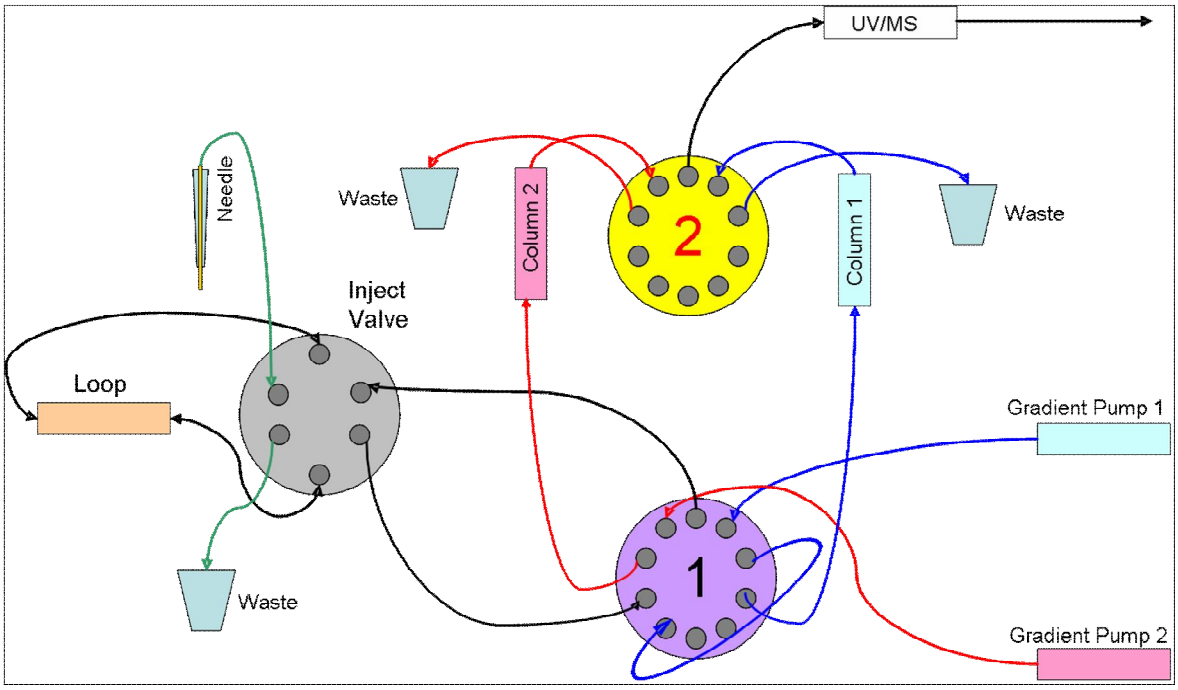


Figure 67: Valving diagram for parallel Direct Injection monolithic separations, allowing loading of one column while the other is eluted. Connections in blue correspond to flow paths for gradient pump 1. Connections in blue correspond to flow paths for gradient pump 2. Green connections denote the injection fluidic path and black connections contain either fluid from gradient pump 1 or 2 depending on valve switching.

Fluidic connections for parallel monoliths were attempted with the original PEEK stators, but sufficient torque for each connection could not be obtained without potential damage to the stator. Thus both stators were exchanged for stainless steel or titanium models.

Direct injection on monolithic columns and parallel LC are intended for fairly low complexity samples that have been subject to extensive prefractionation prior to analysis. Examples from intact protein ion exchange followed by monolithic reversed phase are shown in Figure 68 (for a direct injection example) and Figure 69 (for a parallel LC example). The former is a PC12 lysate, used to generate the SILAC data described in the quantitation chapter. The latter is a wild-type *Leishmania* sample, used to provide identifications for the UV map example, also described in the quantitation chapter.

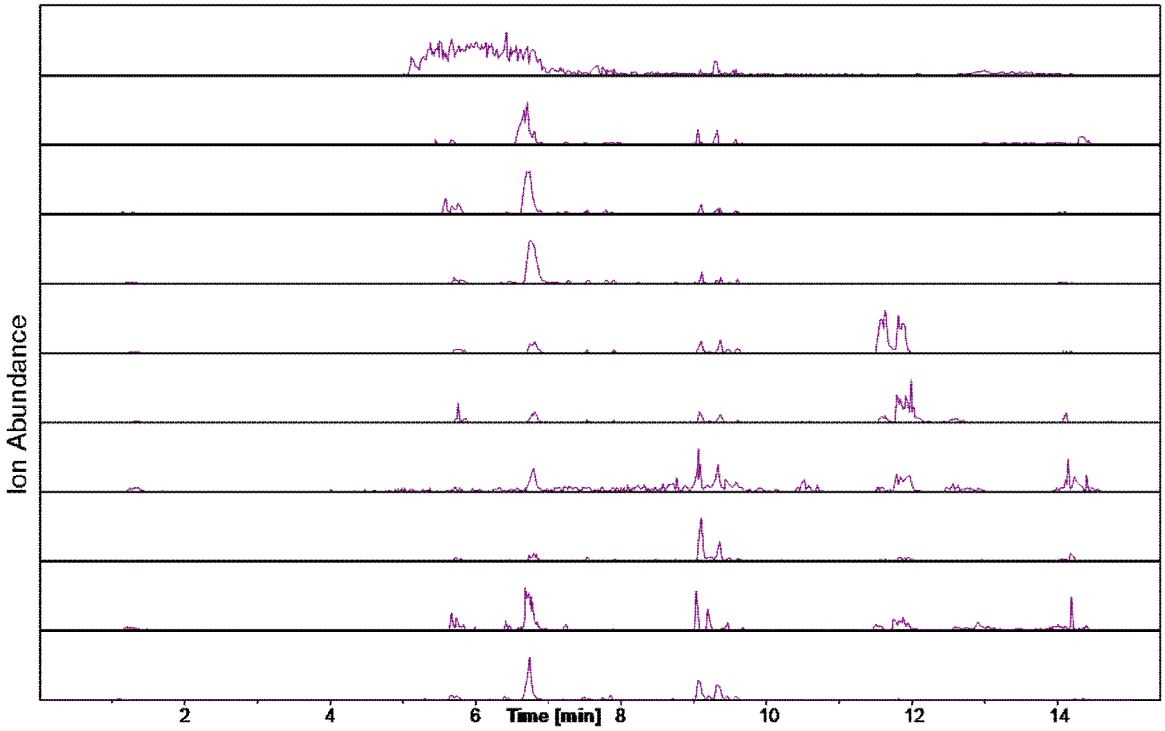


Figure 68: Stackplot of successive direct injection separations from a 2D intact LC separated SILAC labelled PC12 protein lysate. Note that the elution times of peaks common to all separations are identical.

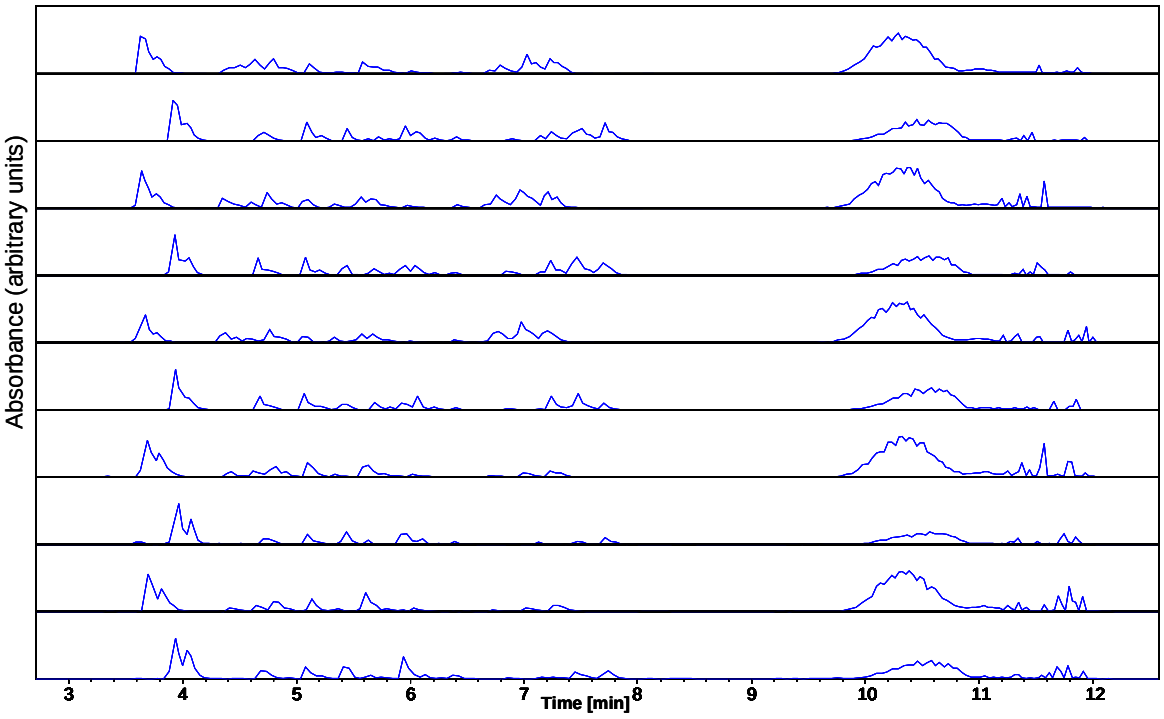


Figure 69: Stackplot of successive parallel LC separations of peptide fractions from a *Leishmania* intact protein LC separation. Although some peaks are common to all runs, characteristic of autolysis fragments of the trypsin used to digest the proteins, some peaks corresponding to *Leishmania* proteins appear and disappear in each run. The shift between alternating chromatograms is again due to unavoidable differences in dead volumes and packing material between column paths.

The final development of the parallel system involves a 7 minute gradient, for a total run time of 12 minutes (for examples of BSA digests separated on each LC channel see Figure

70 and Figure 71). Experimental data shows that, including post-processing of the sample, 4.8 samples can be run in one hour. Scores are slightly lower for a BSA standard in parallel mode than in direct injection monolithic LC, using both the Proxeon steel needle source and the MicroIon sprayer, but the decrease in run-time allows significantly more samples to be analysed in the same amount of time (see Figure 72). The reduction in score is primarily due to the loss of hydrophilic peptides which pass through the column during loading – in a direct injection system they are passed directly to the MS system. This, of course has the significant drawback that anything else in the injection will also be passed to the MS, including salts and detergents which would normally be removed by washing. Indeed, one of the significant benefits of the parallel system is that in a direct injection scheme, the curtain plate of the MS must be cleaned every two days, due to build up of salts, whereas the parallel system can run for extended periods without intervention due to the extensive washing of the columns while they are offline.

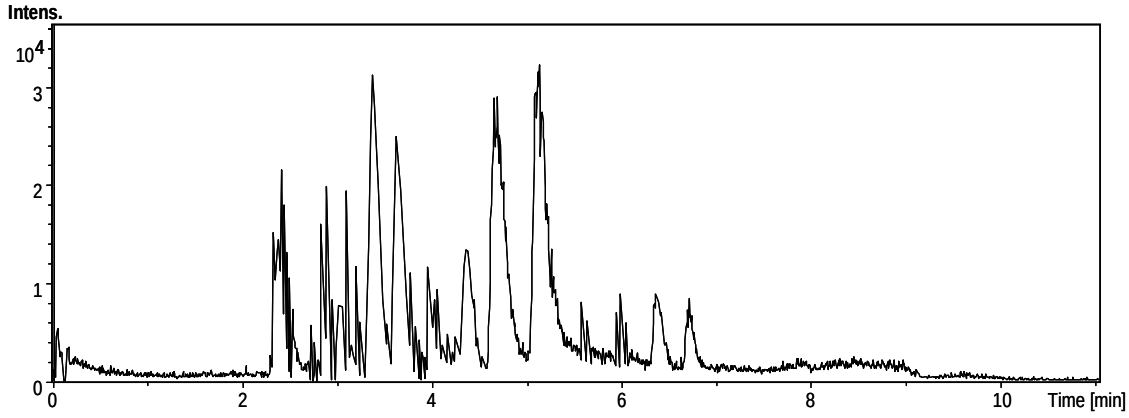


Figure 70: Base peak chromatogram from a BSA separation on column A in the direct injection parallel system. The total separation time is 12 minutes, and loading time is eliminated between runs.

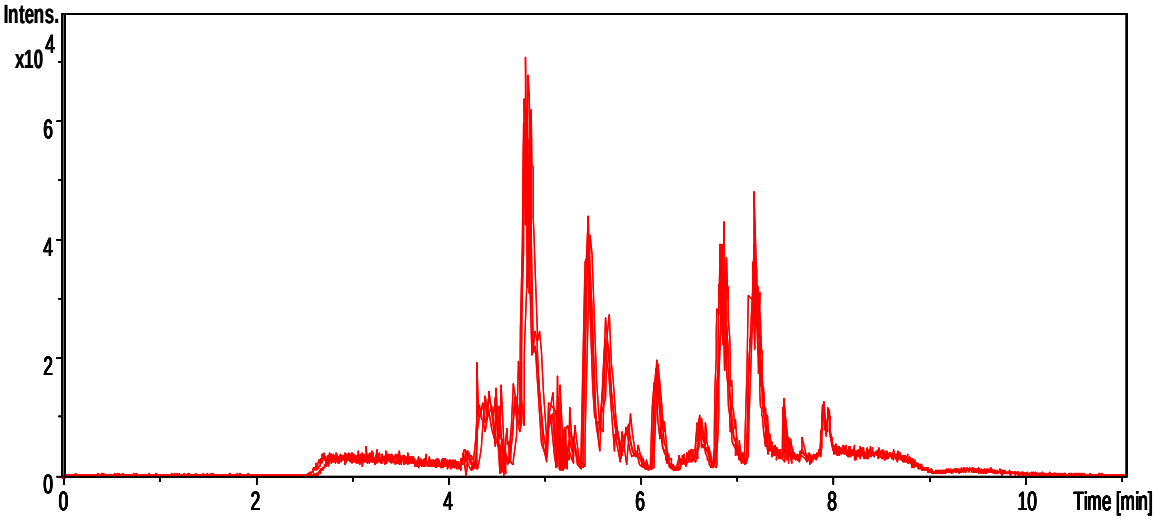


Figure 71: Base peak chromatogram from a BSA separation on column B in the final implementation of the direct injection parallel system. Although the chromatogram is slightly different from that in Figure 70, this is predominantly due to the longer path-length in the column resulting in a delay prior to elution of the first peak.

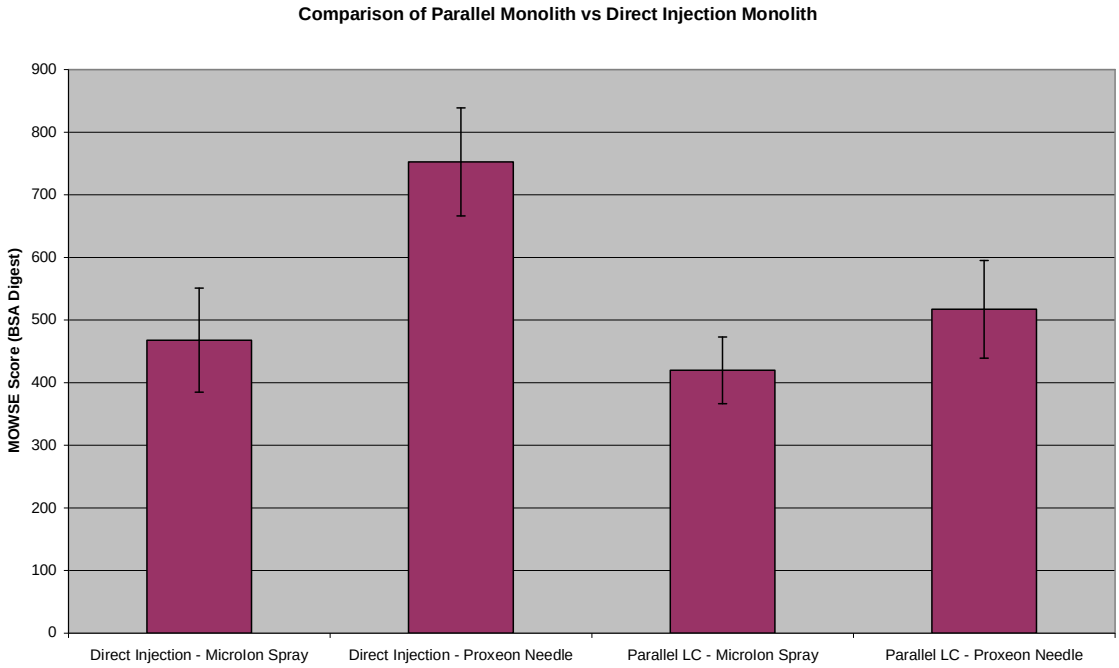


Figure 72: Mowse scores for 500 fmol BSA digest analysed by four different methodologies: direct injection monolithic separation with the MicroIon sprayer and Proxeon stainless steel needle, and parallel separation with the MicroIon sprayer and Proxeon stainless steel needle.

No discussion of the use of polystyrene divinyl benzene resins for peptide and protein separations would be complete without reference to Christian Huber's development work on both pellicular and monolithic columns. Early papers were predominantly focused on the separation of oligonucleotides, but later pellicular derivitised and underivitised columns were applied to the separation of peptides and proteins, giving extremely sharp peak widths [220]. Interestingly, particles derivitised with C18 chemistry were the most effective method for separating both proteins and peptides, including larger proteins such as ovalbumin and carbonic anhydrase, which, in general, separate very poorly on derivitised silica columns. The first application of PS-DVB monolithic columns to proteomics work occurred in 2001 [217], with the application of 200 µm x 6 cm un-derivitised monoliths applied to standard samples, including a transferrin digest and a standard mixture of intact proteins. The standard temperature run throughout Prof. Huber's separation appears to be 80ºC, which was lowered to 60ºC in the work presented here due to anticipated degradation of the columns, but which will contribute to the extremely high resolution obtained. A paper by Oberacher et al. [52] further describes in detail the separation characteristics of monolithic columns, especially focussed on the low resistance

to mass transfer and consequently high resolution of separations common to these columns. Additionally the reproducibility of column manufacture is described, and shows the variability inherent in polymerisation and pore forming during column preparation. Rather than using, as in earlier papers, 6 cm x 200 µm columns packed by the researchers, the study used the same commercially available 5 cm x 200 µm columns from LC Packings (now Dionex) used herein, as well as octadecyl derivitised hand-packed pelicular columns, and shows that monolithic columns provide a smaller theoretical plate height, leading to higher resolution.

The paper which originally describes the modification to the ESI source sprayer using 90 µm OD/20 µm ID capillary was published in [221], and was applied to monolithic separations of proteins, peptides and DNA. Interestingly, TFA is used throughout as an ion pair, which is generally regarded as resulting in significant signal degradation in MS, but which does provide excellent chromatography, which, it is expected, was the reason for its use in a paper focussed on the chromatographic performance of monolithic PS-DVB columns. The use of TFA as an ion pair is reprised in a later paper, presented by Walcher et al. [222], which describes both the effects of temperature and ion pair on the retention of peptides on monolithic PS-DVB columns. Although increasing temperature is a well characterised method of improving chromatographic separation, the conclusion that HFBA and TFA provide lower limits of detection for a variety of peptides than formic acid flies in the face of most other papers on the matter (e.g. [207], [223], [224]). It is possible that the specific conditions, i.e. PS-DVB monoliths, coupled with the modified electrospray source previously described in [221] applied to an HCT ion trap with its unusual reversed configuration source (the voltage is applied to the curtain plate, and the needle is grounded, rather than the voltage being applied to the electrospray needle with a grounded curtain plate), may be able to more effectively evaporate the unpaired TFA than standard sources.

An obvious question upon considering the use of monolithic columns is their performance in comparison to conventional silica resin columns. Toll et al. [225] show conclusively that, when applied to identical gradient lengths, 100 µm monoliths have substantially better performance with higher average MOWSE scores than 75 µm, 3 µm particle size octadecyl derivitised silica bead columns. When applied to 200 µm monoliths, the MOWSE scores for more abundant proteins are consistantly higher than those for C18 columns, but for less abundant proteins, scores are lower. This is probably a consequence of poorer desolvation of the increased solvent volumes resulting from the higher flow rates used in these monoliths. As data shown in this thesis demonstrates, the use of the Proxeon stainless steel

needle significantly increases the MOWSE scores produced by both protein standards and complex mixtures at a cost of reproducibility, which implies that the performance of 200µm monoliths can be rescued by the implementation of a different ESI source. A paper by Batycka et al. [215] describes a comparison between short gradients (9 minutes) performed on a 200 µm monolithic PS-DVB column and standard 30 minute gradients performed on a 75 µm, 3 µm particle size octadecyl derivitised silica bead column. There are a number of interesting features to the experimental design presented. First, the paper describes the electrospray source used for the comparisons as being a coated glass Picotip (New Objective), despite an accompanying table showing a nebuliser assisted source for the monolith, and a nebuliser unassisted source for the silica column. This is critical, because, as described above, the type of source used has a significant effect on the quality of data returned. Furthermore, the drying gas is listed as being 0.5 L/min for the silica column, and 5L/minute for the monolithic column. The latter is standard for a capillary flow source, but the former is extremely low. A value of 10 L/min is considered normal, as reducing the gas flow substantially results in poor desolvation of solvents, leading to 'spiking' of random, high intensity ions in the trap. A later paper [226], which applied an array of monolithic columns to the separation of peptides and proteins shows early application of the type of 'parallel LC' described in this chapter. Four columns were applied in a direct injection configuration with a single injector feeding them with samples. This allows extremely rapid parallel separation of multiple protein and peptide samples, but the requirement for a separate detector coupled to each column limits the utility of the system for electrospray, due to the requirement for an equal number of mass spectrometers. In the paper, UV detectors, which are substantially cheaper than a mass spectrometer, were used. One application where this configuration may be very effective is in LC/MALDI. Using a multi-head spotter would allow extremely rapid analysis of multiple samples.

5.1.5. Conclusions

Analysis of peptides by rapid gradients on monolithic columns is an effective way of improving the throughput of proteomic separations. Sensitivity is enhanced with the use of steel nanospray needles, and reproducibility is enhanced with the use of the Huber group's modified electrospray source. The conclusions of the evaluation of electrospray sources was believed to be of general interest to the proteomic community and has been submitted for publication. Reproducibility of separations on the direct injection parallel system is very good, and throughput of analysis is much improved. Reliability of the system is such that multiple 384-well plates of samples can be run consecutively without intervention save

replenishment of solvents. Coupled to a fast scanning mass spectrometer such as the HCT Ultra (Bruker) parallel LC allows very fast analysis of a large number of samples. Drawbacks to the system are the use of formic acid to both load and elute the sample. TFA provides sharper peaks and improved retention of the peptides on the stationary phase, but due to its ion suppressing qualities, its use was avoided in the buffers. Further work should investigate the use of very low concentrations of TFA for LC-MS/MS purposes on this system.

6. Chapter 6 – Quantitation for Intact Protein LC

6.1. Aims:

Any new general proteomics technique must be compatible with quantitative methods. A variety of methodologies are available for protein quantitation, but most are designed for post-digestion labeling, which is clearly incompatible with an intact protein separation technology. The development of quantitative methodologies for intact protein separation using LC was focused on:

1. Investigation of the potential for UV map-based quantitation and gel-equivalent analysis using standard 2-D gel quantitation software.
2. Testing the isotopic labeling strategies SILAC, iTRAQ, ICAT, ICPL, ExacTag, and DiMethyl, and assessing if modifications to protocols intended for separation after proteolytic digestions could eliminate the necessity for the initial digestion step.
3. To test the efficacy of label-free methodologies, particularly with Top-Down workflows.

6.2. Introduction

Effective, accurate, and reproducible quantitation is of paramount importance to proteomics. As described in Chapter 1, there are three principle methodologies for the quantitation of proteomic samples: fluorescence and absorbance, isotopic labeling, and 'label-free' approaches. One or more methods in each of these three different categories have been applied to intact protein separation: quantitation based on the UV absorbance of the separated chromatograms; isotopic labeling based on the SILAC, iTRAQ, ICAT Dimethyl and ExacTag reagents; and finally the Progenesis software from NonLinear Dynamics for label free analysis (applications described in section 7)

6.2.1.1. UV absorbance based quantitation.

Proteins and peptides absorb ultraviolet light with maxima at 214 (for the peptide bond) and 280 (for ring structures, especially tryptophan) [227]. Chromatograms at these wavelengths are normally obtained during a separation to observe elution of protein and

peptide species from the separation column. For a multidimensional separation, these can be extracted from the software, processed, and displayed as two dimensional gel-style maps. These maps are sufficiently similar to 2-D PAGE gels to enable analysis using the same software.

6.2.1.2. Isotopical Label Based Quantitation

The different reagents used for isotopic quantitation have been discussed in depth in the introduction. In summary they can be categorized into two groups: isobaric (same mass) or non-isobaric (different mass) reagents. Isobaric reagents (such as ExacTag or iTRAQ) fragment into differential mass reporter ions during tandem MS, non isobaric reagents (such as SILAC or ICAT) display a characteristic mass shift in the MS spectrum.

The amino acid specificity of an isotopic labeling strategy has a significant impact on its suitability for intact protein labeling. Most isotopic labels (with the exception of SILAC and dimethyl labelling) are either thiol-reactive reagents, labeling cysteines, or NHS esters, labeling lysines and the N termini of peptides. The ExacTag protocol is a special case where the lysine labeling version of ExacTag contains a reagent that converts the amine group of lysine to a thiol group, and then labels it with the thiol reagent. There are considerably more lysines in the average protein than there are cysteines, and therefore quantitation using lysine labeling reagents is of higher statistical reliability. Indeed, a sizeable percentage [143] of proteins do not contain cysteine residues at all. The particular benefit of cysteine labeling approaches is to intact protein labeling work: most of the reagents are not especially pH sensitive, and saturation labeling of proteins is routine (e.g. [142, 228]). Furthermore, cysteine labeling reagents do not pose problems for subsequent tryptic digestion, a significant drawback associated with the use of lysine and N-terminal labeling reagents. Lysine labeling reagents are commonly used post-digestion, and thus often label a peptide twice, once on the N-terminus, and once on the lysine residue, since trypsin cuts C-terminally after lysine and arginine. In a protein labeling situation most of the benefits of lysine labeling are negated, due to the necessity of labeling prior to separation and hence digestion. First, tryptic digestion of proteins is predicted to fail at labeled lysine residues and only cut at arginine residues, leading to the production of significantly larger peptides with more complex fragment spectra, and secondly, the reagent will only label lysine residues and unmodified protein N-termini, leading to fewer labels and weaker statistical strength.

6.2.1.3. Label free methodologies

Label free methodologies are growing in popularity for several reasons: they require no special processing of the sample, they are not affected by labeling efficiency and they do not suffer the 'multiple proteins per peak' problem of UV and fluorescence based quantitation, where it is impossible to ascertain with certainty the species responsible for a particular peak. The current methods for performing label free quantitation are described in section 1.7.3.8, but the most commonly used are DeCyder MS (GE Healthcare) and MSight (Proteome Informatics Group) [149]. All current label free quantitation software is based on gel analysis software. The principle is based on graphing mass against time for an LC/MS chromatogram, where peak intensity is described by the intensity of a spot, analogous to the spots on a 2D PAGE gel. The resulting graphs may then be aligned and differences in spot area can be assessed. Currently, there is no software capable of producing quantitation over multiple LC runs, and, because most complex samples are analysed by multidimensional LC, label-free methods are most suited to low-complexity samples that can be analysed by a single dimension of liquid chromatography. We have therefore primarily applied label free quantitation to top-down analysis, and the research is described in section 7.4.1.2. The software used was an early release beta version of the Progenesis LC/MS software (NonLinear Dynamics), which is a development of their SameSpot 2D PAGE matching software.

6.3. Methods and materials

6.3.1.1. UV-based quantitation

Leishmania donovani lysates were grown by Christina Nuala and lysed by the author as described in 2.2.2, followed by separation using the on-line parallel separation methodology described in section 2.5. After fraction collection into 384 well plates, samples were analysed by direct injection LC/MS.

200 µg of pentamidine resistant *Leishmania donovani* lysate were separated using the methodology described above. A 16 fraction ion-exchange separation was performed, with 24 fractions collected in the reversed phase dimension, giving a total of 384 fractions. Collected fractions were subjected to tryptic digestion and analysed by LC-MS/MS as described in 2.6.3. A total of 499 unique significantly scoring proteins were identified using the MASCOT software (Matrix Science) version 2.1.

Chromatograms were obtained from each of the 16 ion exchange fractions separated by monolithic reversed phase from each of the samples. Each chromatogram was exported as a set of time - intensity pairs and combined using SigmaPlot 10.0 (Systat Software) into a contour plot. Number of contours was set to maximum, and line thickness was set to minimum, to produce as smooth a map as possible. The intensity scale was set to a maximum of 175 for both images, to prevent saturation of the image, which interferes with spot detection and quantitation. Each image was exported as a 'Microsoft print image' and saved as a tagged image file format (TIFF) using the Microsoft Document Viewer. Both images were then negativised, and converted to 16-bit greyscale TIFF using Adobe Photoshop. Alpha blending was excluded. The resulting images were then processed using the DIA software (GE Healthcare) using a target spot number of 500.

6.3.1.2. Isotopically based quantitation.

6.3.1.2.1. Reagents and Samples

General reagents were obtained as described in section 2.1. Cleavable ICAT and iTRAQ reagents were obtained from Applied Biosystems, ExacTag was supplied by Perkin Elmer (UK). ICPL reagents were supplied by Bruker Daltonics (UK). For ICAT, iTRAQ and ExacTag experiments, two stock solutions of 6 proteins were prepared for methods development for quantitative proteomics as shown in Table 13. Further to successful labelling with of the standards, *Leishmania* lysates were produced for ICAT and iTRAQ samples as described in section 2.2.2.

Table 13: Ratios of concentration in six protein mixes. Two stock solutions were prepared containing the protein amounts shown in the table dissolved in ddH_2O. The solutions were then aliquotted into 100 µg lots with 10 µL in each.

Protein	Amount in A	Amount in B	Ratio
BSA	650µg	650µg	1:1
Myoglobin	150µg	1500µg	1:10
Carbonic Anhydrase	1500µg	150µg	10:1
Cytochrome C	300µg	1500µg	1:5
Ovalbumin	1200µg	300µg	4:1
α-lactalbumin	900µg	600µg	3:2

6.3.1.2.2. Isotopic Labelling Protocols

Details are given for all labelling protocols except for ExacTag and ICPL, since these were carried out as per the supplied documentation.

SILAC samples

SILAC samples were produced in by Christian Preisinger and Alex von Kriegsheim at the Beatson Cancer Research Institute, and were supplied as lysates as described in sections 2.2.3 and 2.2.4. Samples for duplex quantitation were wild-type and lysine $^{13}C_6$ labelled K562 cells, which are cells from a patient in the final stages of chronic myeloid leukaemia. In this state, the immune cells are in what is known as blast crisis, where the cells are poorly differentiated, and contain a constitutively active BCRAbl protein, the result of a chromosomal translocation. This change is anti-apoptotic and pro-survival and pro-proliferation. Cells were harvested 6 hours after treatment with Glivec which inhibits the tyrosine kinase activity of BCRAbl, resulting in apoptosis. Samples for triplex quantitation were wild-type, arginine $^{13}C_6$ and $^{13}C_6{}^{15}N_4$ labelled PC12 cells, grown in a serum free medium. PC12 cells, when stimulated with neural growth factor (NGF), differentiate into neuron-analogues, and, when stimulated with epidermal growth factor (EGF), proliferate. They are therefore an ideal cell type for the study of the biochemical pathways in control of differentiation and proliferation.

ICAT protocol

The ICAT protocol used in this thesis is adapted from the 'Cleaveable ICAT Reagent Kit for Protein Labelling' (Applied Biosystems, UK; part number 4333373). The protocol is intended for a peptide LC-MS/MS experiment, and therefore required significant modification for use with protein separation. 100 μg of the control sample (6 protein mix A or wild-type *L. donovani* lysate) and 100 μg of the test sample (6 protein mix B or pentamidine resistant *L. donovani* lysate) were diluted to (in the case of the 6 protein mix) or resuspended in (in the case of the *L. donavani* lysate) 80 μL with 50 mM Tris-HCl, 0.1% w/v SDS. 2 μL of 50 mM TCEP was added to each sample, followed by boiling for 10 minutes to reduce cysteine residues. 20 μL of acetonitrile was added to light and heavy ICAT labelling reagent vials, followed by vortexing. After cooling, the reduced control

sample was added to the light reagent vial and the reduced test sample was added to the heavy reagent vial followed by vortexing. The labelling reaction was allowed to proceed for 2 hours at 37ºC. After labelling, the samples were mixed together, and 90 µL of 95% TFA was added to cleave biotin. Cleavage continued for 2 hours, followed by neutralisation of the acid with a few drops of 7 N ammonium hydroxide. Desalting was performed with a desalting spin column (Pierce, UK), followed by two dimensional protein separation as described in sections 2.3.2 and 2.4. Once separated, samples were digested and analysed by MS. For the 6 protein mix, direct MALDI spotting of digested samples from the collection plates was employed (see section 2.6.1). For the leishmania samples, analysis was performed using parallel monolithic LC-MS/MS of the digests (see section 2.7).

iTRAQ protocol

The final iTRAQ protocol was adapted from the 'Applied Biosystems iTRAQ Reagents Application Kit – Protein' (part number 4370075). This protocol is intended for application to SDS PAGE gels, and therefore deviation from the protocol was necessary in the final stages.

25 µg of the control sample (6 protein mix A or wild-type *L. donovani* lysate) and 25 µg of the test sample (6 protein mix B or pentamidine resistant *L. donovani* lysate) were (for the 6 protein mix) diluted to 20 µL with 0.1% w/v SDS, 0.5 M triethyl ammonium bicarbonate or resuspended from an acetone precipitated pellet with 0.1% w/v SDS, 0.5 M triethyl ammonium bicarbonate, were reduced with TCEP. Samples were reduced using 2 µL of 50 mM TCEP for 1 hour at 60ºC, followed by alkylation using 1 µL of 200 mM iodoacetamide for 30 mins at room temperature in the dark. 20 µL of ddH_2O was added to each sample. 50 µL of ethanol was added to each of a light and heavy iTRAQ reagent vial (the particular reporter ion used for light or heavy labelling was dependant on the experiment), and vortexing was performed to ensure that all reagent was dissolved. The light iTRAQ reagent solution was added to the control sample, and the heavy iTRAQ reagent solution was added to the test sample. After vortexing, the samples were allowed to react at room temperature for two hours. After the labelling reaction completed, samples mixed, then evaporated to dryness in a Speedyvap centrifugal vacuum dryer (Eppendorf, UK), followed by resuspension in 100 µL of 50 mM Tris-HCl pH 8.5, 0.1% w/v SDS. After resuspension, the mixed, labelled sample was desalting on a protein desalting spin column (Pierce, UK), followed by two dimensional protein separation, as described in sections 2.3.2 and 2.4. The separated protein samples were collected in 384 well plates and

subjected to tryptic digestion, followed by direct spotting of the digested samples on to MALDI target plates for MALDI MS analysis (see section 2.6.1).

6.3.1.2.3. Dimethyl Labelling

The dimethyl labelling protocol was adapted from [229], and was performed by Fahsai Kantawong using 6 protein mixes A and B prepared by the author. 90 µL of 100 mM sodium acetate was added to each 10 µL aliquot containing 100 µg of protein. 10µL of 4% formaldehyde was added to the control sample (6 protein mix A), and 10 µL of 4% D_2 labelled formaldehyde was added to the test sample (6 protein mix B). 10 µL of 1 M sodium cyanoborohydride was added to each sample, followed by vortexing and incubation at room temperature for 5 minutes. Labelled proteins were pooled together to form a differentially labelled mixture. A complex separation was not performed on the mixture, because the experiment was performed simply to test labelling efficiency, and thus, 4 µg of trypsin in 25 mM ammonium bicarbonate was added to the sample for digestion, and the sample was analysed as described in section 2.6.3.

6.3.1.2.4. Data Analysis

Data analysis for isotopic labelling was performed by one of two methods, WarpLC 1.1 (Bruker Daltonics) or ProteinPilot (Applied Biosystems), depending on the instrument that was used to collect the data: WarpLC was used for data obtained from the HCT Ultra, while ProteinPilot was used to quantitate data from the 4700 Proteomics Analyser and QStar Pulsar i,. Although Mascot searches are an option with ProteinPilot, quantitation cannot be performed in this mode, and therefore the Paragon algorithm must be used to generate quantitative data.

6.4. Results and Discussion

6.4.1.1. UV Mapping

Separation of intact proteins by LC provides many of the benefits of 2DGE, such as basic quantitation using protein intensity imaging. UV maps can be obtained from the reversed phase separations and utilized for 2D-gel-like relative quantitation (see Figure 73). The DIA software from the DeCyder suite (GE Healthcare) was employed to perform spot detection and ratio of abundance calculation (see Figure 74) on wild type and pentamidine resistant *Leishmania*. The gel appears streaky in the vertical dimension due to the low

number (16) of ion exchange fractions used. This could be improved by increasing the number of ion exchange fractions to closer match the resolution of the column, but with a corresponding increase in analysis time. Spot detection was set to 500 to prevent merging of the more intense spots, although this excludes a number of the low intensity spots visible with increased contrast. Most spots detected were within the standard 1.5-fold abundance threshold used for single DiGE gels, implying that most of the proteins are of equal concentrations. Analysis of the complete plate of pentamidine-resistant *Leishmania* resulted in 499 unique protein identifications, the more clearly distinguishable of which are shown in Table 14. Many spots which had high scores did not match clear spots in the UV, and conversely, some of the spots shown in the table have surprisingly low scores, given the intensity of the spot.

Figure 73: Pseudogel generated from UV absorbances detected at 214nm during a two dimensional protein separation of pentamidine resistant *Leishmania.* Reversed phase chromatography is plotted on the x-axis, and the y axis describes ion exchange fraction number. Overall intensity (corresponding to the level of UV absorbance) is relatively low to avoid saturation of the abundant spots, but it is clear that there is high resolution separation in the reversed phase dimension with many distinct spots. The ion exchange dimension is streaky due to the low number of fractions obtained, and can be improved by obtaining additional fractions. The gel was further analysed using 2D gel quantitation software, as shown in Figure 74.

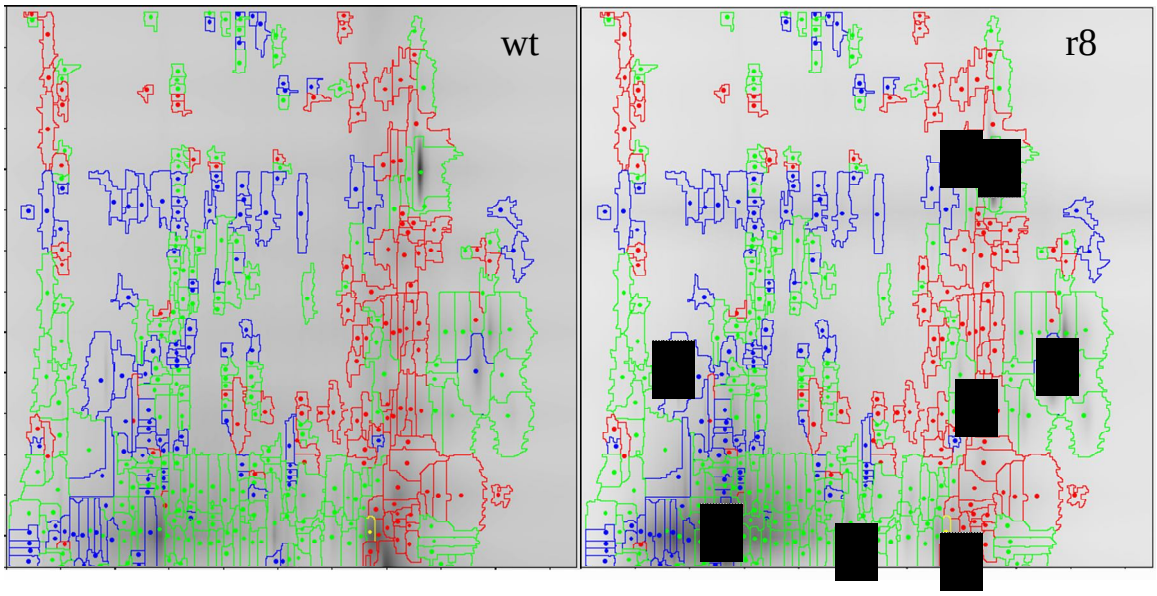


Figure 74: UV (214nm absorbance) pseudogel map of wild type (left) and pentamidine resistant (R8) *Leishmania donovani* analysed using the DeCyder DIA software (GE Healthcare). Spot detection software was set to detect 500 spots. 390 spots were detected, with 100 decreasing and 102 increasing in intensity in pentamidine resistant *Leishmania.* Reversed phase chromatography is plotted on the x-axis, and the y axis describes ion exchange fraction number. The numbers on the right hand gel map correspond to protein IDs listed in Table 14.

Table 14: Selection of protein identifications, matched to spots from the UV quantitation. Volume ratios of wild type vs pentamidine resistant (R8) are shown for the selected spots. The ratio marked with a * describes a peak that was not recognized as a spot by the DeCyder software. 'Cov.' refers to the sequence coverage of the identified peptides, while 'Peps' refers to the number of peptides observed.

Spot Number	Volume Ratio (WT/R8)	Accession	Protein Name	MW	pI	Cov.	Peps	Score
1	6.32	LinJ32.3460	nucleoside diphosphate kinase b	16613	7.78	19%	4	109
2	*	LinJ32.3460	nucleoside diphosphate kinase b	16613	7.78	11%	2	79
3	1.98	LinJ28.2920	Heat-shock protein hsp70, putative	71224	5.35	6%	3	252
4	-6.1	LinJ31.0010	5-methyltetrahydropteroyltriglutamate--homocysteinemethyltransferase, putative	85875	6.04	7%	5	360
5	1.66	LinJ13.0090	carboxypeptidase, putative	57101	5.48	5%	2	104
6	1.87	LinJ36.2400	xylulokinase, putative	53885	8.16	5%	4	34
7	1.61	LinJ09.0970	calmodulin, putative	16814	4.1	22%	5	185
8	1.07	LinJ08.0110	hypothetical protein, unknown function	69111	5.8	6%	3	41

Pentamidine resistance results in a number of physiological changes to *Leishmania*, some of which can be identified by the use of this technique. Nucleoside diphosphate kinase B (NDK-B) catalyses the phosphorylation of diphosphorylated nucleosides [230]. The strong upregulation (6.32-fold) of NDK-B observed in this study has also been identified in DiGE experiments (R. Burchmore, personal communication). NDK-B is involved in maintaining an appropriate balance between triphosphorylated and diphosphorylated nucleosides in the cell and especially the mitochondria. Pentamidine is an anti-parasitic drug that accumulates in the mitochondria, and resistance to the drug in *Leishmania* is believed to be mediated by modification to the mitochondrial membrane, which results in decreased uptake of the drug

[231]. The significant upregulation seen in this study could be related to an attempt to maintain essential mitochondrial function in the face of decreased membrane potential.

Heat shock protein 70 is a chaperone protein involved in maintenance of correct protein folding. Searle et al. [232] describe the characterization of the gene using bioinformatics, and describe the presence of calmodulin and ATP binding domains. An increase in expression (1.98-fold) of this protein may be related to the increases in abundance for calmodulin and nucleoside diphosphate kinase B. Additionally, the modification to the mitochondria may result in a decrease in metabolic rate and increased cell stress, which would lead to a requirement for improved maintenance of the cell's protein repertoire.
Calmodulin is a calcium binding protein involved in many cellular processes that was shown to be increased in abundance by 1.61-fold in pentamidine resistant *Leishmania*. Searle and Smith [232] recognized a calmodulin-binding domain in HSP70 in the closely related *Leishmania major*, and upregulation of this protein, may be in association with increased stress response and the requirement to trigger multiple downstream effects.

It should be noted that the up- and downregulation of the proteins described is recognized on the basis of a single experiment, and should not be considered clear biological evidence of phenotypic change. Instead, they serve as an indication of the type of data obtained from the quantitative analysis of proteins using intact protein liquid chromatography. Further, higher resolution replicates should be applied to more rigorously identify protein changes in differing organisms.

6.4.1.2. Isotopic labeling methodologies.

6.4.1.2.1. ICPL

ICPL is a non-isobaric lysine label specifically intended for intact protein labelling. Despite this, after following the provided protocol with the supplied protein standards, only 60% efficient labelling was observed across the gamut of peptides identified via tandem MS.

An example of the labelling efficiency is shown in Figure 75. The peptide, identified as having a sequence of VTEQESKPVQMMYQIGLFR was chosen due to the presense of a proline residue immediately following the lysine residue, which in both labelled and

unlabelled peptides blocks tryptic digestion. Lack of cleavage at the lysine precludes the unlabelled peptide being digested into two smaller peptides, making comparison of ion abundances easier. This leads to an unlabelled peptide size of 2284.26Da, and labelled peptide sizes of 2395.44 and 2389.44. The integrated peak areas of the diprotonated peptide are:

- 1143.13 (unlabelled) --- 1.0×10^6 counts
- 1198.72 (heavy labelled) --- 1.8×10^6 counts
- 1195.72 (light labelled) --- 3.5×10^5 counts

Total labeled peptide counts are therefore: 2.1×10^6 of 3.2×10^6 total counts, or 67%. Although the ionization potential of the peptide is affected by the derivitisation of the arginine residue, the size of the peptide should minimize the effects, and even taking this into account a significant amount of the peptide remains clearly unlabelled. A later peak with a mass close enough to 1143.13 to be included in the extracted ion chromatogram is unrelated to the three described.

Interestingly, the labelling significantly affects the retention of the peptide on the reversed phase column used to perform the separation, with a chromatographic shift of 7 minutes. The ICPL reagent is the most expensive available for the number of reactions, and despite a number of further attempts to perform the reaction, improved labeling efficiency was not observed. It was therefore decided to investigate other methods.

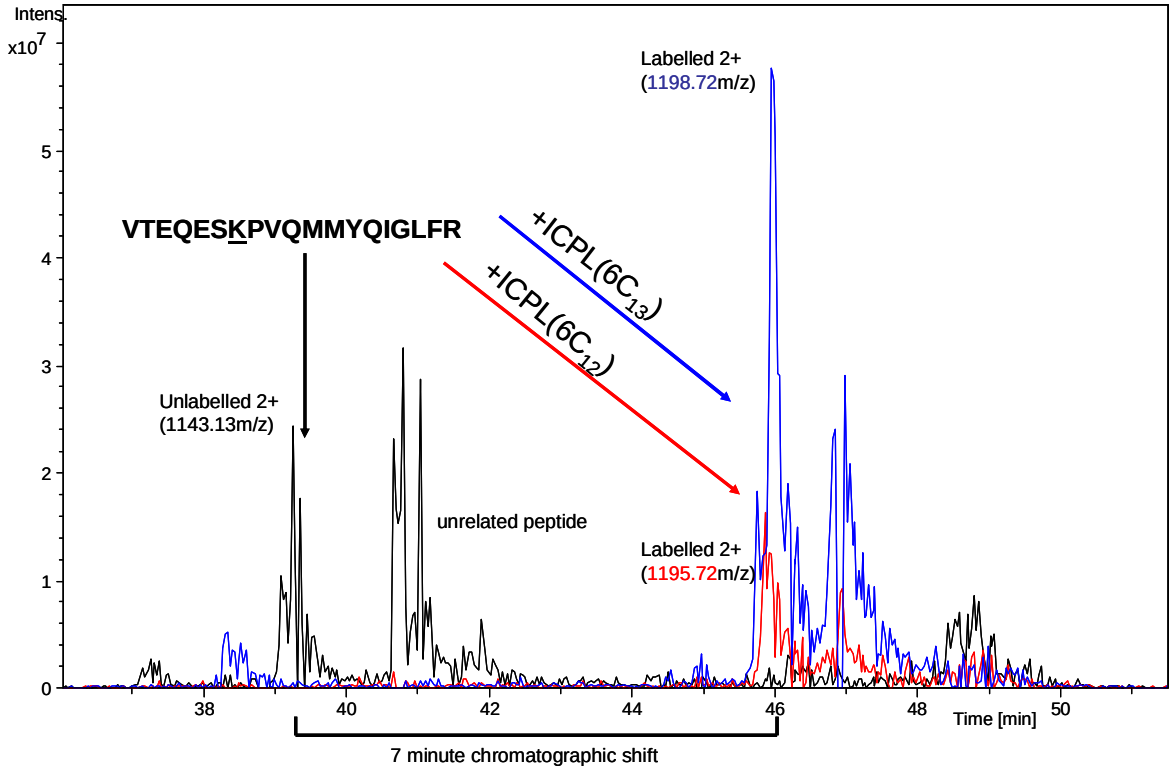


Figure 75: ICPL labelling of an example ovalbumin peptide. The peptide VTEQESKPVQMMYQIGLFR was inefficiently labelled in the test mix, the unlabelled peptide's abundance (at 1143.13) is slightly higher than the 'light' labelled peptide, and less than half the abundance of the 'heavy' labelled peptide.

ICPL was originally described in 2005 by Schmidt, Kellermann and Lottspeich [233]. In the original paper, the reagent is described as a nicotinoyloxy-succinamide ester labelled with either 4 deuterium atoms, or 4 hydrogen atoms for heavy and light labels respectively. The commercial version of the labelling reagent uses ^{13}C isotopes, rather than deuterium, and has a larger mass shift – 6 Da rather than 4. Efficient labelling was shown with myoglobin after a two hour reaction, and it is therefore surprising that poor results were obtained in this study. An interesting feature of the reagent is the substantial shift to acid pI resulting from the replacement of the amine groups of the protein with a non-polar species, which results in a marked change in 2D gels, and will also affect retention on ion exchange columns.

6.4.1.2.2. ICAT

ICAT is inherently an intact protein labelling solution, and therefore less substantial modification to the protocol was required for compatibility with the two dimensional LC separation process.

ICAT was originally designed as a biotinylated peptide with a deuterated linker resulting in a mass shift of 8 Da from the undeuterated (light) reagent. Other than small enhancements, such as replacing the 8 deuterium atom label with 9 C13 atoms and adding an acid cleaveable linkage to allow removal of the biotin after labelling, little difference may be observed between the commercially available and originally published reagents. Gygi [142] describes a 70% recovery of labelled peptides from the initial sample. Some sample losses are inevitable in any complex proteomic workflow, but it is unclear whether the losses are due to imperfect binding and release from the streptavidin column used for affinity purification of labelled peptides, or whether substantial amounts of the initial proteins are left unlabelled under the standard conditions.

The most significant modification to the protocol performed in this work was the removal of the affinity purification step – thus the general workflow consisted of reduction of cysteines, labelling with the cICAT reagent, mixture of the two samples, cleavage with TFA, neutralisation of the solution with NaOH, dessication in a centrifugal vacuum dryer, resuspension in ion exchange compatible buffer and separation.

ICAT was designed as both a quantitation methodology and a means of reducing the complexity of proteomic samples. The key feature allowing the complexity reduction is the addition of a biotin tag to the peptide tag, allowing labelled species to be purified using a streptavidin column. Purification does not strictly require digestion of the sample prior to purification, but would only serve as a means of excluding the small number of cysteine free proteins and any unlabelled proteins, which, in the latter case, should be negligible in a good experiment. When applied to a protein digest, the purified sample would contain only cysteine labelled peptides, significantly reducing its complexity. The obvious drawback of this is that identification of a protein relies on acquiring high quality tandem MS data from a far smaller number of peptides, and that because separation is performed post digestion the technique is clearly unsuitable as an intact protein labelling methodology. By eliminating the streptavidin purification step and the digestion and simply labelling the proteins then cleaving the biotin from the sample, it was hoped that identification could be enhanced with the use of the non-cysteine labelled peptides, while quantitation could be achieved with the labelled peptides. The use of mass spectrometer parameters available on the HCT Ultra that preferentially target isotopically labelled pairs was implemented for all separations.

When applied to a complex mixture (lysates comprising wild-type and glucose transporter knockout *Leishmania mexicana*) separated in two intact protein dimensions, followed by digestion and peptide analysis, no clear, manually verifiable, labelled identifications were observed, despite high intensity precursors observable in the ion chromatograms. This highlights an issue with complex lysates, intact protein separation and cysteine labels such as ICAT: labelled peptides are rare occurances among the large number of unlabelled, non-cysteine containing peptides. Since the MS was set to preferentially fragment pairs separated by 9 Daltons, and the resolution of the instrument is poor, misassignments were common, and the spectra obtained from correct pairings were predominantly from low abundance peptides that were insufficiently intense to provide sequence-assignable fragmentation data.

To generate information about efficiency of labelling and to develop the methodology, the technique was applied to two quantitatively different six protein mixes, which were separated by a single dimension of monolithic reversed phase protein separation. Although Mascot was capable of identifying several ICAT labelled proteins, the resolution and charge assignment was insufficient to correctly select several pairs of labelled peptides, such that the WarpLC 1.1 software (Bruker Daltonics) was unable to provide any quantitative information. In the example shown in Figure 76, the GLVLIAFSQYLQQCPFDEHVK peptide appeared at 891.5 m/z. It was selected as the light labelled partner to the 896.1m/z peak, corresponding to a 4.5m/z shift, equal to 9Da as both peptides were assigned a 2+ charge state. Unfortunately in this case, the 891.5m/z peptide is triply charged, and is the heavy labelled partner to the more clearly triply charged 888.5m/z peak. An additional, unexpected problem with the method is a result of the biotinylation. In common with the standard methodology, there is no removal of the biotin tag following its acid cleavage. This forms the largest peak in the total ion chromatogram, and is present from then on throughout the rest of the separation, until the high organic wash removes the biotin from the column (see Figure 77 and Figure 78). This poses a problem for the ion trap as there is a limit to the ion density of a single peptide in the trap. The fill time is therefore attenuated to avoid saturation which leads to reduced accumulation of other ions.

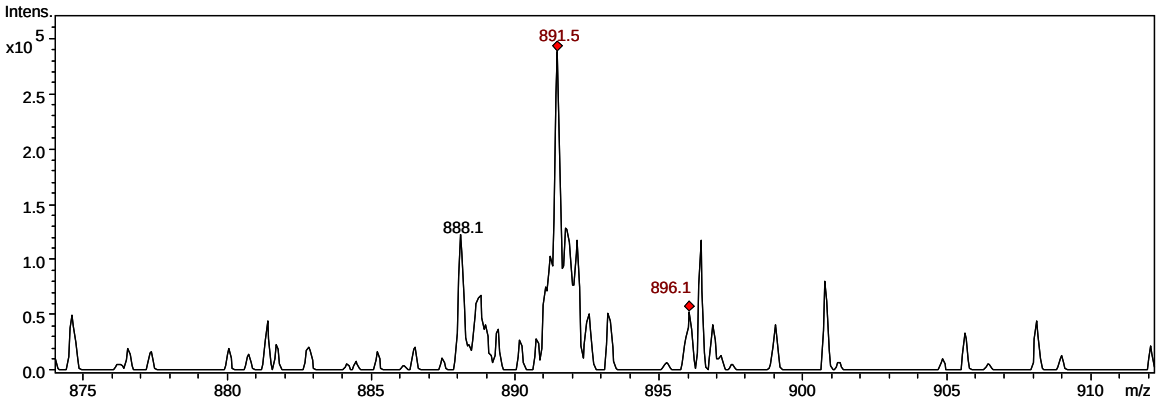


Figure 76: Illustration of the difficulties of using targetted fragmentation to enrich for analysis of ICAT labeled peptides. In this example, a heavy labeled BSA peptide, GLVLIAFSQYLQQCPFDEHVK with an m/z of 891.5, was incorrectly identified as matching the 896.1 peak (4.5m/z difference, corresponding to the difference between doubly charged peaks), when the peak, which is poorly resolved, is actually 3+ and matches to the 888.1 peak 3m/z below it.

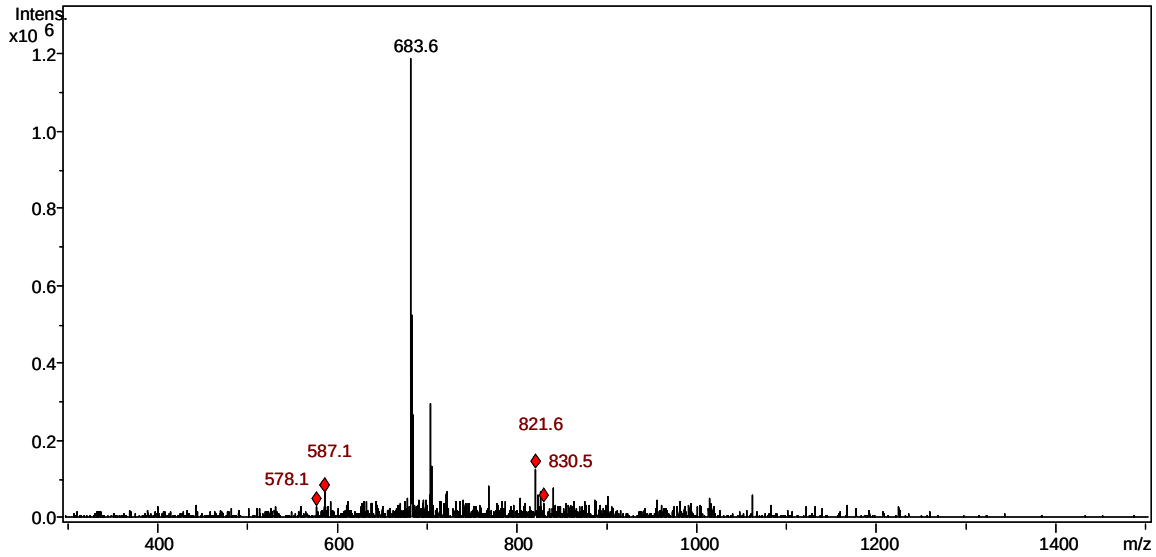


Figure 77: MS spectrum showing biotin contamination of the sample. The biotin plus part of the cleaved peptide linker corresponds to the highly abundant peak at 683.6m/z. Potential ICAT pairs (red) are consistently low in abundance, leading to poor identification of charge state and incorrect selection for fragmentation.

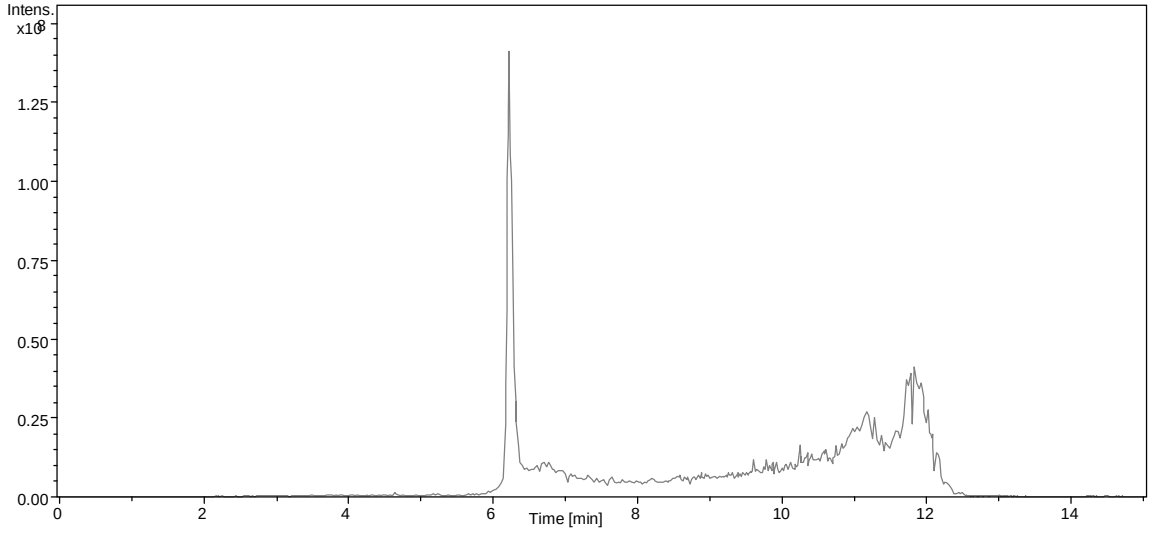


Figure 78: Extracted ion chromatogram of the biotin peak at 683.5m/z. Ion abundance reaches 10^8 in the most intense part of the peak, denoting an extremely concentrated species. Tailing is also observed throughout the remaining 5 minutes of the run.

Labelling efficiency is good, however, with any possible unlabelled peptide indistinguishable from noise. As an example, the molecular mass of the labelled peptide described earlier is 2670.39, with a corresponding unlabelled mass of 2434.24. The corresponding triply charged labelled and unlabelled masses equate to 891.5 and 812.41 respectively, and if extracted ion chromatograms are acquired for these, the peak corresponding to the labelled peptide is clearly visible, whereas the 812.51m/z unlabelled peptide is not seen to produce a distinct peak (see Figure 79).

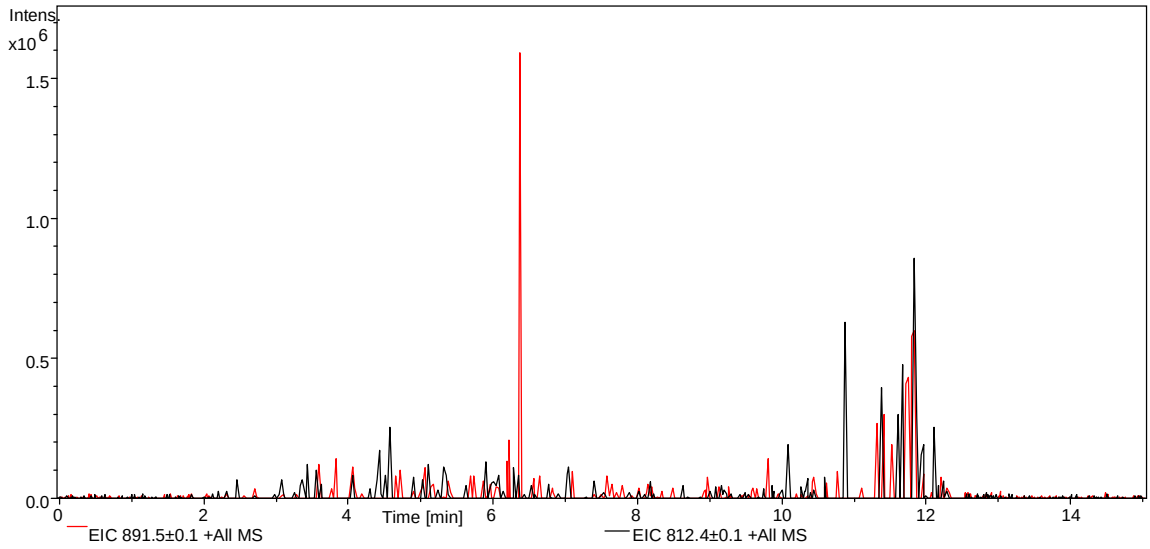


Figure 79: Comparison of extracted ion chromatograms from the peptide GLVLIAFSQYLQQCPFDEHVK, for which the triply charged ICAT labelled version (891.5) is shown in red, and the unlabelled triply charged unlabelled version (812.5) is shown in black. The abundant peak at 6.4mins corresponds to the point at which elution of the peptide occurred and it was selected for fragmentation. High intensity peaks from 11 to 12 minutes are a result of increased chemical noise during the acetonitrile wash.

The use of a higher resolution MS to provide more accurate identifications of labelled pairs, coupled with stricter ion abundance requirements prior to fragmentation may provide significantly better results, as would the removal of biotin prior to analysis. The latter could be easily performed by affinity purification with a streptavidin column, following acid cleavage. Collection of the flow-through would provide a biotin-free sample, ready for further analysis. However, the overall low number of quantitative species in a sample significantly reduces the viability of this method.

There are no specific examples of ICAT performed without affinity purification to provide comparison to the results described, but an effective use of the technology was described in [34]. ICAT was used to elucidate changes that occur to microsomes from naïve and differentiated human myeloid leukaemia cells (HL-60s). Standard protocols were used in the experiment, which used the older, non-cleaveable ICAT reagents, which would result in the identification of biotin-labelled peptides. 491 proteins were identified and quantified in 3, non-identical experiments, showing the utility of the technique. However, this is still poorer than many DiGE experiments, where thousands of spots are identified, although the total number of identifications is often lower due to the selection of only those spots consistently differentially regulated. It is to be expected that although quantitative data is obtained for all peptides identified (or at least the vast majority), the loss of information due to removal of all non-labelled peptides limits the use of the technique to specific samples rich in cysteine residues.

ICAT has also been applied by Leifso et al. [184] to quantify the differences in proteomic expression between *Leishmania* promastigote and amastigote life cycle stages, and is described in further detail in section 3.2.3.11.2.

6.4.1.2.3. iTRAQ

The standard iTRAQ protocol is intended for peptides, and some steps in the protocol, most obviously the pre-labelling digestion, but notably also the reaction solution consisting of 70% v/v ethanol, pose a problem for its application to proteins. The latter is problematic primarily because high percentage organic conditions are an ideal method of precipitating proteins.

The ITRAQ reagents, as originally described by Ross et al. [147] are a family of methylpiperazine acetic acid N-hydroxy succinimide esters modified to provide an isobaric

mass in MS mode but fragment into four different reporter ions (one per label) when collision-induced dissociation is applied. Reporter ions have masses of 114-117 and the balancers are 28-31 daltons in mass, followed by the NHS ester. They are therefore relatively small tags for peptides.

Several different modifications to the standard protocol were attempted with the ITRAQ reagent before a successful result was obtained. The initial experiments involved solubilising the reagent with dimethyl formamide (DMF) prior to addition to the protein sample, because this solvent has a significantly higher protein dissolution capacity than ethanol. This was attempted with a standard mixture of protein, and although some labelling was observed, the majority of protein precipitated on addition of the reagent. Analysis of the supernatant and the precipitant revealed partial labelling in the former case, and no labelling in the latter.

Follow up attempts gave interesting results. Samples were labeled with reagent dissolved in DMF and more vigorous attempts (the two hour labeling reaction was performed in a sonicating water bath) were made to reduce precipitation of the protein. After the reaction, the samples were frozen at -20ºC overnight, prior to tryptic digestion and analysis by MS. Mascot searches of the data show extensive labeling of trypsin, suggesting that the reaction continued well into the tryptic digestion. This implies that the reaction proceeds when the reagent is dissolved in DMF, but at an extremely slow rate.

Weise [234] describes a protocol for intact protein labelling using iTRAQ reagents, but the critical step (the addition of the reagent and the solvent in which it was dissolved) is not described – merely that the reagent was added. This implies that the protein solution was added directly to the dry reagent. NHS esters have a half-life in water measured in seconds, making the labelling reaction dependant on extremely rapid mixing. The development of a more robust method was therefore attempted. A protocol provided by Applied Biosystems using 50% ethanol in the reaction mixture, and one quarter (25 µg) of the usual amount of protein per sample was evaluated. This protocol, when applied to a mixture of 6 proteins, separated by intact protein LC, digested and analysed by MALDI ToF MS (see section 2.6.1) demonstrated extremely efficient labelling (see Figure 80, Figure 81 and Figure 82).

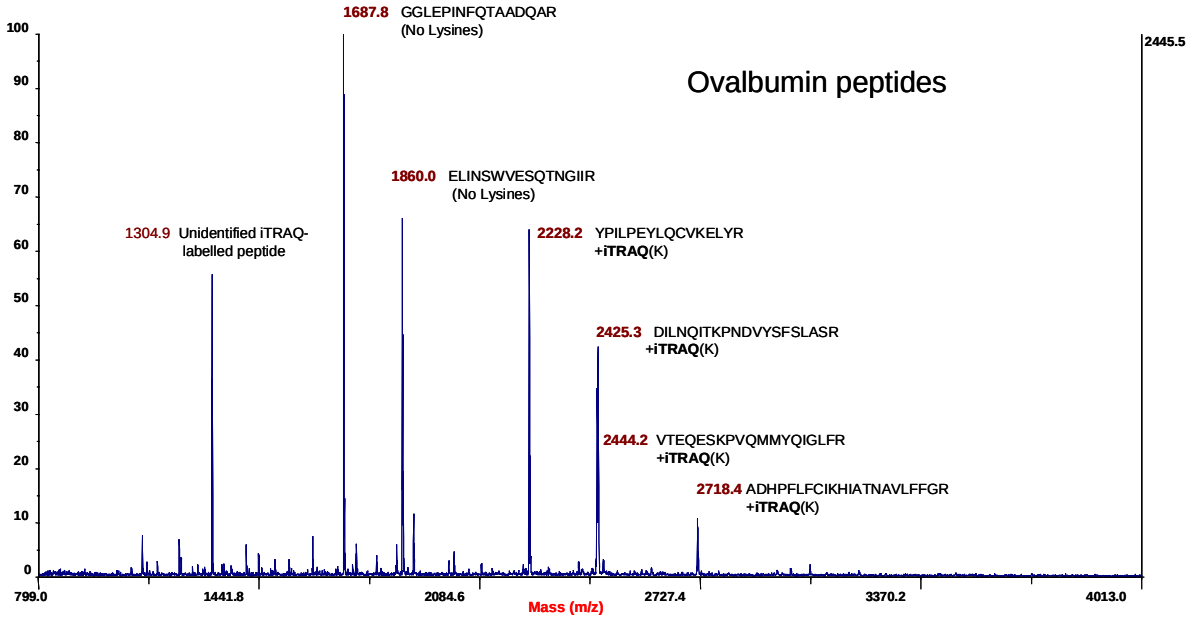


Figure 80: A single MALDI spot containing multiple iTRAQ labeled peptides from ovalbumin. Two peptides (those of highest abundance) contain no lysines, and are therefore unable to be labeled. All of the other intense peaks in the spectrum were manually and computationally confirmed to be labeled.

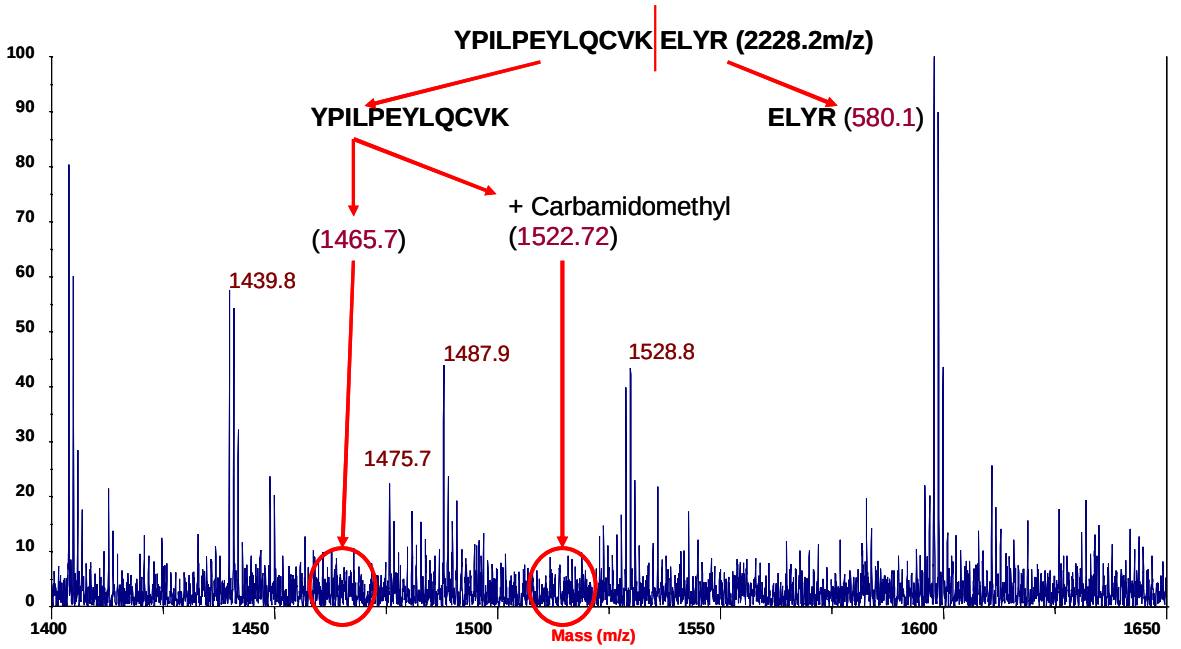


Figure 81: Tryptic digestion of the iTRAQ labeled peptide YPILPEYLQCVKELYR (mass 2228.2) is inhibited at the position 12 lysine. Unlabelled peptide will correctly cleave at the KE peptide bond, forming two fragments of masses 1465.7m/z and 580.1m/z. Although the latter is below the low-mass cut-off of the instrument, the former will be seen, if extant. Due to the iodoacetamide treatment performed during the reaction, the alkylated cysteine will push the peptide's mass to 1522.7m/z. Neither the 1465.7 or 1522.7 masses is observable above the noise level of the instrument, whereas the labelled peptide's signal to noise is 801/1 showing a very high level of labelling efficiency.

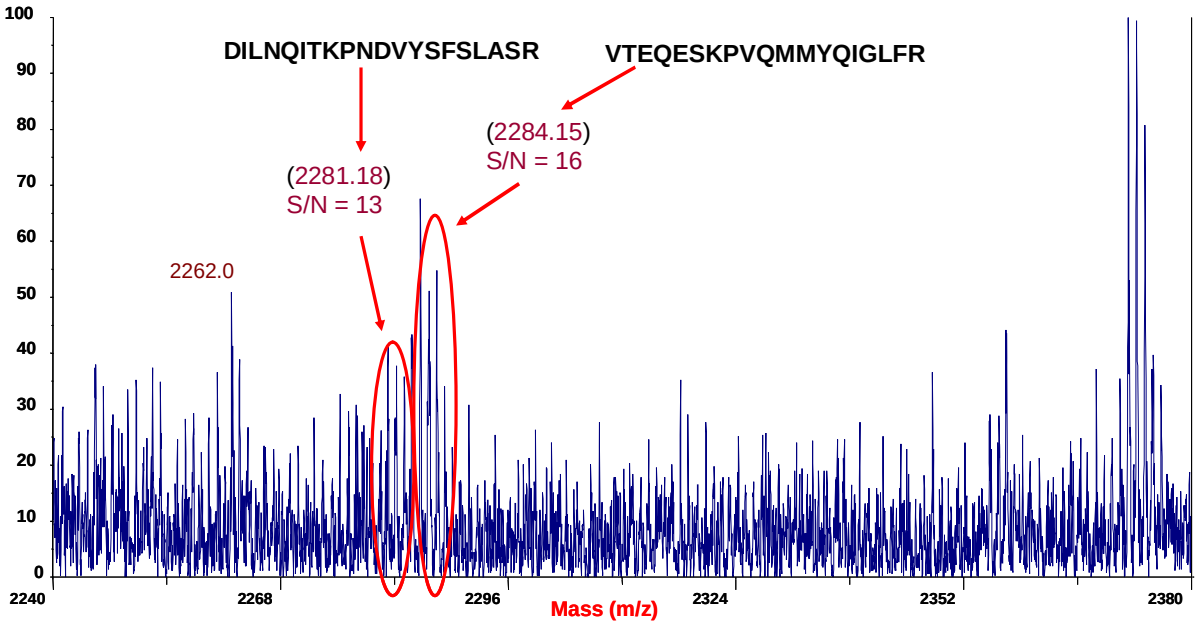


Figure 82: An example of slightly less successful labeling efficiency. The iTRAQ labeled peptides DILNQITKPNDVYSFSLASR (2425.3m/z) and VTEQESKPVQMMYQIGLFR (2444.2m/z) both contain lysine residues, but these are masked by the adjacent proline residues. The unlabelled peptides are both present, but at very low signal to noise (13/1 and 16/1 respectively, compared to 448/1 and 49/1 respectively for their labelled equivalents).

The peptide modification due to the 4-plex iTRAQ reagent has a mass of 144, which is small enough to cause few problems for analysis as there is little significant mass shift due to labelling. Two problems do exist for mass spectral analysis, however. One is common to peptide iTRAQ labelling – that the conditions for good quality sequencing fragmentation are different from those that emphasise the reporter ion peaks, and the second is that iTRAQ labelling of intact peptides precludes digestion at those points, resulting in cleavage only at arginine residues.

A number of different strategies have been proposed to deal with the problem of setting appropriate collision energies that provide good sequence and good fragmentation. A very effective method, pioneered by Smith [235] relies on high-energy fragmentation of precursors to provide the highest quality quantitation. From this data, an 'include list' of target peptides with good quantitation is prepared, and the MS is optimised for good quality sequence data from these peptides. Although this is a highly effective method for analysis it requires double the sample, and takes twice as long. An alternative approach, is to slightly increase the standard collision energies for unknown, 2+, 3+, 4+ and 5+ charge state precursors, which provides a balance of sequence data and quantitation data. Alternatively, a MALDI ToF/ToF based method can be used, which was the method decided on for this study. Due to the high energy (1 keV) of ions entering the collision cell in the 4700 Proteomics Analyser, fragmentation is of higher energy than the equivalent on

ESI-MS instruments, producing not only good quality quantitative peaks, but also internal fragments which may provide enhanced sequence data.

Missed cleavage of peptides due to label-induced blocking of trypsin activity can be a drawback when samples are analysed by MALDI instruments, due to the monoprotonated charge state of the majority of ions. Monoprotonation of a larger peptide may result in a mass to charge ratio outside the mass range of the MS. However, some benefit may be obtained from identification of smaller peptides concatenated due to missed cleavage, which are shifted above the lower mass range of the instrument. When samples labelled using the intact protein methodology are analysed by electrospray instruments, the same increase in predicted peptide size due to missed cleavage results in an increase in charge state. The resulting ions may be too highly charged for isotopic resolution of the peak on some ESI instruments, which leads to misassignment of charge state.

Labelling of the same lysates studied in the ICAT experiment (wild type and glucose transporter knockout *Leishmania donovani*) failed to produce more than two verifiable protein identifications, despite the existence of a number of clearly labelled peaks with excellent fragmentation data, using either Mascot (Matrix Science), or the ProteinPilot software (Applied Biosystems) (see Figure 83, Figure 84 and Figure 85).

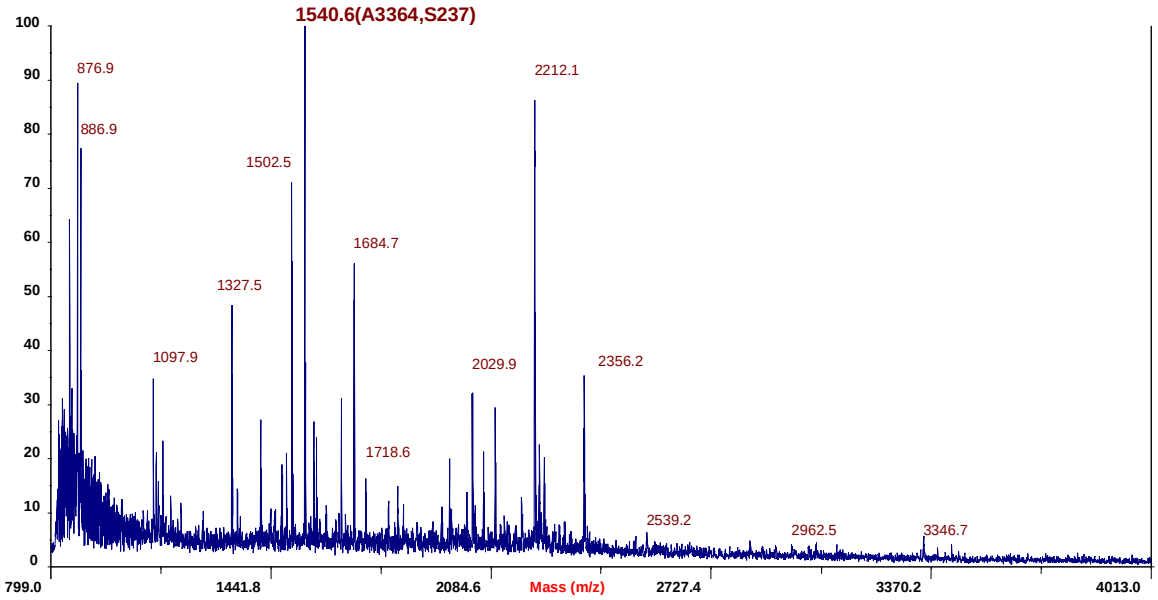


Figure 83: MS spectrum from a complex MALDI spot that nevertheless failed to provide any identifications, despite the high signal-to-noise of most peaks. The 1540.6 peak, in particular, has excellent signal-to-noise (237 to 1), yet, as shown in figs 7 and 8 was not able to be sequenced.

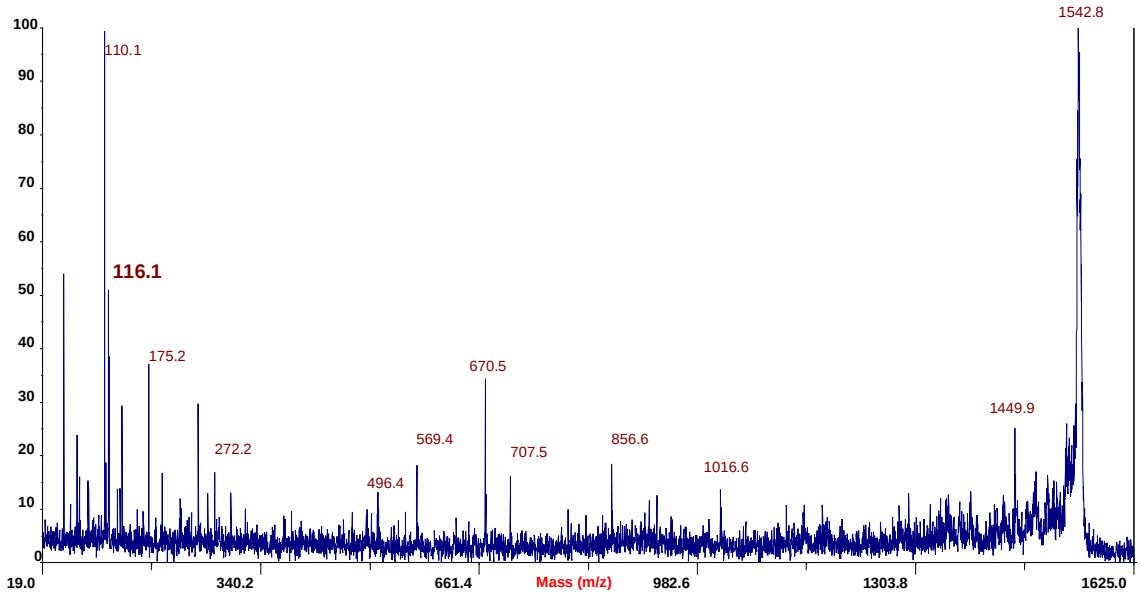


Figure 84: Fragmentation spectrum of an unidentified protein, derived from two dimensional separation of an iTRAQ labeled *Leishmania* lysate. The peptide is clearly labeled the iTRAQ signal ion is shown in boldface, the 117 peak is too close to 116 to be displayed in this image).

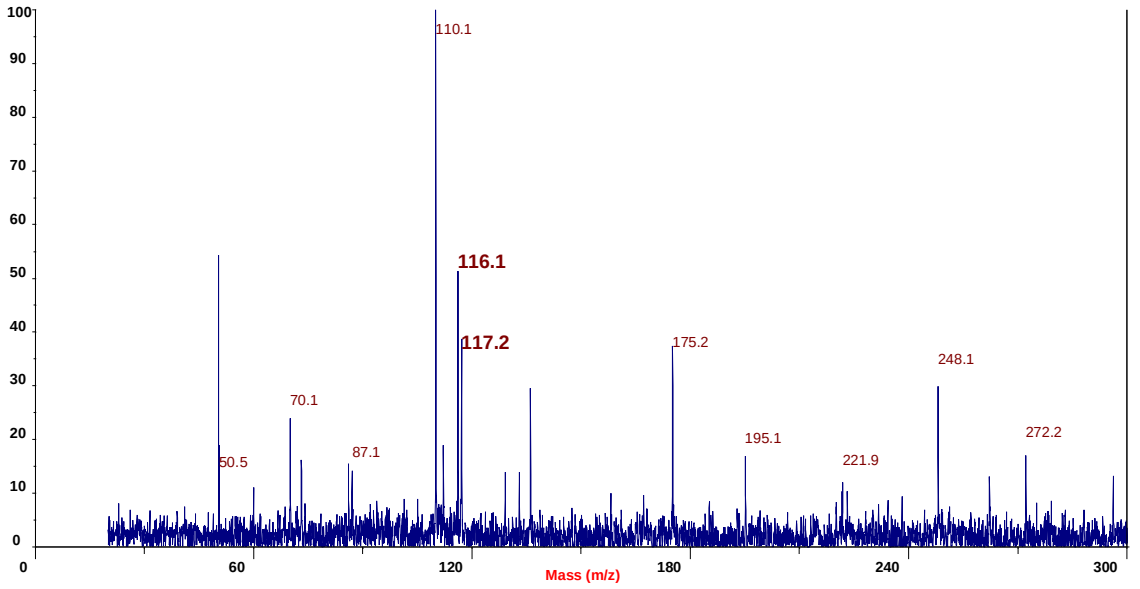


Figure 85: Closer analysis of the peptide shown in figure 6 demonstrates more clearly the labeling of the peptide, with distinct 116 and 117 peaks corresponding to the chosen reporter ions for this experiment.

The two proteins identified with confidence were kinetoplastid membrane protein and ubiquitin. Kinetoplastid membrane protein is downregulated 37% in the glucose-transporter knockout line. Since the function of the protein is suggested as being regulation of lipid bilayer pressure [236], it is likely that the reduction in the total number of membrane proteins resulting from the knockout of the family of glucose transporters

results in changes to the pressure, and less KMP is required to maintain turgor. The latter, ubiquitin, is a signaling molecule that targets proteins for degradation in the proteasome. It is downregulated 28% in the glucose-transporter knockout line, suggesting that the turnover of proteins is decreased due to a lower metabolic rate resulting from inability to uptake glucose. Due to the fact that these were the only two proteins confidently identified, it is difficult to ascertain whether the values have been normalised to a global standard of all reporter ions identified, as some difference in protein amounts between samples is inevitable, although ProteinPilot provides no information on normalisation.

iTRAQ has been applied to *Leishmania donovani* at a peptide level by Rozensweig [237], to track proteomic changes during chemically-stimulated differentiation from the promastigote to amastigote life cycle stages. The experiment was successful in identifying 20.9% (1713) proteins by combining data from both experimental replicates. 969 proteins were identified at all 7 time points. The 4-plex reagents were employed, and unstimulated promastigotes were used as a standard to correlate two 4-plex experiments of three time points each. An experimental replicate of the entire experiment was performed, and the total number of identified proteins common to both analyses was 931 (54% of the original 1713). This level of reproducibility highlights the growing importance of experimental replicates in large scale proteomics experiments, but, as the data in this paper (published in 2008) shows, using multiple replicates to improve data stringency is still uncommon.

6.4.1.2.4. SILAC

The SILAC labeling procedure is entirely unaffected by the separation methodology used to analyse the samples, as the proteins are labeled in culture, therefore no modifications to the protocol were expected or needed for analysis. Samples analysed consisted of duplex lysine labeled K562 lysates prepared by Christian Preisinger, and triplex arginine labeled PC12 lysates prepared by Alex von Kriegsheim. Samples were prepared as described in 2.2.3 and 2.2.4 respectively, and general introductions to the cell types and experimental design are given in 1.8.3 (for K562 cells) and 1.8.4 (for PC12 cells).

Efficiency of labelling is dependant on the quality of the culture medium, and as high quality stocks are now commercially available, this is not subject to user skill, unlike post-lysis chemical labelling methods. The two sets of SILAC cells were analysed by two different methods. The PC12s were separated using ion exchange, followed by 32 fractions being selected for reversed phase analysis and collection into two 384 well plates, for a total number of fractions equal to 768. Each plate was subjected to digestion using a

TECAN Genesis workstation, and plates were rapidly analysed by parallel monolithic separation into the HCT Ultra (15 minutes per run). Data was collated and analysed by an AutoIT script, which automated the WARPLC 1.1 software, and further parsed using PERL scripts which summarised the results for each protein. This resulted in a list of 292 quantitatively analysed proteins with scores above Mascot's significance threshold for the SWISSPROT database.

The K562 cells were subjected to a more significantly high-resolution separation, into 64 fractions, each of which was separated again by monolithic reversed phase. Four 384 well plates were used to collect the fractions, for a total of 1152 fractions. A TECAN genesis workstation was programmed to automatically digest and spot each plate onto two MALDI target plates, suitable for analysis with the 4700 proteomics analyser. Each plate was analysed by tandem ToF/ToF MS. The data was analysed using the ProteinPilot software and the Paragon algorithm, which resulted in quantitative data for 53 different proteins.

Most proteins identified using the Paragon algorithm were not significantly differentially regulated, a reassuring observation based on the short treatment time of the cells. Those proteins with differential regulation determined to be at two-fold or greater are shown in Table 15 and Table 16.

Table 15: This table shows significantly upregulated proteins identified from ToF/ToF analysis of a 2D intact protein separation of SILAC labeled K562 cells. Labels were applied to untreated, and K562 cells after 6 hours of treatment with Glivec (see section 1.8.3). Results are from the ProteinPilot software (Applied Biosystems, Warrington, UK). N is the rank of the peptide in the total number of identifications. Unused and Total are a measure of the confidence of the identification (Total may score higher than Unused if multiple proteins match the same peptides). A score of 2 or greater in Unused or Total denotes a 95% confident identification.

N	Unused	Total	% Cov	Accession #	Name	H:L
13	4	4	15.7	O95881|TXD12_HUMAN	Thioredoxin domain-containing protein 12 precursor (EC 1.8.4.2) (Thioredoxin-like protein p19) (Endoplasmic reticulum protein ERp19) (ERp18) (hTLP19)	0.1135
27	2	2	11.4	P06702|S10A9_HUMAN	Protein S100-A9 (S100 calcium-binding protein A9) (Calgranulin-B) (Migration inhibitory factor-related protein 14) (MRP-14) (P14) (Leukocyte L1 complex heavy chain) (Calprotectin L1H subunit)	0.2626
9	4	4	3.5	P49588|SYAC_HUMAN	Alanyl-tRNA synthetase, cytoplasmic (EC 6.1.1.7) (Alanine--tRNA ligase) (AlaRS) (Renal carcinoma antigen NY-REN-42)	0.4094

Table 16: This table shows significantly downregulated proteins identified from ToF/ToF analysis of a 2D intact protein separation of SILAC labeled K562 cells. Labels were applied to untreated, and K562 Cells after 6 hours of treatment with Glivec (see section 1.8.3). N is the rank of the peptide in the total number of identifications. Unused and Total are a measure of the confidence of the identification (Total may score higher than Unused if multiple proteins match the same peptides). A score of 2 or greater in Unused or Total denotes a 95% confident identification.

N	Unused	Total	% Cov	Accession #	Name	H:L

11	2.55	2.55	21.2	P09651|ROA1_HUMAN	Heterogeneous nuclear ribonucleoprotein A1 (Helix-destabilizing protein) (Single-strand RNA-binding protein) (hnRNP core protein A1)	2.0001
5	4.62	4.62	27.7	P46109|CRKL_HUMAN	Crk-like protein	2.0167

Interesting quantitative differences observed include increases in the abundance of several antioxidant proteins involved in regulation of the formation of disulphide bridges [238]. An example is Thioredoxin-like protein p19 which was found to be upregulated 9-fold in Glivec-treated K562 cells. This is likely to be in response to cellular stress as a result of the initiation of apoptosis. Alanyl-tRNA synthetase is upregulated significantly (2-fold) in Glivec-treated cells. Although the remaining tRNA ligases were not observed, it is likely that the upregulation of Alanine-tRNA synthase is a response to the induction of apoptosis pathways in drug-treated cells as the cell begins to express proteins involved in the shutdown of cellular processes.

The downregulation of CRK-L is interesting, due to the protein's association with the generation of macromolecular machinery at focal adhesions [239]. It is known that overexpression of CRK-L results in the transformation of cells, and downregulation of the protein in response to Glivec is likely to be associated with the treated cells' entry into apoptosis. Heterogeneous nuclear ribonucleoprotein A1 is responsible for shuttling mRNAs between the nucleus and cytoplasm [240]. Twofold downregulation of this protein is likely to be associated with its additional role in negatively regulating the translation of the subset of proteins induced using XIAP IRES elements, which are translated during apoptosis [241]. The S100 A9 calcium binding protein is upregulated fourfold under Glivec treatment of the cells. The S100 proteins are markers for the inflammatory response, and are released by phagocytes [242]. Expression of this protein is a marker of the cell progenitors of the K562s, as they are derived from leukocytes. The function of S100 proteins is to reduce the inflammatory response [243], likely in response to increased cell stress, and after the initiation of apoptosis.

The triple labeled SILAC sample analysed by third dimension peptide LC/MS/MS analysis resulted in the identification of 113 quantitative identifications for EGF treated samples, and 103 quantitative identifications for NGF-treated samples. Due to the single experiment performed, a cut-off of 2-fold was chosen as significant, as any smaller deviation is within the expected level of biological noise. Table 17, Table 18, Table 19 and Table 20 describe the differences equal to or greater than twofold in NGF and EGF determined in this experiment.

Table 17: Proteins upregulated in EGF-treated PC12 cells. This experiment applied triplex SILAC labeling to PC12 cells either untreated, treated with EGF or treated with NGF, with untreated peptide intensities applied as the denominator in all cases. Boldface figures are ratios of protein abundance in EGF stimulated PC12 cells compared to untreated PC12 cells. Italic figures are ratios of protein abundance in NGF stimulated PC12 cells compared to untreated PC12 cells and are included for comparison.

Well	Protein Name and Species	Accession	Score	MW [kDa]	pI	SC [%]	# Pept.	Avg. H/L	Avg. M/L
395	tropomyosin 5; TM-5 [Rattus sp.]	gi|9653293	347	29.158	4.6	23	6	**83.63**	*7.73*
370	secretogranin 2	gi|12083601	147	70.987	4.6	6	7	**25.66**	
394	rCG62531, isoform CRA_h	gi|149048018	161	17.556	4.6	17	4	**12.9**	*7.33*
346	Scg2 protein	gi|38181552	127	66.622	4.5	4	6	**2.93**	*12.98*
204	Rho GDP dissociation inhibitor (GDI) alpha	gi|31982030	108	23.450	5.0	15	3	**2.79**	*7.67*
447	PDGF associated protein	gi|1136586	119	20.523	6.3	16	2	**2.58**	*7.06*
156	ionized calcium binding adapter molecule 2	gi|21553105	128	17.070	7.5	17	4	**2.58**	*2.18*
35	PREDICTED: similar to polyubiquitin	gi|109497721	146	8.821	9.8	31	4	**2.27**	
157	poly(A) binding protein, cytoplasmic 1	gi|19705459	162	70.884	10.0	8	4	**2.25**	*5.41*

Table 18: Proteins downregulated in EGF-treated PC 12 cells. This experiment applied triplex SILAC labeling to PC12 cells either untreated, treated with EGF or treated with NGF, with untreated peptide intensities applied as the denominator in all cases. Boldface figures are ratios of protein abundance in EGF stimulated PC12 cells compared to untreated PC12 cells. Italic figures are ratios of protein abundance in NGF stimulated PC12 cells compared to untreated PC12 cells and are included for comparison.

Well	Protein Name and Species	Accession	Score	MW [kDa]	pI	SC [%]	# Pept.	Avg. H/L	Avg. M/L
205	D-dopachrome tautomerase	gi|13162287	154	13.239	6.1	21	3	**0.49**	*14.45*
322	stathmin 1	gi|8393696	223	17.278	5.7	30	6	**0.49**	*0.45*
395	rCG62531, isoform CRA_c	gi|149048013	250	33.488	4.6	16	5	**0.45**	*0.72*
419	nucleobindin 1	gi|16758210	349	53.474	4.9	29	13	**0.37**	*0.79*
300	hypothetical protein LOC683313	gi|155369696	163	59.555	8.9	5	3	**0.31**	*0.35*
36	rCG50668, isoform CRA_a	gi|149031956	134	25.532	10.2	24	5	**0.31**	*0.03*
347	Lysozyme homolog AT-2, bone - rat (fragments)	gi|539969	100	3.162	4.4	57	1	**0.26**	*0.8*
155	ATP synthase, H+ transporting, mitochondrial F0 complex, subunit F6	gi|16758388	117	12.487	9.9	26	4	**0.25**	*0.28*
10	polyubiquitin	gi|1050930	133	11.234	5.3	35	3	**0.25**	*0.25*
36	rCG33578	gi|149054207	185	35.220	4.6	17	4	**0.11**	*1.01*
36	rCG50690	gi|149031970	130	21.249	4.9	11	2	**0.07**	*0.03*

Stathmin 1 is found to be downregulated 2-fold in both EGF and NGF-treated cells. This is surprising, as phosphorylated forms of stathmin were found to be required for completion of mitosis [244]. It is described as an abundant protein, and the score of 223 reflects this, but the identification of the protein in only a single well, and the control of its regulation by an increase in phosphorylation may have resulted in the majority of the protein being shifted to another part of the separation, where identification was not successful. D-dopachrome tautomerase, a macrophage migration inhibitory factor protein was identified as being downregulated in EGF treated cells, but significantly upregulated in several wells under NGF treatment conditions. Calandra and Roger [245] describe MIF as a strong inhibitor of apoptosis, which does not match the observed changes, since proliferation is induced in EGF-treated cells, whereas NGF-treated cells are known to differentiate.

Perhaps, as a cytokine, the significant upregulation is a result of cell-cell communication required to set up neural connections. Tropomyosin is an actin-binding protein, and its regulation is associated with both proliferation and differentiation [246]. Significant upregulation was observed in both EGF and NGF-treated cells, which is in accordance with expectations, as modification of the actin cytoskeleton occurs in both proliferation and differentiation.

Secretogranin 2 (Scg 2) is a chromogranin protein that assists in the maintenance of osmotic equilibrium in chromaffin granules [247]. It is expressed in the adrenal medulla, which matches the origin of PC12 cells, and has been identified twice as being differentially regulated, once as being significantly upregulated in EGF cells, although the appropriate matching NGF stimulated cell peaks were not observed, and secondly as being upregulated threefold in EGF cells and 13-fold in NGF cells. The significant upregulation is likely to be due to changes in osmotic pressure in response to the cells' preparation to undergo mitosis. However, multiple identifications in other spots do not show significant upregulation which may be evidence of spurious quantitation, or that there is modification to a single isoform of the protein. Polyubiquitin is downregulated fourfold in both EGF and NGF-stimulated cells, which is likely to be a consequence of a general decrease in protein degradation resulting from preparation for significant protein expression changes. In addition to degradation, polyubiquitination is implicated in a number of other cellular processes, and downregulation may be related to the function of the multiple signaling cascades activated due to growth factor stimulation [248]. In general there are no obvious trends in the data showing up- or down-regulation, but slight upregulation in cytoskeletal and cytoskeletal-related proteins is observed, e.g. myosin (1.41 and 1.25) together with tropomyosin, is likely to be associated with the morphological changes, both due to mitosis and differentiation, that the cells are preparing to undergo. Many of the proteins identified and the patterns of quantitative differences observed are intriguing, and further work allowing appropriate replicates to be performed are appropriate.

Table 19: Proteins upregulated in NGF-treated PC12 cells. This experiment applied triplex SILAC labeling to PC12 cells either untreated, treated with EGF or treated with NGF, with untreated peptide intensities applied as the denominator in all cases. Boldface figures are ratios of protein abundance in NGF stimulated PC12 cells compared to untreated PC12 cells. Italic figures are ratios of protein abundance in EGF stimulated PC12 cells compared to untreated PC12 cells and are included for comparison.

Well	Protein Name and Species	Accession	Score	MW [kDa]	pI	SC [%]	# Pept.	Avg. H/L	Avg. M/L
253	D-dopachrome tautomerase	gi|13162287	113	13.239	6.1	47	3	*0.83*	**36.57**
35	polyubiquitin	gi|1050930	146	11.234	5.3	24	4	*1.46*	**33.3**
182	D-dopachrome tautomerase	gi|13162287	157	13.239	6.1	28	4	*0.86*	**17.5**
229	D-dopachrome tautomerase	gi|13162287	115	13.239	6.1	21	3	*1.62*	**16.47**

205	D-dopachrome tautomerase	gi|13162287	154	13.239	6.1	21	3	*0.49*	**14.45**
346	Scg2 protein	gi|38181552	127	66.622	4.5	4	6	*2.93*	**12.98**
395	tropomyosin 5; TM-5 [Rattus sp.]	gi|9653293	347	29.158	4.6	23	6	*83.63*	**7.73**
204	Rho GDP dissociation inhibitor (GDI) alpha	gi|31982030	108	23.450	5.0	15	3	*2.79*	**7.67**
394	rCG62531, isoform CRA_h	gi|149048018	161	17.556	4.6	17	4	*12.9*	**7.33**
447	PDGF associated protein	gi|1136586	119	20.523	6.3	16	2	*2.58*	**7.06**
157	poly(A) binding protein, cytoplasmic 1	gi|19705459	162	70.884	10.0	8	4	*2.25*	**5.41**
181	D-dopachrome tautomerase	gi|13162287	211	13.239	6.1	38	8	*1.94*	**3.61**
182	hypothetical protein LOC290230	gi|157823059	140	18.656	5.8	18	4	*0.862*	**3.02**
588	PREDICTED: similar to Huntingtin interacting protein K	gi|109468719	131	17.119	4.9	30	3	*1.97*	**3.01**
277	endoplasmic reticulum protein 29	gi|16758848	109	28.614	6.3	12	5	*1.81*	**2.34**
156	ionized calcium binding adapter molecule 2	gi|21553105	128	17.070	7.5	17	4	*2.58*	**2.18**

Table 20: Proteins downregulated in NGF-treated PC12 cells. This experiment applied triplex SILAC labeling to PC12 cells either untreated, treated with EGF or treated with NGF, with untreated peptide intensities applied as the denominator in all cases. Boldface figures are ratios of protein abundance in NGF stimulated PC12 cells compared to untreated PC12 cells. Italic figures are ratios of protein abundance in EGF stimulated PC12 cells compared to untreated PC12 cells and are included for comparison.

Well	Protein Name and Species	Accession	Score	MW [kDa]	pI	SC [%]	# Pept.	Avg. H/L	Avg. M/L
322	stathmin 1	gi|8393696	223	17.278	5.7	30	6	*0.49*	**0.45**
300	hypothetical protein LOC683313	gi|155369696	163	59.555	8.9	5	3	*0.31*	**0.35**
442	NSFL1 (p97) cofactor (p47)	gi|14010837	126	40.655	4.9	7	3	*0.67*	**0.35**
155	ATP synthase, H+ transporting, mitochondrial F0 complex, subunit F6	gi|16758388	117	12.487	9.9	26	4	*0.25*	**0.28**
10	polyubiquitin	gi|1050930	133	11.234	5.3	35	3	*0.25*	**0.25**
180	Rho GDP dissociation inhibitor (GDI) alpha	gi|31982030	128	23.450	5.0	21	5	*1.14*	**0.17**
36	rCG50668, isoform CRA_a	gi|149031956	134	25.532	10.2	24	5	*0.31*	**0.03**
36	rCG50690	gi|149031970	130	21.249	4.9	11	2	*0.07*	**0.03**

It was hoped that the two different methodologies would give comparable results, since directly spotting from a 384-well plate and analysing it by MALDI is significantly more rapid even than parallel monolithic separations. Unfortunately, the data does not bear this out, although further replicates are required to complete the findings, but it seems that even at a resolution of 64 x 24 fractions, directly spotted to MALDI target plates, it is advantageous to use LC/MS of a 32 x 24 fraction separation as a means of identifying quantitative differences. The most significant advantage of a 64 x 24 fraction separation, however, is that only 10% of the sample is used for analysis, leaving the remainder for further investigation, especially for those spots that clearly contain multiple proteins. 10% of the sample was chosen as this is the standard for proteins excised from 2D gel spots, so that direct comparison may be drawn between the two techniques. Intensity of precursors in MS mode was fairly low in comparison to the average intensity of spots excised from gels, and this may be a consequence of poor digestion or resuspension from the plate.

Two further intact protein labelling methodologies were attempted by associated researchers, ExacTag and DiMethyl labelling.

6.4.1.2.5. ExacTag

ExacTag is an isobaric labelling technology recently developed by Perkin Elmer. The company provides two different labels, a thiol labelling reagent, and a lysine labelling reagent. The former is of less interest as it shares the limitations of ICAT, especially that quantitation data must be derived from a small number of cysteine-containing peptides. Trials were therefore performed with the lysine labelling reagent, which, in fact is the same as the thiol labelling reagents with the addition of a lysine derivatisation step that converts its amine group to a thiol. This work was carried out in collaboration with Marek Kuzbicki, who performed the labelling reaction. Analysis of the data and subsequent bioinformatics was performed by the author.

The major drawback of this procedure is that missed cleavages will occur in the same way as iTRAQ with the added complication that the mass of the isobaric tag is extremely large (at 1046.48Da). The analysis of ExacTag labelled peptides is somewhat arcane. Quantitation is performed using the provided software, which takes an MGF file (mascot generic format, a peaklist used to search for protein IDs), and the results of a Mascot search as parameters. The Mascot search results are performed *without* the ExacTag modification enabled, and when searching with the modification enabled, no matches to labelled peptides have been obtained in any analysis performed to date. It is likely that the reason for this is the size of the mass tag, in that along with the sequence of the labelled peptide, the label also fragments into a peptide ion series, which results in an inability to correctly sequence the original peptide. If so, this is a major hindrance to the use of the tag, especially the lysine labelling version, which will label more frequently. It means that all protein identifications are based on the sequences of unlabelled peptides, which equates to those peptides containing no lysines. Those peptides assigned for quantitation are highly complex (see Figure 86), and there are a large number of masses in the reporter ion region that are clearly not reporter ions (see Figure 87). Additionally, the masses of the reporter ions are one Dalton smaller than anticipated. Poor calibration is unlikely to be the culprit for the change in mass, since the MS/MS accuracy window used in Mascot searches is too small to allow for the change, and multiple highly significant hits were observed.

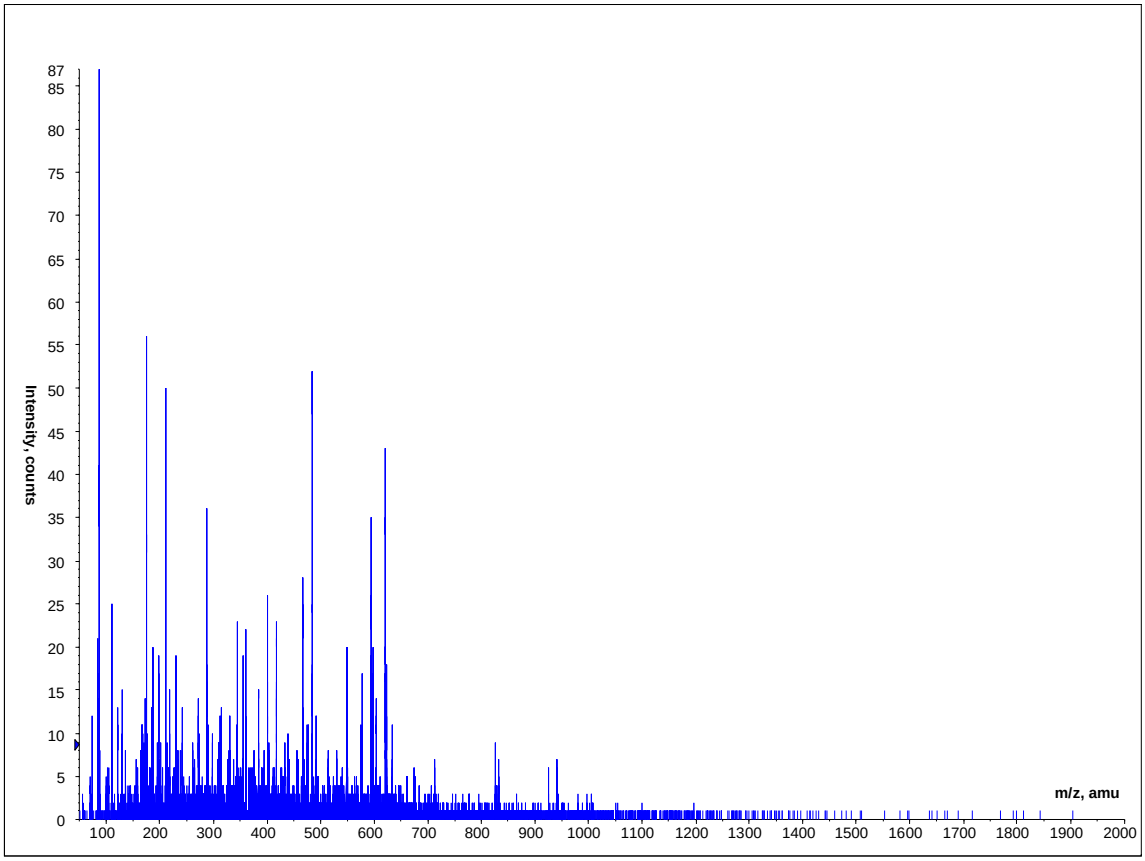


Figure 86: Fragmentation spectrum of the ExacTag labelled peptide SLGKVGTR from Ovalbumin. Note the complexity of the spectrum, which is significant for a small peptide. This is likely to be due to the additional complexity resulting from the large peptide tag, which in turn makes the spectrum very difficult to analyse.

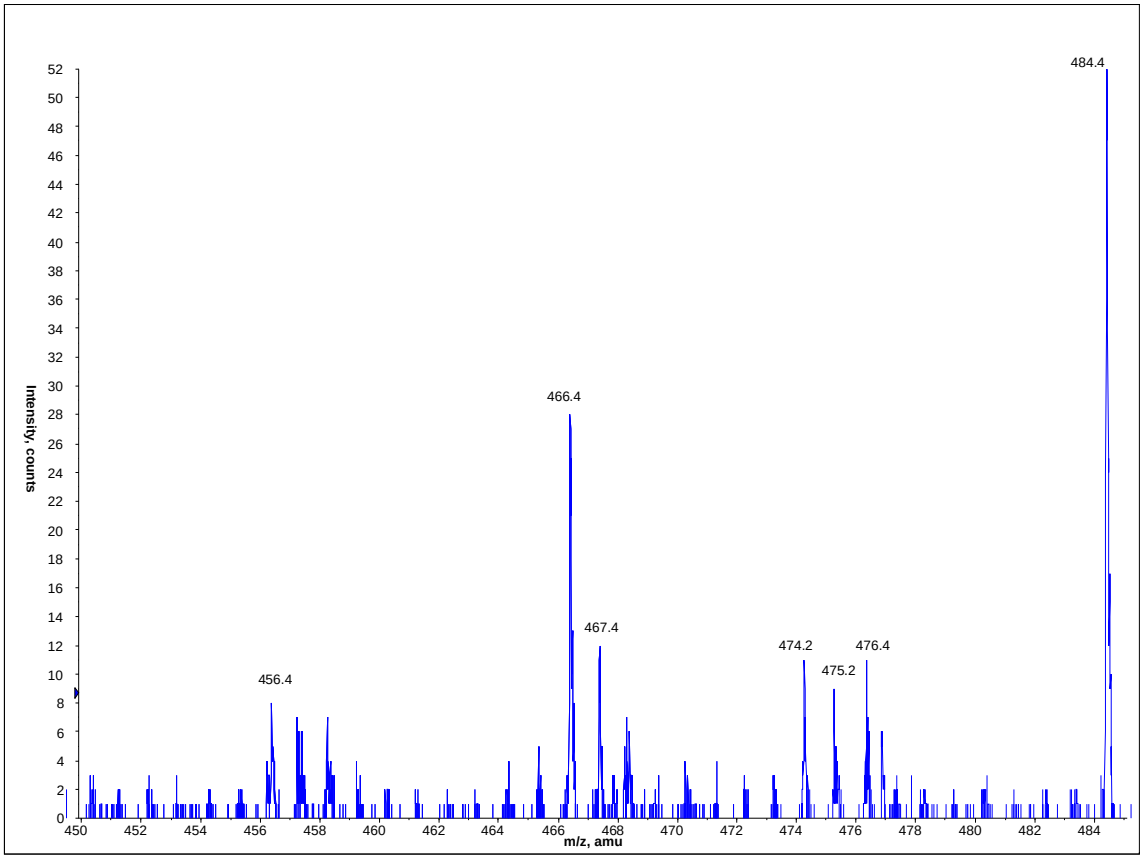


Figure 87: Reporter ion area of the spectrum shown in Figure 86. The labels used in this experiment were 457 and 475. The reporter ions appear to be one Dalton smaller than anticipated. The 466 peptide peak is unrelated to the experiment.

6.4.1.2.6. DiMethyl Labelling

Dimethyl labelling was developed by Chen [229, 249] and is a non isobaric protein and peptide labelling methodology based on the double methylation of lysine residues using either formaldehyde or deuterium labelled formaldehyde. This method has a number of distinct advantages over other labelling methodologies, primarily that it is comparatively inexpensive and very rapid. Its low cost and high availability, indeed are great benefits, as a large excess of reagent can be used to improve labelling efficiency. The application of dimethyl labelling to intact protein labelling was conceived by the author, and work was carried out in collaboration with Fahsai Kantawong, who performed the labelling reaction. Subsequent data analysis and bioinformatics was performed by the author.

When applied to a mixture of standard proteins, efficiency of labelling is excellent (see Figure 88 and Figure 89), with no detection of an unlabelled peak by the MASCOT software. Comparison between the extracted ion chromatograms of the light labelled peak of the peptide KVPQVSTPTLVEVSR, (mass 1666.64, triply charged mass 556.55)(see Figure 90) and the predicted unlabelled peak (mass 1638.61, triply charged mass

547.203)(see Figure 91) show a single distinct peak in the case of the former, and no significant peak in the case of the latter. The uneven and relatively high intensity chromatogram for the unlabelled peak is a result of the single C_{13} isotopic species of a singly charged polymer contaminant at 546.13, which interferes somewhat with the chromatogram.

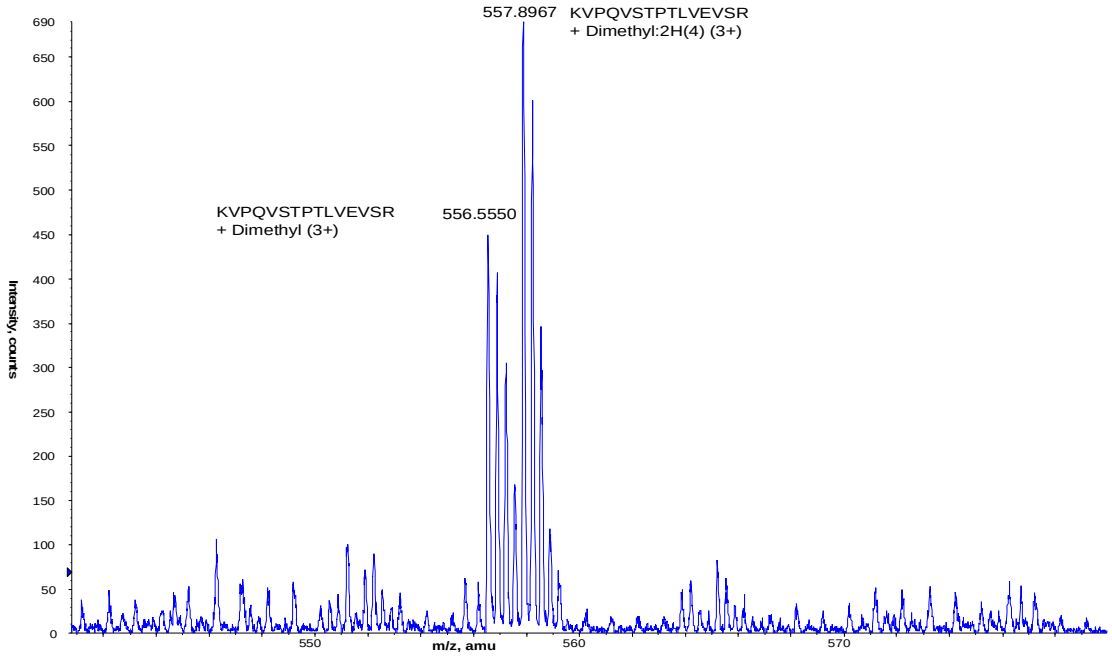


Figure 88: Dimethyl labelled pair derived from the triply charged peptide KVPQVSTPTLVEVSR. The distance between the monoisotopic peaks of each peptide is 1.3, corresponding to the 4 Da mass differential between heavy and light labeled proteins.

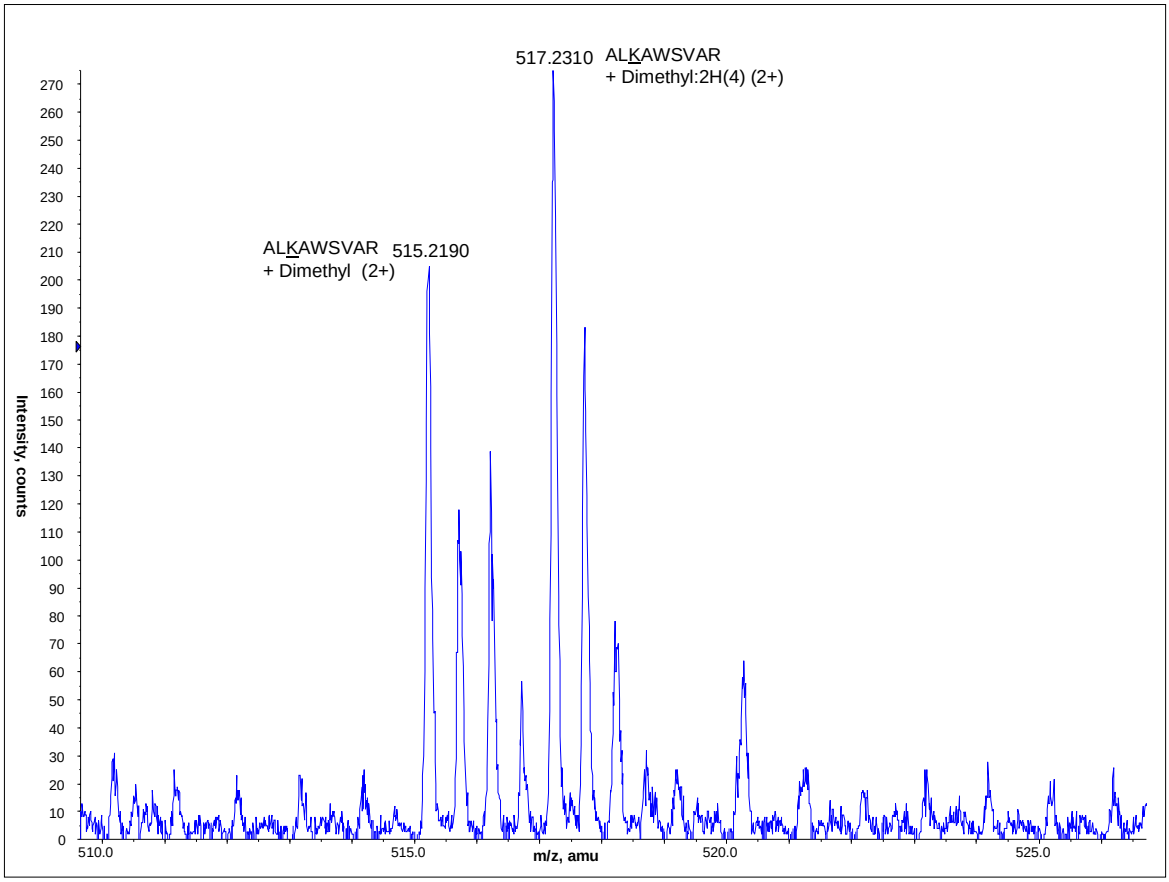


Figure 89: A second example of a successfully dimethyl-labelled peptide pair, this time a doubly charged peptide. Note the 2 m/z mass shift resulting from the 4 Da difference between heavy and light labels.

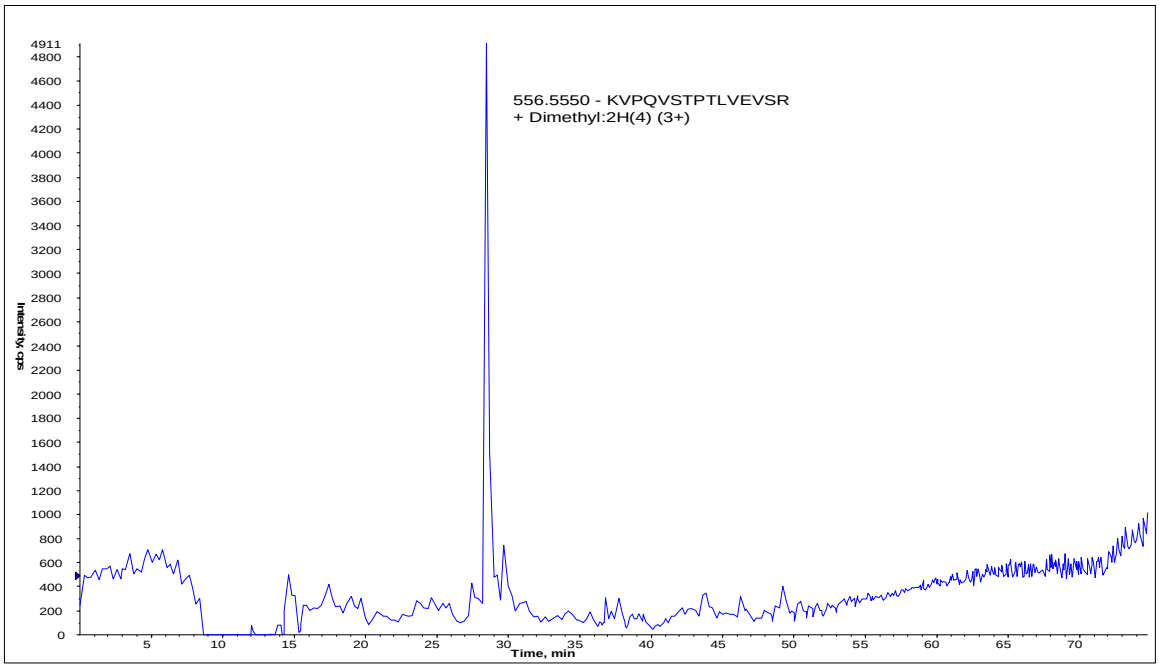


Figure 90: Extracted ion chromatogram of the triply-charged light-labelled peptide KVPQVSTPTLVEVSR, for comparison with the predicted unlabelled peptide shown in Figure 91. Note the distinct, high intensity peak resulting from elution of the peptide.

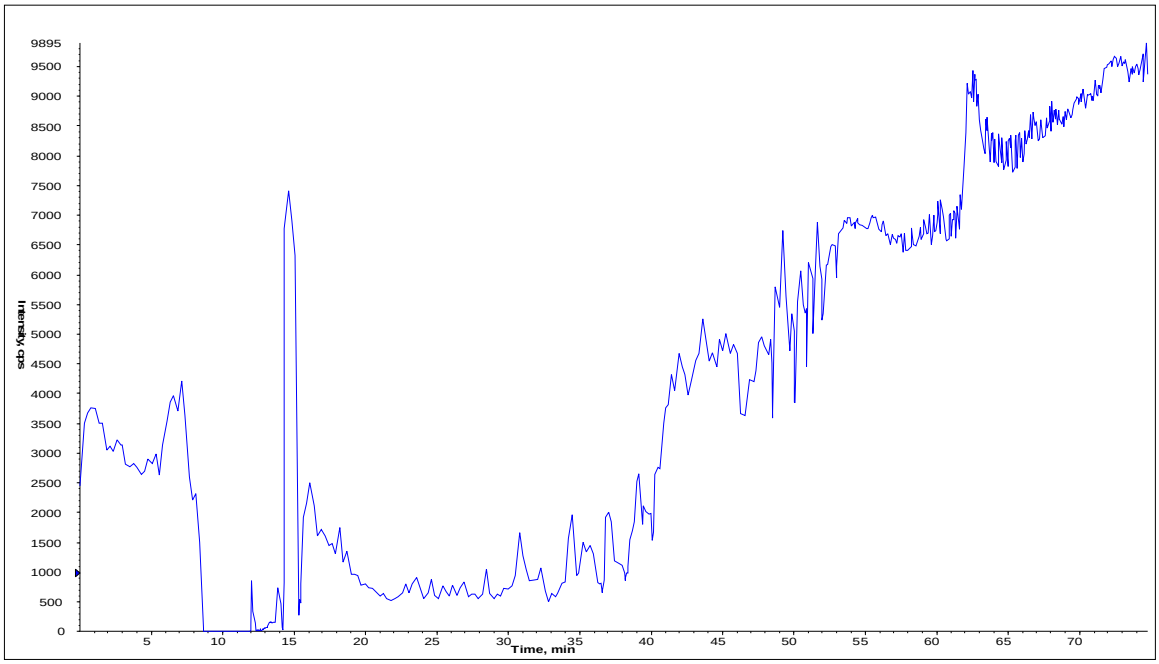


Figure 91: Extracted ion chromatogram of the predicted 547.203m/z of the 3+ charged, but unlabelled peptide KVPQVSTPTLVEVSR. Although the number of counts for this peptide rises to significant levels towards the end of the chromatogram, this is due to its similarity to the single C13 peak of a polymer contaminant, and there is no significant peak whose monoisotopic mass is equal to 547.203, showing that labeling is highly efficient.

Dimethyl labelling was first applied in a proteomics context to E7 transformed human uroepithelial cells both untreated and treated with arsenic [229]. The duplex label was further expanded in 2006 [249] to a quadruplex label with the use of d3 sodium cyanoborohydride. With the use of LysC for digestion purposes (which cleaves only at the C terminus of lysine residues), a mass difference of at least 4 Da is guaranteed. However, the major benefit of iTRAQ over this approach is that all the peaks are combined in MS, allowing each sample to contribute to the overall intensity of a particular peptide. With quadruplex dimethyl labelling, four times the amount of sample is required to provide the same signal. Additionally, the benefits of using a LysC digestion protocol to allow assurance of a 4Da shift cannot be applied in a protein labelling context, as labelling is performed prior to digestion.

6.5. Conclusions

The different methods of applying quantitation to intact protein separation: label free quantitation, isotopic labelling, and fluorescence/absorbance analysis, all have a number of advantages and disadvantages that must be considered before planning an experiment.

All current label-free methodologies fail to support multidimensional analyses, and combination of multiple label free analyses into a single multidimensional analysis map is non trivial, especially due to the unavoidable peak splitting observed in any fraction collection regimen. A small shift in retention time in different separations is likely to cause the apex of two protein peaks to be split between two fractions, especially the very sharp peak widths provided by PS-DVB based columns, despite their superior reproducibility. Thus, even if the abundance of both peaks is identical, in one fraction the software would report upregulation and in the next downregulation, as the apex of one peak is compared to the slope of the other, and vice versa. Combination of isobaric peaks with the same retention time across multiple runs is a potential way of dealing with this situation, but one of the primary reasons for performing a multidimensional separation is that isobaric peaks may, in fact, be different species with different physiochemical properties. Thus great care must be taken to ensure the rigour of any grouping together of similar peaks.

Most of the time allotted to quantitation, therefore, has focussed on isotopic labelling procedures, since in these cases, measurement of relative abundance is tied directly to observed peptides. With any fluorescence or absorbance based quantitation technology, a differentially expressed peak or spot will often represent multiple proteins, with the probability inversely proportional to the resolution of the separation and the number of fractions collected. The major limitation of isotopic analysis is, of course, that the complete set of fractions must be analysed to produce complete quantitation data, which becomes very time-consuming when applied to high resolution multi-dimensional separations, whereas fluorescence/absorbance methods provide the ability to limit analysis to only those samples containing quantitatively different proteins.

Hybrid methods [250] using, for example, DiGE minimal labelling coupled to ICPL saturation labelling to achieve the benefits of most methods are, on closer inspection, somewhat superfluous. Fluorescent labelling is performed to allow selection of specific fractions, and ICPL allows deconvolution of the specific proteins producing the observed quantitative difference. However, isotopic labelling is likely to show multiple quantitative differences in a particular fraction, and is a tacit admission that resolution of the separation is not sufficient to provide fractions each containing a single protein quantifiable by fluorescence. If fluorescence alone does not provide effective quantitation as it fails to take into account multiple quantitative differences in co-eluting species, it suggests that isotopic labelling alone, coupled with full analysis of the separation, is required for effective and complete quantitation. If, on the other hand, fluorescence provides effective quantitation

for, say, the most abundant species in a fraction, and the remainder may be safely disregarded, then simply labelling the sample with fluorophores would provide the same quality of quantitation.

Due to the exceptional reproducibility and high resolution of the chromatography with the PS-DVB based columns used for intact protein separation, UV analysis is worthy of further development work. Of course, for effective quantitation, multiple, high resolution replicates of each separation must be performed. This is likely to be very time-consuming, and expensive in terms of sample, at 200µg per separation, but once a high-quality map of quantitative differences across a separation can be produced, analysis of the specific, differentially regulated fractions is potentially very rapid, depending, of course, on the number of differences observed.

Isotopic labelling reactions, however, are also time-consuming. To perform multiple replicates in a single separation requires the use of either iTRAQ or ExacTag, with the potential for either two (for 4plex iTRAQ), four (for 8plex iTRAQ) or five (for 10plex ExacTag) binary replicates from a single experiment. However, protein labelling with iTRAQ requires four times the amount of reagent per 100µg of sample compared to peptide labelling, making this type of labelling experiment expensive, and the peptides generated by ExacTag are too large to be effectively analysed by standard mass spectrometers. Dimethyl labelling is the least expensive, and also the smallest lysine protein tag, but it is at best a quadruplex reagent (based on [249]), and if used in the manner described in the paper, will suffer the same isotope pattern overlap observed in O18 labelling, due to the small mass gap between labels (+0, +2, +4 and +6). However, this labelling methodology should be good for duplex experiments, which comprise the majority of quantitative proteomics experiments.

Of all the quantitation experiments performed during this project, SILAC provided the most detailed results. This is likely to be due to the nature of metabolic labelling of essential amino acids, as labelling is applied during culture of the cells and therefore incomplete labelling can be essentially eliminated by extended doubling times of cell culture. Ong [33], for example, shows that 5 doubling times are necessary for more than 97% labelling efficiency. This leads to a highly efficient, unbiased labelling technique that is unaffected by lysis conditions or any form of separation performed. SILAC may be multiplexed for up to three conditions [251] (as in the experiment described in section 6.4.1.2.4), which is less than that of iTRAQ (4 or 8-plex) or ExacTag (2 to 10-plex), but

sufficient to compare a control sample, a test sample, and a pooled standard for improved matching between multiple experimental replicates. If research can be performed on a cell line compatible with the SILAC labelling procedure, it is recommended that this type of quantitation is applied. For clinical samples and cell-lines incompatible with metabolic labelling, an alternative methodology must be employed. If a preferred method were to be chosen for differential analysis of SILAC incompatible cells by multidimensional intact protein separation, the choice would be between UV-map based quantitation, iTRAQ or dimethyl labelling.

For UV-map based quantitation, the preferred methodology would be to use low-bore columns, such as 1mm or 500µm in the first dimension, and 200 or 100µm in the second dimension for multiple statistical replicates so as to limit sample consumption while still providing good signal to noise for quantitation. Very high resolution would be required, with one fraction collected every 20 seconds in the first dimension, leading to roughly 60 ion exchange fractions for further separation. Eluents could either be passed to waste, which would require a final large-load separation for protein identification, similar to the 'prep gel' from DiGE, or collected in layers in the same vessels, relying on the reproducibility of the chromatography to ensure collection accuracy. To reduce the separation time for each replicate from two days to 10 hours, it would be highly advantageous to apply a parallel protein separation system to the small bore colums. MS identification time, is, of course, in addition to this, but may be rapid analysis by MALDI, or fast peptide separation using monoliths. This quantitation methodology is recommended if MS time is at a premium, and there is free access to an appropriate separation system.

An iTRAQ experiment, however, has the capability to perform at least two replicates (four with the recently released 8plex reagent) simultaneously, due to the isobaric nature of the reagent. Biological replicates are labelled, mixed together, and samples are deconvoluted after analysis using the reporter ion in the fragmentation spectrum. Analysis of all samples is required, but fraction collection from the protein separations need not be of the same resolution as for UV quantitation, since a third dimension peptide separation will be applied to the samples (a single 384 well plate has been sufficient for labelled separations during this research). This methodology is recommended if sufficient reagent is available, and if a suitable mass spectrometer and separation system capable of expediting the analysis of hundreds of samples while providing fragmentation data in the iTRAQ reporter mass range is available (this is generally not the case on ion trap instruments due to the loss of signal below 33% of the m/z of the precursor).

Finally, dimethyl labelling provides a useful alternative for routine investigations if machine time is sufficient to perform multiple replicates of the separations, as this technique is most robust when applied to duplex samples.

7. Chapter 7 - Separation of proteins for enhancement of Top-Down analysis.

7.1. Aims:

Top down analysis of proteins is an important but challenging area of proteomics. Traditionally, it is performed using extensively purified proteins, but 2DLC of intact proteins provides the possibility of performing on-line analysis of complex mixtures using Top-Down methods.

The aims of this study were:

1.) To apply PS-DVB monolithic columns to protein separation, online with high-resolution FT-ICR MS.
2.) To investigate 2DLC's application to top-down analysis.
3.) To investigate electron transfer dissociation (ETD) and proton transfer reaction (PTR) and their applications to top-down analysis.

7.2. Introduction

The vast majority of proteomics experiments performed to date have relied on tryptic digestion of proteins prior to analysis. Digestion has been shown to have huge benefits for the analysis of proteins: firstly that it is possible to identify proteins based entirely on the masses of the tryptic fragments (PMF) and secondly that conventional chromatography techniques are most effective when applied to peptides. There are, of course, many drawbacks to this technique, not only that information on protein isoforms is lost, but especially that the complexity of the sample is increased by an order of magnitude [252].

Direct analysis of intact proteins (top-down) has many benefits for the analysis of protein species. Sample complexity is held at the minimum possible, the intact precursor can provide many clues as to the identity of the protein, and post translational modifications are easier to identify due to the shift in the precursor mass. Of course in-MS fragmentation of the protein must still be performed to provide sequence data and hence unambiguous identification of the protein, and another benefit of top-down is that a sequence of product ions is generated without requiring recompilation of the intact protein from multiple tryptic fragments. In addition to the CID fragmentation normally used for peptide fragmentation,

alternative fragmentation techniques such as electron capture dissociation (ECD) [109] and electron transfer dissociation ETD [110] can be employed to improve sequence coverage and confidence of not only overall protein but also PTM identifications. The electron transfer and capture fragmentation techniques are of particular interest as their efficiency increases proportionally to the charge-state of the precursor, which for intact proteins ranges from 6 to 60 and above. A further advantage of top-down analysis is that information about the populations of differentially post-translationally modified proteins can be directly obtained.

Conventional top-down analysis is performed on ultra-high resolution FT-ICR instruments, commonly giving resolutions of 200,000 or greater on 9+ Tesla instruments. In this study we used the 12Tesla Apex IV FT-ICR (Bruker Daltonics), which is fitted with ECD and a mass resolving quadrupole coupled to a collision cell for precursor isolation and collision induced dissociation (CID), although some success has been obtained using modified Qq-ToF type instruments fitted with extended range quadrupoles [253].

As an alternative to the ICR instruments, the addition of ETD to comparatively inexpensive ion trap instruments, such as the HCT Ultra, has allowed efficient fragmentation of larger ions, but the poor resolution (approximately 3000) of these instruments has made effective analysis impossible. To overcome this limitation, recent development work on ion traps [119] resulted in a method of stripping charges from protein and peptide species, based on the transfer of protons from peptides to radical species. This technique, known as proton transfer reaction (PTR) is described more fully in the introduction. In practical use, after fragmentation of an intact protein by ETD, PTR can be applied to the resulting mixture of highly charged ions. As the rate of reaction, like ETD, is proportional to the charge state of the ions, highly charged species are rapidly stripped down to lower charge states, capable of being resolved by the ion trap (1-4+), while peptides with charge states low enough to be deconvoluted by the trap react more slowly. Although this rapidly forces larger fragments out of the trap's mass range, sufficient sequence can be generated to either completely sequence smaller proteins, or sequence the N and C termini of larger proteins.

Although Top-Down analysis may be described as analysis at the lowest complexity available, heavily modified proteins are still extremely challenging to analyse. Conventionally, complex separation methodologies are employed to produce purified protein samples for analysis. Alternatively, as described in section 7.3.1.2, it is possible to

separate and analyse intact proteins on-line with high signal to noise and narrow peak widths using PepSwift monolithic PS-DVB columns. In addition to the HPLC and MS, sample collection and infusion is assisted by a NanoMate (Advion). The NanoMate comprises a robotic sample handler, a microchip-based electrosprayer, and a post-column flow splitter, which can be used to divert a portion of the sample for high resolution fraction collection, while the remainder is electrosprayed into the mass spectrometer. On-line mass profiling from the NanoMate-assisted electrospray into the MS provides candidates for further analysis, which, selected from the simultaneously collected fractions, and applied using the 'off-line electrospray' function of the NanoMate, where samples are robotically aspirated and infused at low flow rate. PepSwift monolithic reversed phase columns are based on aromatic ring chemistry, leading to similar but distinct separation properties when compared to conventional aliphatic hydrocarbon (C4, C8, C18 etc.) resins. In addition to protein standards, and the complex samples discussed in other chapters, these columns have been employed to separate histone proteins derived from biological samples for use with top-down analysis.

Histone proteins are important targets for research, due to their intimate association with DNA and the mechanisms of transcription. Additionally, disruption of normal histone function is implicated in many cancers. Analysis of histones, however, is not amenable to traditional 'bottom-up' proteomics. Complex gradients must be employed to provide partial separation of histones [4, 254-257], and the extremely basic amino acid composition of these proteins leads to digestion into extremely short peptides by trypsin or extremely long peptides for other proteases.

7.3. Methods and materials

Chromatography was performed using an UltiMate® 3000 (Dionex, Camberley, UK), coupled to a NanoMate (Advion) robotic electrospray and fraction collection device to provide electrospray, concomitant fraction collection and off-line reinfusion of collected fractions. ESI MS was performed using an Apex IV Qe 12T FT-ICR (Bruker Daltonics, Bremen, Germany), with the assistance of Colin MacKay for methods development, histone separations and *L. donovani* analysis, and an HCT Ultra PTM Discovery System (Bruker Daltonics, Bremen, Germany) was used for ETD/PTR analysis of SILAC labelled K562 cells. A generalised workflow of the separation and analysis using the FT-ICR MS,

including an optional prior dimension of ion exchange, is shown in Figure 92.

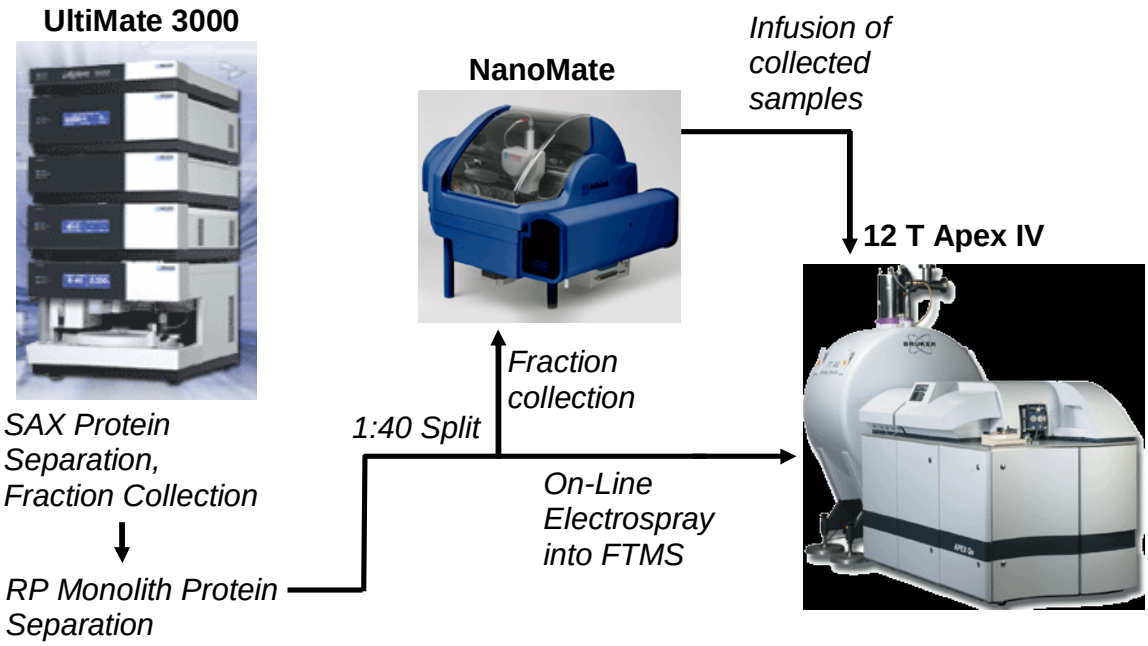


Figure 92: Workflow for FT-ICR MS and MS/MS of protein samples. Complex samples are first separated by strong anion exchange. 1 minute fractions are collected and reinjected for second dimension reversed phase analysis, but for simple samples, such as protein mixtures or histone samples, this can be omitted. Reversed phase separation is performed with a 5cm x 500µm PepSwift monolith. This separation is performed on-line with a 12Tesla Apex IV FT-ICR mass spectrometer, to generate mass profiles of eluted proteins. Concomitant fraction collection using an Advion Nanomate allows subsequent selection of protein species and further fragmentation analysis to produce sequence data.

7.3.1.1. Reagents and standards

The 6 protein mix used for methods development consisted of bovine serum albumin (5 mg/mL), cytochrome c (2 mg/mL), α-lactalbumin (6 mg/mL), ovalbumin (6 mg/mL), carbonic anhydrase (10 mg/mL) and myoglobin (1 mg/mL). All protein standards were obtained from Sigma (Dorset, UK). Histone samples for PTM analysis were produced by precipitation from cell lysates using sulphuric acid as described in 2.2.5, and supplied in solution by Bernard Ramsahoye. General reagents were obtained as described in section 2.1. Wild-type *L. donovani* cells were produced by Christina Nuala, and lysis was performed by the author as described in section 2.2.2. SILAC labelled K562 lysates for ETD/PTR analysis were prepared by Christian Preisinger as described in 2.2.3.

7.3.1.2. Chromatography

Reversed phase separations were used for the 6 protein mix and histone samples. Aqueous phase consisted of 2% v/v acetonitrile, 0.1% v/v formic acid, and 97.9% v/v ddH_2O.

Organic phase consisted of 80% v/v acetonitrile, 0.08% v/v formic acid, and 19.92% v/v ddH_2O. Gradients were applied to a PepSwift (Dionex UK) 500 µm x 5 cm PS-DVB monolithic column at a flow rate of 20 µl per minute, maintained at a constant temperature of 60ºC. UV detection was performed at 214 and 280 nm. Gradient details are shown in Table 21.

To provide an additional dimension of separation for complex samples, ion exchange chromatography was performed on 200 µg of *L. donovani* lysate and 1 mg of SILAC labeled K562 lysate. Columns used were coupled ProPac™ SAX 2 mm x 250 mm and SCX 2 mm x 50 mm ion exchange columns, at a flow rate of 200 µL per minute. Fractions were collected at a rate of one per minute in 96 well microplates and reinjected for a second dimension of reversed phase separation as described above. The ion exchange gradients used are described in section 2.3.

Table 21: Gradient table for reversed phase intact protein separations applied to top-down analysis.

Time (min)	MeCN (%)
0.0	2%
20.0	56%
20.1	72%
24	72%
24.1	2%
27	2%

7.3.1.3. FT-ICR ESI-MS/MS Conditions

The FT-ICR MS was maintained in positive ion mode with a scan rate of 1Hz for all analyses. For both LC-MS and infusion, ESI was performed using the NanoMate with an electrospray voltage of 1.6 kV, and spray was assisted with nebulising gas at 0.4psi. Detection range was 500-3000 m/z. For MS/MS of reinfused samples, ECD was performed with a current applied to the cathode of 1.8 A, and a reaction time of 0.00060 seconds.

7.3.1.4. Ion Trap ESI-MS Conditions

Samples were introduced to the HCT Ultra through a stainless steel low-flow nebuliser (Bruker Daltonics, UK) at 4 kV with a nebuliser gas pressure of 15 psi and drying gas of 5 L per minute. Source temperature was set to 250ºC. The HCT was maintained in ‘enhanced

mode' with a scan rate of 8000 amu per second. The mass range applied to detection was 100-3000 m/z, and 5 scans were summed to produce each MS spectrum. For MS/MS of reinfused samples, the CI source for both ETD and PTR was maintained at a temperature of 60ºC. The ETD reagent was activated with the application of a 3 µA current, and the PTR reagent was activated with the application of a 17 µA current. ETD and PTR reactant accumulation times were maintained by MS control, and set to 100,000 counts and 150,000 counts respectively.

7.4. Results and Discussion

The preferred method for analysing proteins top-down directly on an FT-ICR instrument is by infusion; numerous fragments are generated from the protein by ECD or CID and therefore the signal-to-noise of any individual ion in the fragment spectrum is consequently poor, even if signal for the precursor is strong. Continuous infusion of the sample provides the time required to sum multiple scans for improved signal to noise.

To provide enough material for direct infusion of collected fractions, a 500 µm PepSwift monolith was selected, due to its superior loading capacity and higher flow rates. To maintain the option for on-line fragmentation, as well as the generation of on-line mass profiles to provide targets for analysis, a 1:40 split was implemented, with the majority of sample collected at 10 second increments in microplates, and ~500 nl per minute electrosprayed into the FT-ICR.

A 2 µl aliquot of a mixture of 6 proteins – bovine serum albumin (5 mg/mL), cytochrome c (2 mg/mL), alpha lactalbumin (6 mg/mL), ovalbumin (6 mg/mL), carbonic anhydrase (10 mg/mL) and myoglobin (1 mg/mL) was placed in a 96-well plate and injected directly onto a 500 µm monolithic column (see Figure 92). Separation was performed with a standard 20 minute gradient. Half height peak widths for the separation are ~15 seconds, leading to insufficient time for on-line fragmentation due to the low scan rate of the MS. Therefore wells were selected by reference to the initial on-line analysis for infusion and more in-depth analysis. ECD of one smaller protein matched well to cytochrome c, although sequence assignment required removal of the N-terminal methionine, manual assignment of the N-terminal acetyl glycine and the cysteine-bound haem group at position 14 (see Figure 93). Interestingly, to provide successful matches, the mass window was required to be opened to 1.2Da, which is very wide especially considering the normally very accurate mass measurements produced by the FT-ICR MS. By analysing the mass defects of individual ions, it is clear that roughly 50% of ions have a shift from the predicted mass of

almost exactly 1 Da (see Figure 94). This is likely to be due to poor assignment of the monoisotopic peak, since the software assigns masses to the predicted monoisotopic peak of a single ion population. Therefore, the requirement for a wide mass window for the ions is a result of poor identification of the monoisotopic peak of many species, rather than a limitation of the instrument.

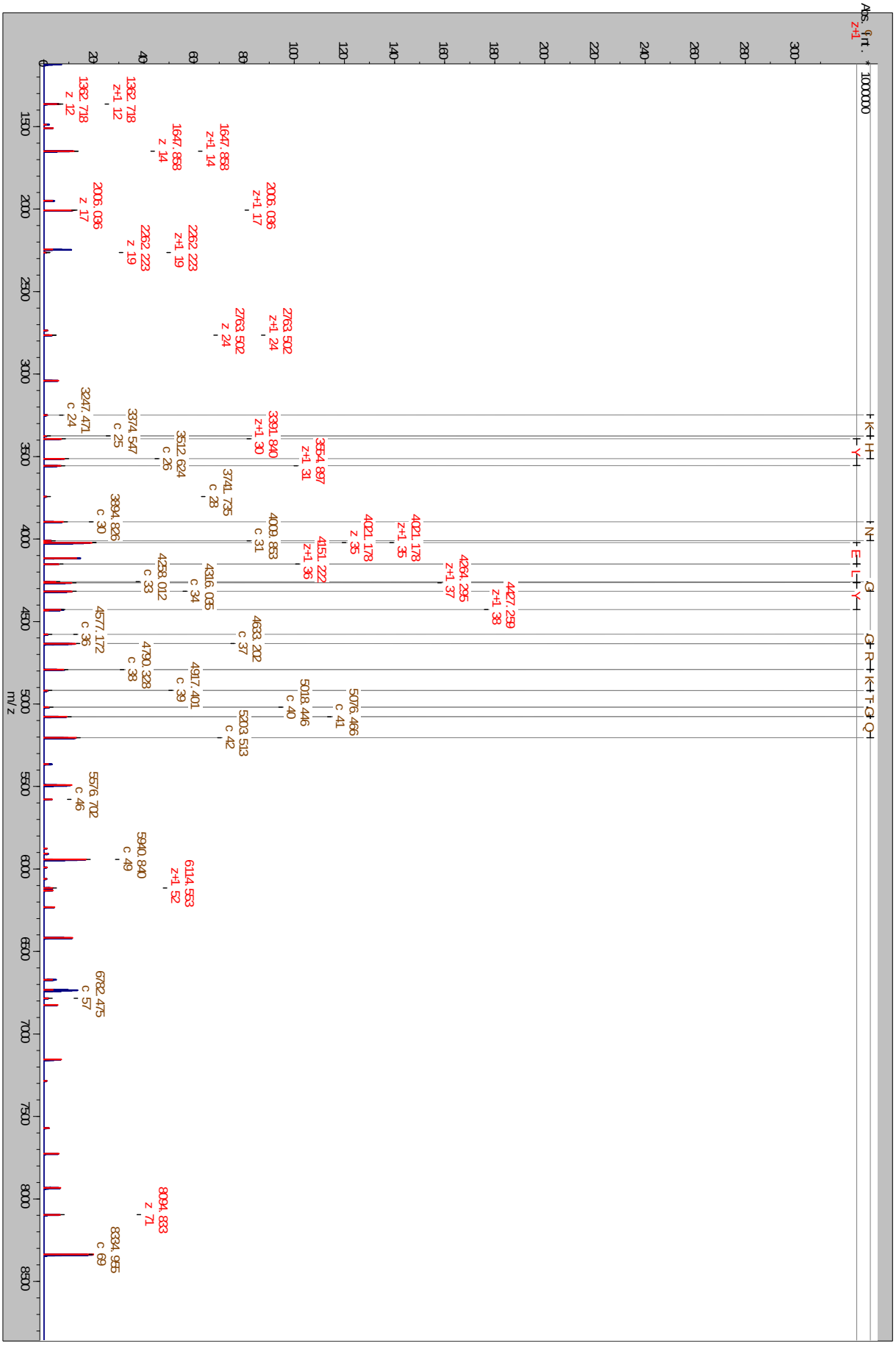


Figure 93: Section from the ECD fragmentation spectrum of Cytochrome C from a protein mix separation. Note that much of the sequence is assigned, with a large number of c and z ions matched and enough to provide partial sequence runs.

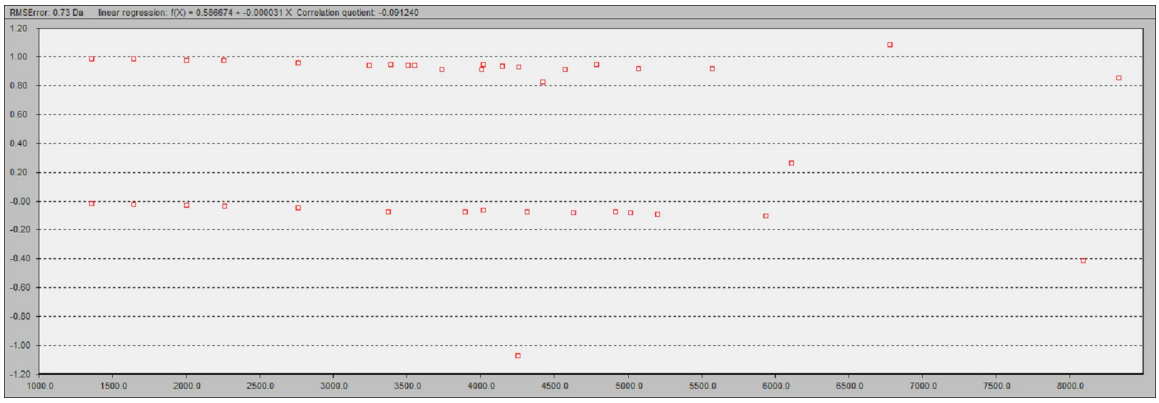


Figure 94: Mass error in Daltons of MS/MS fragment ions matched to corresponding sequence fragments of cytochrome c.

7.4.1.1.1. Analysis of Leishmania by FT-ICR

Of course, a single dimension of separation is insufficient for effective separation of biological samples such as whole cell lysates. Several researchers (e.g. [47, 187, 188, 258]) have performed multidimensional chromatography both on and offline, but the generally poor separation quality of the predominantly C4 columns used are coupled with a requirement for relatively large amounts of starting material (e.g. ~3 g of S. cerevisiae for the ALS/RP method described in [259], 10-15 mg of yeast in [187]).

Until recently (2006) Prof. Kelleher's group utilised a prep gel-based first dimensional separation using Acid Labile Surfactants (ALS) rather than SDS, which would cause problems for downstream mass spectrometric analysis. Reversed phase separation involved the use of a 20 minute gradient on a C4 column using standard or non-porous silica resins. Although separation quality is excellent, the initial use of very large amounts of material precludes this type of experiment for anything unable to be cultured in large quantities. The peak capacity of ALS-PAGE/C4 RP is quoted as being approximately 400 peaks, in contrast with the calculated peak capacity of a PS-DVB multidimensional separation, which is, at a conservative estimate, 3600 peaks.

In contrast, more recent experiments utilised the Beckman PF2D system [258], which is based on a chromatofocussing first step, followed by C4 chromatography, and requires up to 10 mg of protein (1-5 mg were used in the most recent papers). Poor resolution on the first dimension (in comparison to both PS-DVB based ion exchange, and the pISEP technique described in section 4), and poor resolution in the second dimension (in comparison to PS-DVB monolithic reversed phase), suggest that significant improvements could be made by switching to the method described in this thesis.

The most recent experiment [187] published by the Kelleher group employs a similar set up to that described above, using ion exchange chromatography, followed by reversed phase separation to analyse proteins on-line (previous examples relied on re-suspension and infusion of collected fractions). In contrast to the PS-DVB strong anion exchange column used to perform separations for this project, a weak anion exchange column was used to separate 15 mg of total protein, followed by a 70 minute gradient on reversed phase separation on a C5 column. Chromatograms are not shown for either dimension of separation, but they are unlikely to be significantly higher resolution than others described by the same group. This approach has an advantage for on-line analysis, in that more time is available for the slow-scanning FT-ICR to acquire spectra, but the significant drawback that a large amount of starting material is used to compensate for the poor signal to noise of each peak.

To test the suitability of the PS-DVB IEX/RP high resolution chromatography method in development, 200 µg of a *Leishmania donovani* lysate was separated by anion exchange and further separated using monolithic columns, with direct coupling to an FT-ICR and simultaneous fraction collection with a NanoMate (see Figure 95).

Analysis of the chromatogram provided a number of targets for further top-down analysis, and collected fractions were selected for infusion on this basis (an example spectrum of an infused protein is shown in Figure 96). ECD was performed on the selected $1021.5547_{(8+)}$ m/z, peak, isolated using quadrupolar mass filtering: the fragment spectrum is shown in Figure 97. Note that the MS spectrum contains a large number of peaks not associated with the protein of interest, presumably from the other protein and peptide species present in this fraction, highlighting the requirement for precursor isolation prior to fragmentation to select the species of interest. The ECD fragmentation spectrum was summed over 700 scans and contains 120 discernable peaks for which charge states could be determined and this spectrum was used to unambiguously identify the protein. However, the peaks obtained after fragmentation are dominated by charge reduced species where an electron has been captured but radical dissociation has failed to occur. At maximum ECD efficiency, approximately 60% of the precursor remains unfragmented or charge-stripped, thus limiting the sensitivity of the technique. At just over 1 second per scan, it is clear that to obtain high quality fragmentation data in chromatographic time requires further refinement of the technique or instrumentation.

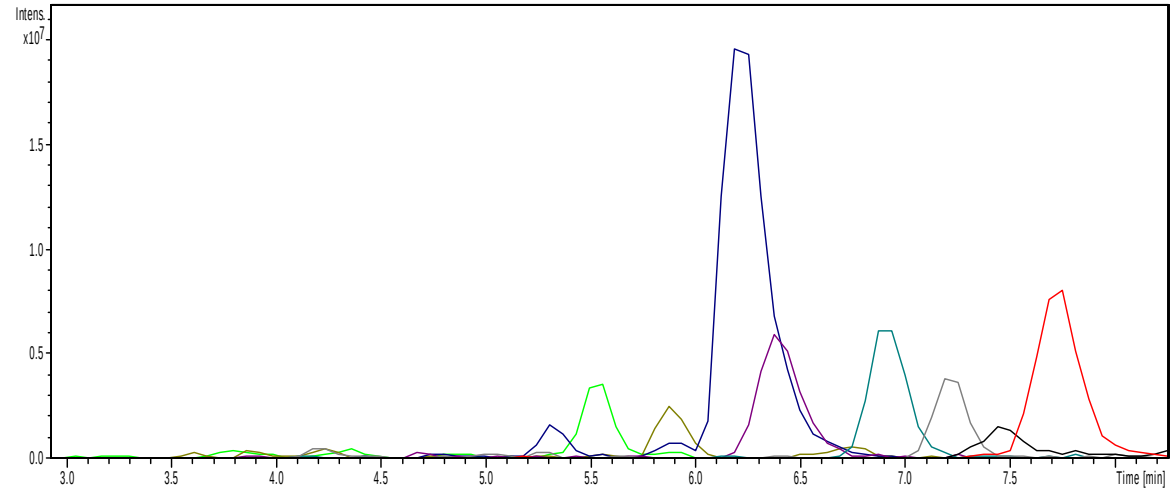


Figure 95: Extracted ion chromatograms for a variety of proteins separated by PS-DVB monolithic chromatography. Initial sample was an ion exchange fraction from a separation of a *Leishmania* lysate. Base peak widths are ~30 seconds except for the very abundant peak, which is evidence of overloading. Different colours correspond to individual small protein species.

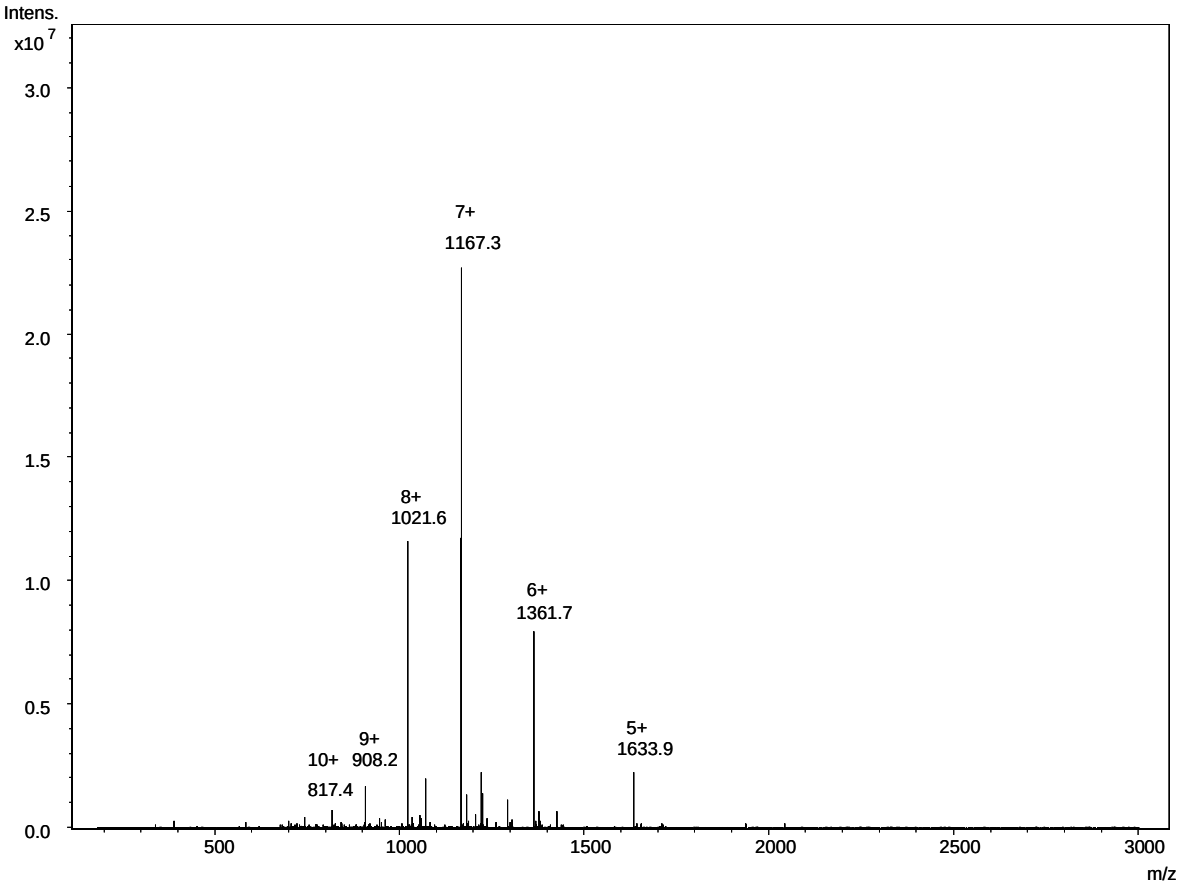


Figure 96: MS of a small protein (8160Da) prior to selection for analysis by electron transfer dissociation. The five major peaks of the spectrum constitute the charge envelope of the protein, ranging from 9+ to 5+. The 8+ peak at 1021.5547 was chosen as the ECD precursor.

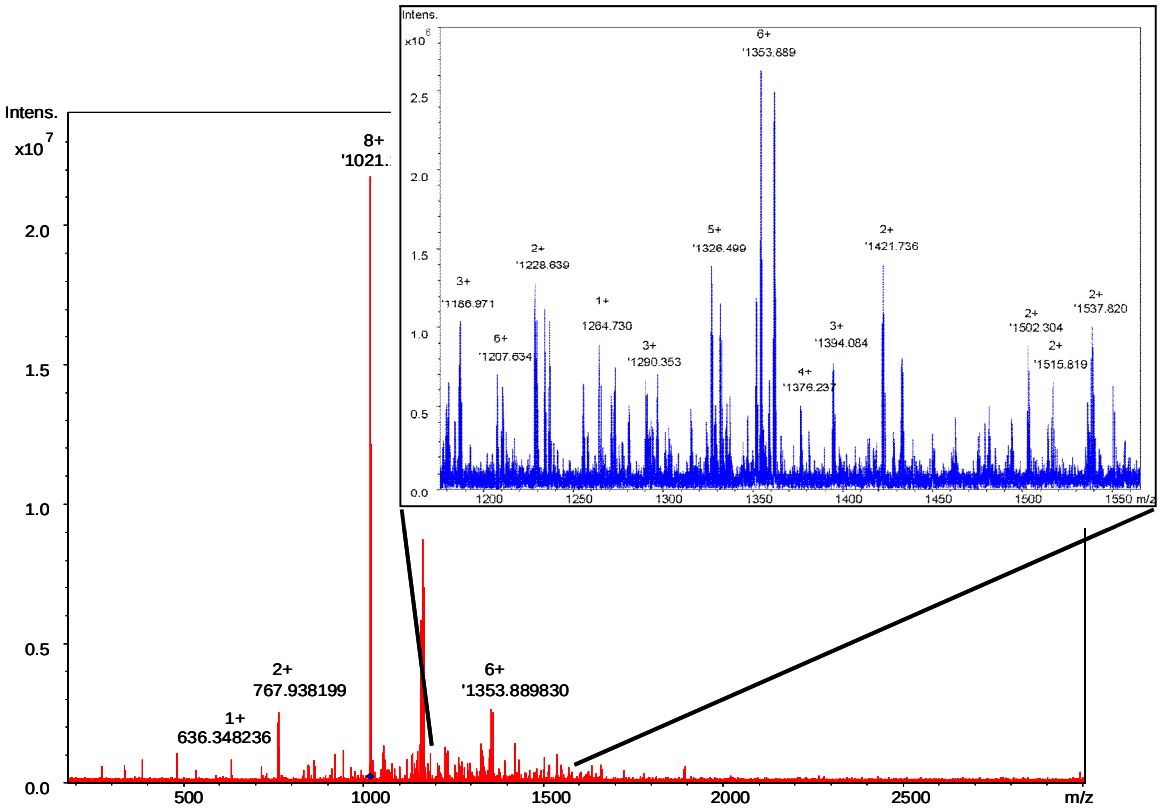


Figure 97: Electron capture dissociation spectrum of a ubiquitin ligase protein, inset shows a short section from 1200 to 1550m/z to demonstrate the very high complexity of the fragment ions obtained. These spectra are a challenge to analyse, but can provide substantially more information than a conventional 'bottom-up' approach.

Identification of the products of intact protein fragmentation are very difficult to interpret for several reasons: unlike typical peptide fragmentation spectra containing predominantly singly or doubly charged ions, ions with a broad spectrum of different ions with charge states (which could be between one and the charge state of the precursor) are observed, which require a high resolution mass spectrometer in order to determine the charge state. In addition, the software available for intact protein analysis is primitive in comparison to peptide fragmentation analysis software such as Mascot and X-Tandem. The only freely available software, ProSight PTM [260], maintained by Dr. Kelleher's group requires either that a set of deconvoluted masses are provided for an organism already in their in-house database, or a candidate sequence is provided to search against. Therefore a pre-loaded database was not available, as *Leishmania donovani* was the organism used for sample generation in this case. Hence, a semi-automated approach was adopted, which required initial manual sequencing followed by the use of ProSight. Manual *de novo* sequencing using the mass list obtained from the deconvoluted spectrum led to a three amino acid sequence that was searched against the *Leishmania infantum* protein database by simple text matching. This provided three possible targets out of the entire predicted proteome. Although poor scores were obtained from two of the sequences ProSight PTM

confirmed a match to truncated Ubiquitin Fusion Protein. Initially, a low, but significant score was obtained, as the sequence matched only using c ions – those initiating from the N-terminus of the protein. However, truncation of the sequence of the original protein to match the mass of the observed precursor (8164Da) resulted in a match of 30 z ions, starting from the C terminus of the fragment and a P-value of 1.8×10^{-96} (see Figure 98). The protein was therefore theorized to be truncated from an original mass of 14673amu to the observed mass of 8164.4amu with an overall mass deviation of 15ppm for all 59 matched ions. Subsequent analysis using the BioTools software (Bruker Daltonics) led to assignment of 84 ions within a mass error of 150ppm (Figure 99). Matches cluster around 0Da mass deviation from expected, but increase in inaccuracy proportionally to the mass of the fragment, although 11 very high mass fragments (ranging from 6465 to 8160) have mass defects very close to 1 Da, again suggesting mis-assignment of the monoisotopic peak.

Figure 98: Output of the software ProSight PTM for a targetted analysis of the masses from the ECD fragmentation of the 8164Da precursor, matched to the truncated sequence for the ubiquitin fusion protein LmjF31.1900. Blue lines denote identified fragmentation products, with southeastern pointing ticks corresponding to z ions, and northwestern pointing ticks corresponding to c ions.

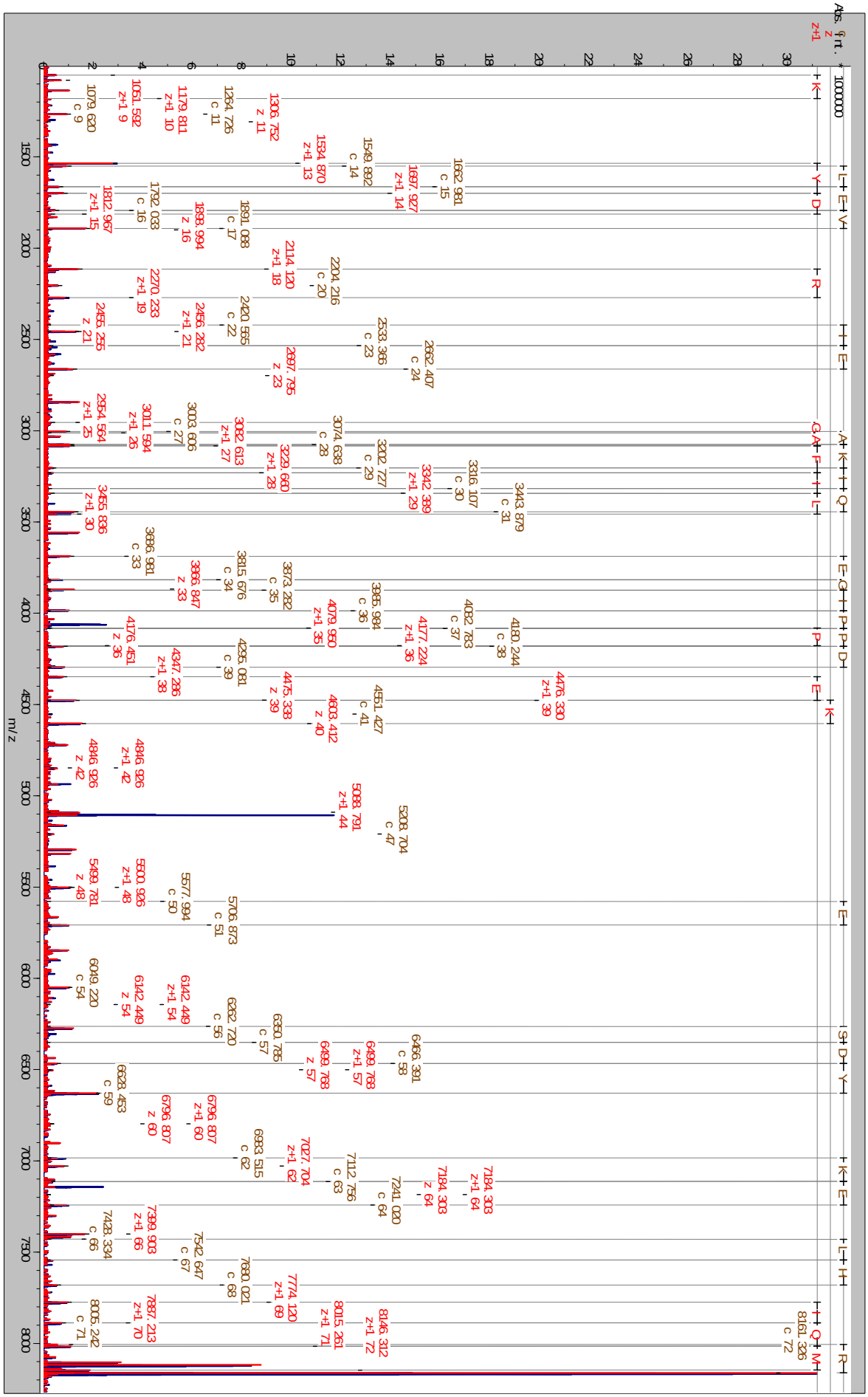


Figure 99: Mass matches of the 8164Da precursor to the amino acid sequence of the *Leishmania* ubiquitin fusion protein. z and z+1 ion matches are coloured red, c ion matches are brown.

Interestingly, the same ubiquitin fusion protein was identified in the *Leishmania* analysis described in section 3.2.3.11.2 using a bottom-up approach as the final analysis step. In that instance, matches to three peptides were observed, all of which occurred within the region covered by the top-down approach. Of course, in a peptide based analysis, it is impossible to determine whether missing peptides from the C terminus of the protein occurred due to post-translational truncation, as in this case, or merely failure of the mass spectrometer to identify those peptides.

Several other proteins were dissociated using either ECD or CID, providing significant numbers of fragments. Unfortunately, although in some cases short sequence runs could be observed, none were discriminatory enough to provide a shortlist for identifications, which highlights the critical requirement for software improvements to enable top down analysis.

7.4.1.1.2. Separation of PTMs/Isoforms with Monoliths, Top-Down Analysis

The identification of unknown proteins by top-down analysis is very challenging, but the use of intact protein analysis to characterize known proteins is a significantly easier task, as matching fragmentation data to a known sequence is substantially more straightforward. Protein characterization can be performed to identify PTMs and different protein isoforms, but in many cases a mixture of differentially modified species are present, making resolution of individual isoforms difficult. The application of high-resolution protein chromatography to these samples can provide an added dimension to reduce the complexity of the sample and generate fractions containing different species. This method of analyis was applied to histone proteins, which exist in a number of isoforms consisting of multiply methylated and acetylated species.

2 picomoles of histone sample were placed in a 96 well plate and injected directly onto a 200 µm x 5 cm monolithic column. Eluent was electrosprayed using an Advion NanoMate directly into a 12Tesla Apex IV FT-ICR MS (Bruker Daltonics). A total ion chromatogram for the separation is shown in Figure 100. The first two large peaks in the chromatogram are truncated forms of histone H4, with the third peak being intact H4. Peak widths are ~10 seconds, except for the H4 peak which is substantially broader. This additional peak width is due to partial separation of histone methylation and acetylation states (see Figure 101). Increasing methylation and acetylation of histones masks charged residues (lysines)

on the surface of the histone, resulting in greater hydrophobicity and subsequently tighter association with the monolithic column stationary phase.

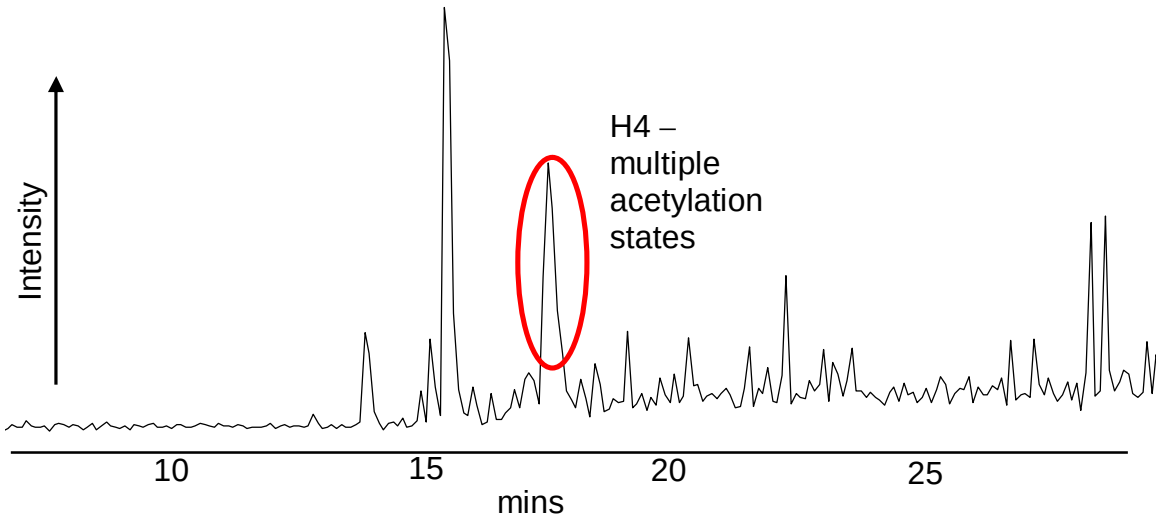


Figure 100: Base peak chromatogram of histone H4. The large peaks correspond to H4 fragments except for the circled broad peak, which denotes intact histone H4. Shouldering of the peak is due to the partial separation of acetylation and methylation states.

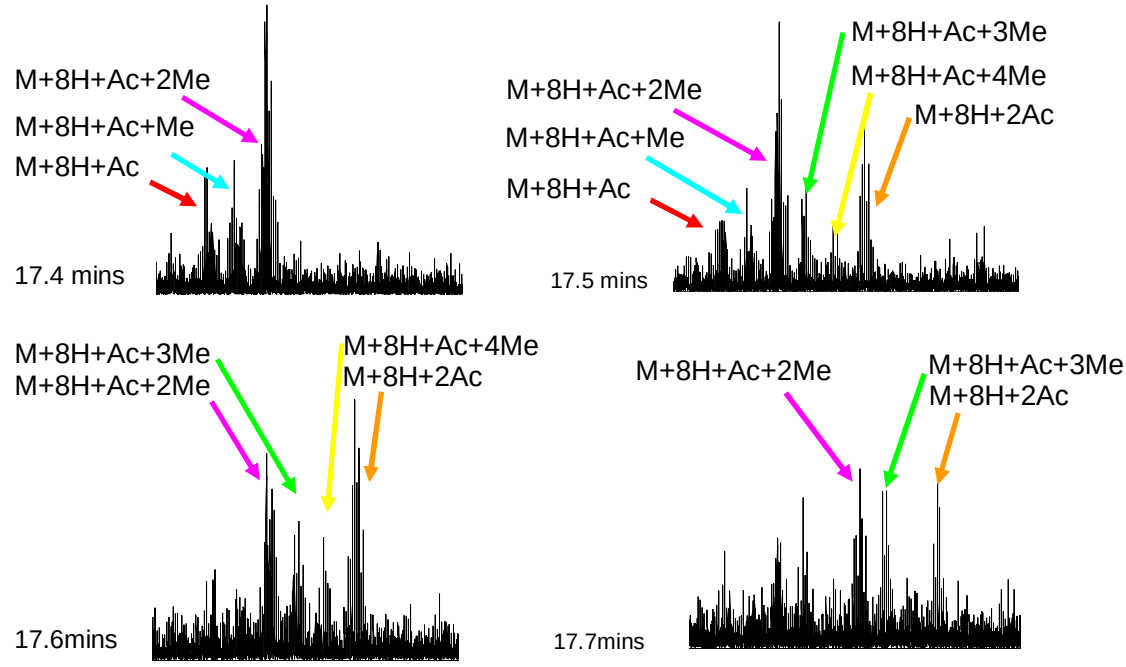


Figure 101: Separation of histone H4. M denotes the unmodified mass of histone H4. Each panel represents 0.1 minute steps through the chromatogram. First visibility of H4 peaks occurs at 17.4 minutes in the separation, and each subsequent scan shows a shift towards higher acetylation/methylation states (subsequently confirmed by MS/MS analysis).

7.4.1.2. Label Free Quantitation for the Assessment of Histone Modifications.

As discussed in the chapter on general quantitation for intact protein separation, label free methodologies are restricted for multidimensional analyses, due to the complexities this raises. A single one dimensional separation is generally insufficient for even a handful of proteins when applied to the 'bottom up' approach, but by maintaining low complexity by using top down analysis, low complexity samples, such as histones, are a promising area of research for label-free quantitation.

In collaboration with NonLinear dynamics, a pre-release version of the Progenesis LC-MS label-free quantitation software has been used to assess quantitative differences between the acetylation and methylation states of a mixture of histones from young (6 week old) and aged (3 month old) mice. FT-ICR of histones was produced in the same manner as for other histone samples in this study, by reversed phase monolithic LC, by collaborators at the SIRCAMS facility in Edinburgh. Multiple runs (three for old mice, 5 for young mice) were analysed by the author using the software, and run-to-run matching was performed in the same manner as for 2DGE matching software, using multiple visual aids to enhance user interaction.

Significant amounts of smaller molecules were quantified using the sample, as well as multiple instances of histone proteins. The resolution of the image generated is insufficient to show isotopic resolution of the histones, and as such they appear as abundant single spots, whereas lower charge-state species are represented as oval patterned streaks (see Figure 102). This is a limitation of the software, in that it looks for gel-like 'spots' which usually correspond, on lower resolution instruments, to peaks, whereas the superior resolution versus time of FT-ICR instruments provides peaks that are represented by oval-shaped sets of vertical bars (the oval shape is a result of the lower intensity of the more less abundant isotopic peaks, as they appear and disappear from the trace more rapidly due to their lower signal to noise), and the vertical bars are a consequence of the very high resolution compared to short chromatographic runs. Mass is represented on the x axis, and time is represented on the y axis, in contrast with most other software.

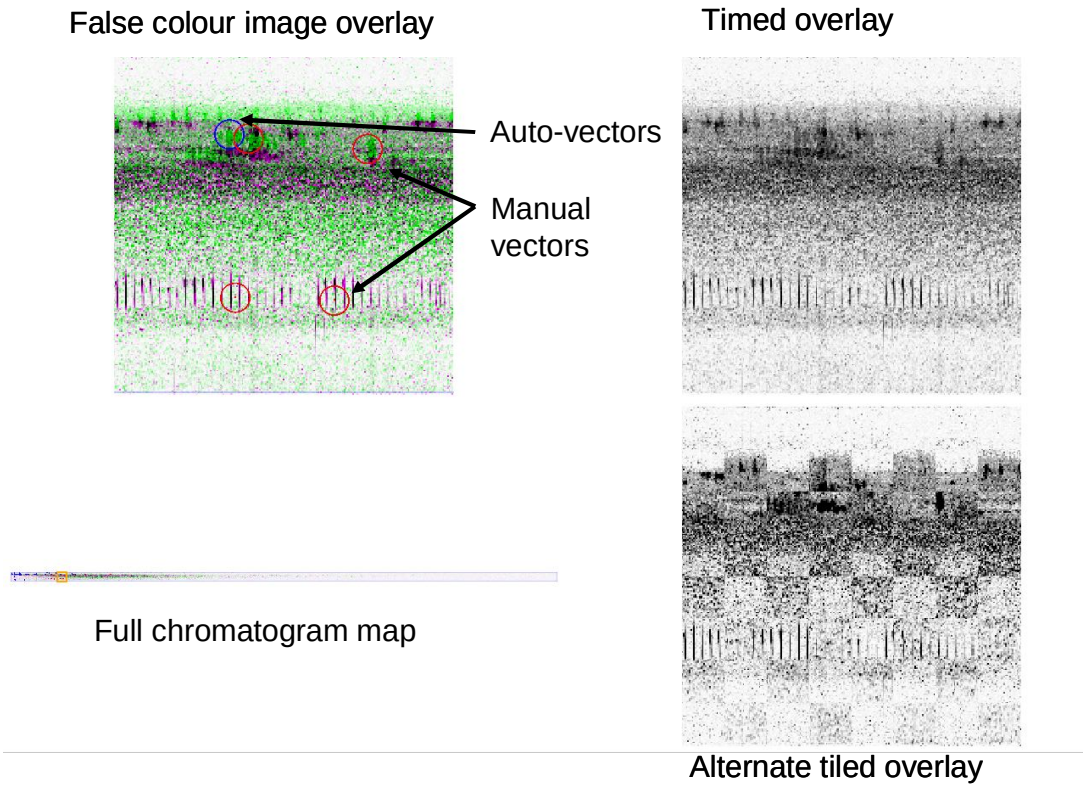


Figure 102: Overview of matching software applied to differential histone samples. Chromatograms are represented as 'gels' with chromatographic time on the y axis and m/z on the x axis. Peaks correspond to spots of intensity corresponding to ion abundance. Note that the full chromatogram map is very short vertically, corresponding to a short chromatographic separation, but very long horizontally, corresponding to the superior resolution of the FT instrument. Three different views of the data are provided to assist matching of the 'gels'. The false colour overlay shows unmatched areas on one 'gel' as green, and unmatched areas on the other 'gel' as purple. Matching regions are black. Minor shifts in retention time are accounted for by using 'vectors' to show the direction and magnitude of the shift. Vectors can be generated automatically by the software (auto vectors), but are assisted with the use of vectors specified by the user (manual vectors). The 'timed overlay' shifts between one image and another and the 'alternate tiled overlay' consists of small square elements of one gel, interspersed with elements of another gel to assist in matching.

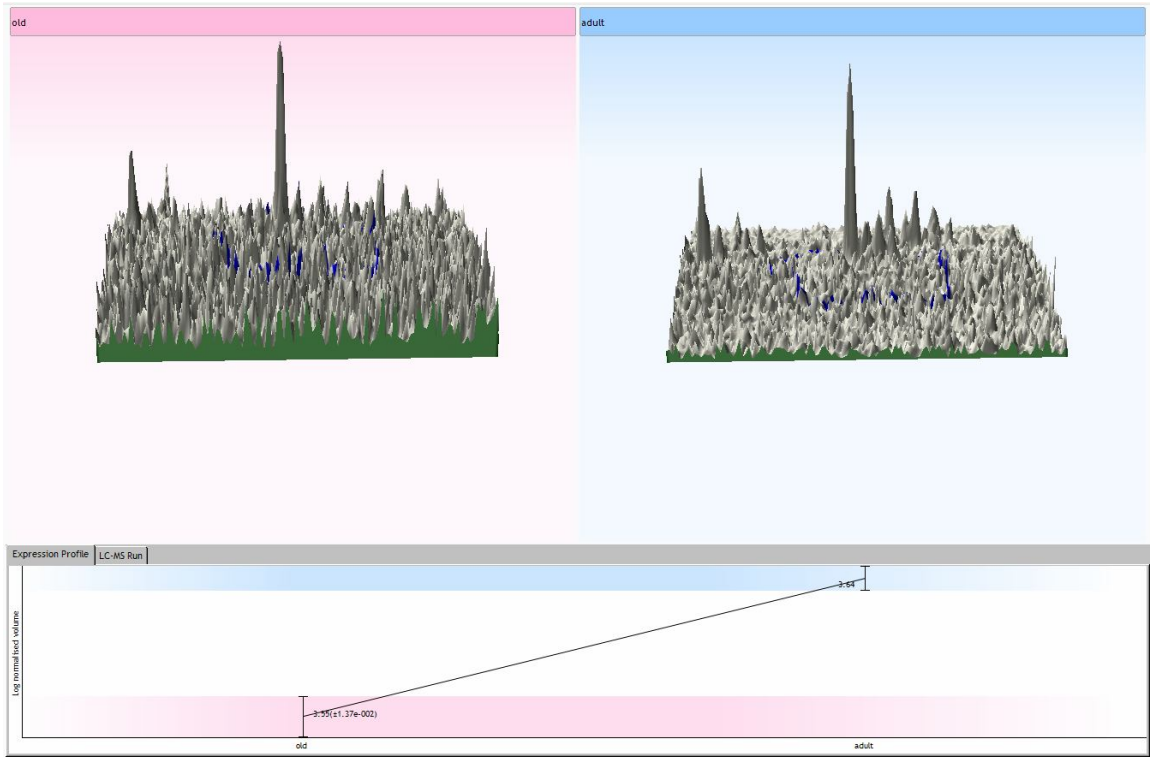


Figure 103: 3D image and graphical display of abundance difference between the cluster of Histone H4 species beginning at 919.5m/z. The distinct row of peaks shown here are evidence of different methylation and acetylation states of the protein. However, these have been grouped together as a single spot, and any information on differential regulation of the PTMs is not observed.

In collaboration with Non-Linear Dynamics, improvements to the software to allow identification of peaks from high resolution data have been recommended, and are now implemented in the new version of the software. True differences in intensity between histone species are difficult to ascertain as the spot detection merges multiple histone species into single identified spots (see Figure 103).

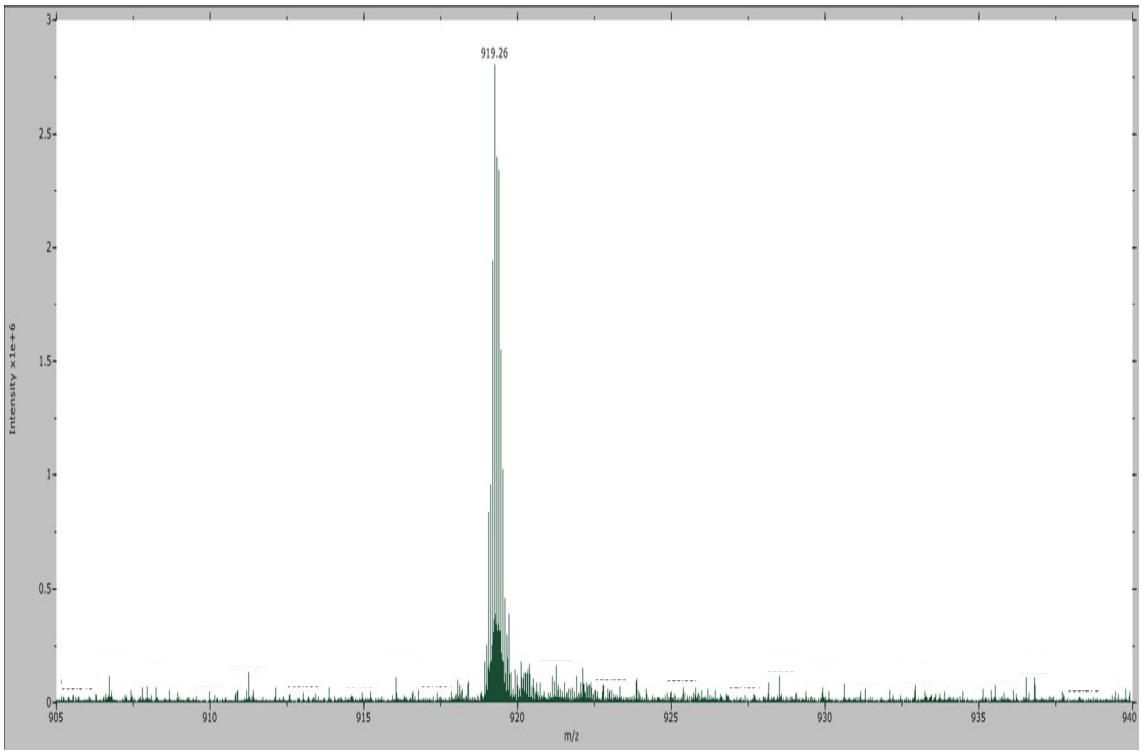


Figure 104: Histone species identified as being consistently differentially regulated 1.2-fold down in three month old mice, as compared to 6 week old mice.

For example, the cluster of peaks from 919.5 to 922.5 was found to decrease 1.2 fold (ANOVA score of 2.73×10^{-7}) in abundance in old mice, as compared to adult mice (see Figure 104). Although interesting, the experiment in general was focussed on PTMs, especially changes in methylation and acetylation of histones due to age. Due to the grouping of different isoforms of histones together into single spots by the software, analysis of specific PTM states was not possible. The reason for the grouping is likely to be the highly charged nature of the proteins, and the small mass of the modifications (which the 9-14+ charge state of the protein causes to appear as roughly 1 m/z differences), so the algorithm that groups isotopic peaks into a single spot recognises the set of histone isoforms as a set of isotopic peaks and groups them together as a single spot.

Histone PTM analysis using FT has been performed extensively by Dr. Kelleher's group. For example, a paper on Histone H3 by Thomas [254] discusses the methylation and acetylation states, which can be elucidated by a combination of ultra-high resolution MS and ECD fragmentation of the different isoforms under a variety of conditions. Separation of the histones was performed on a C18 column using a gradient from 30-60% v/v acetonitrile over at least 80 minutes and provided only partial separation of the species, and a further dimension of weak cation exchange was performed to additionally separate acetylated species. Of course, due to the salt content of the resulting fractions, further purification and desalting must be performed post cation exchange. The fact that

monolithic columns provide partial separation of histone isoforms on a one dimensional 10 minute gradient demonstrates the significant reduction in time and complexity provided by this methodology. The application of an additional dimension of weak cation exchange prior to RP separation using either derivitised monolithic or pelicular PS-DVB stationary phases suggests that improvements could be made to the process, but this will, of course provide an additional requirement of separation time. The use of RP as a second dimension allows direct interfacing with MS, as well as a convenient MS compatible solvent allowing direct infusion post fraction collection.

As a contrasting example, papers on histone H2B [255], histone H2A [261]and a recent paper on histone H4 modification state during the cell cycle [256], describe purification of the histone samples with the use of standard reversed phase gradients from 5-90% v/v acetonitrile (in the case of the former) and from 5 to 85.5% v/v acetonitrile (in the case of the latter). H2B is described as eluting as a single peak, and therefore further purification was not required. H2A's elution profile is not described, but H4 is described as eluting in a single peak containing multiple isoforms, which were pooled resulting in contamination with other histones. It is clear that in some circumstances, histone proteins can achieve adequate separation with the use of a single (long) dimension of standard column chemistry separation, but the convenience of monolithic separation means that multiple histone species and isoforms can be separated over a single 10 minute gradient.

7.4.1.3. Top Down Analysis with Non FT Instruments

The limitation of non FT instruments for top down analysis is not that proteins cannot be observed, or that fragments cannot be generated, but that the resolution of the instrument is insufficient to confidently identify the charge states and therefore the molecular masses of protein fragments. Analysis of fragmented proteins therefore requires reducing the charge states of the fragments to the point where they can be confidently determined by the instrument in question. Although ETD has recently been applied to Q-ToF instruments [262, 263], the original development work was performed on an ion trap type instrument [110], and these remain the only instruments commercially available with an ETD CI source.

Although the original description of PTR by Coon [119] required the addition of a second chemical ionisation source to the standard ETD source, this requires a significant modification to the instrument. An elegant solution developed by Ralf Hartmer at Bruker

Daltonics [120] uses the standard fluoranthine ETD reagent under a higher excitation voltage to generate a fluoranthine anion – an excellent proton acceptor, which fulfils the same function as the benzoic acid derivative does for the system described by Hunt, thus providing fragmentation and charge stripping from the same CI source. This is implemented as a software upgrade to the standard HCT spherical ion trap with ETD. In practice, the mass range of the HCT (from 50 to 3000 Da) in 'enhanced' mode provides a scan rate of 8000 amu per second and a resolution sufficient to identify 3+ ions. This provides a theoretical maximum of 100% sequence coverage at 2 x 9000 Daltons assuming that series of c and z ions can be determined, or a maximum coverage of larger proteins equivalent to 9 kDa in from each terminus.

Although ETD is a translation of the ECD fragmentation available on FT-ICR instruments, fragmentation with ETD was theorised to be significantly more efficient than ECD [264]. To test this, the efficiency of fragmentation of cytochrome c was assessed: 2 μM cytochrome c was electrosprayed into the MS (see Figure 105), the most abundant, highly charged peak was isolated with a mass window of 2 amu (see Figure 106) and ETD was performed on it (see Figure 107). Prior to fragmentation the intensity of the precursor (averaged over a period of two minutes) was $8x10^6$ counts. After fragmentation, the precursor's intensity (again averaged for two minutes) was $8.75x10^4$, thus demonstrating that fragmentation in this case is 98.9% efficient (unless leakage from the trap has occurred). The total intensities of the precursor and charge-stripped species are $8.75x10^5$. If the charged stripped species are included in the efficiency calculation (since they do not provide any sequence data), fragmentation efficiency decreases to 89.06%, which is significantly higher than the ~30% for ECD.

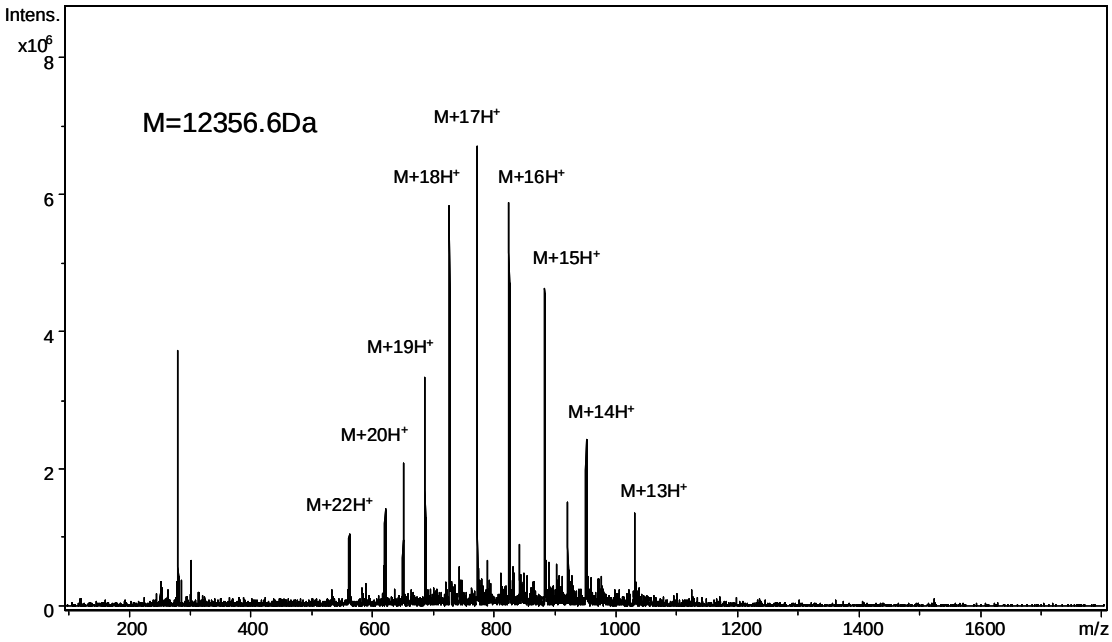


Figure 105: Spectrum derived from infusion of 2µM cytochrome C to a spherical ion trap. Note the high charge states in the characteristic 'mass envelope'.

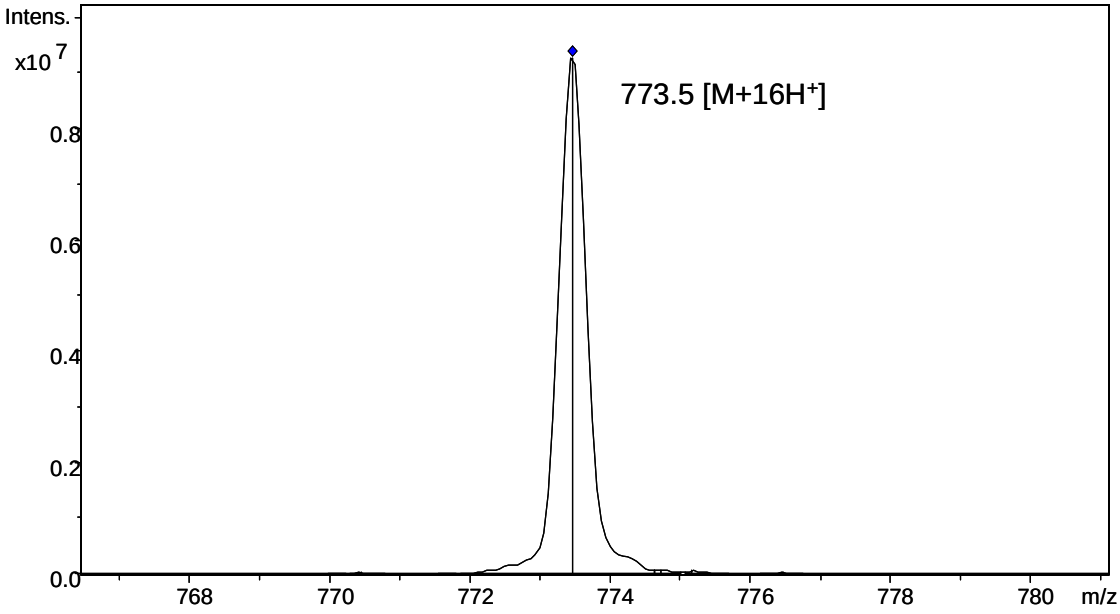


Figure 106: Isolation of the 16+ charged peak from the charge envelope shown in Figure 105. Note the lack of isotopic resolution. The 773.5m/z peak was chosen as it is the most consistently abundant ion observed.

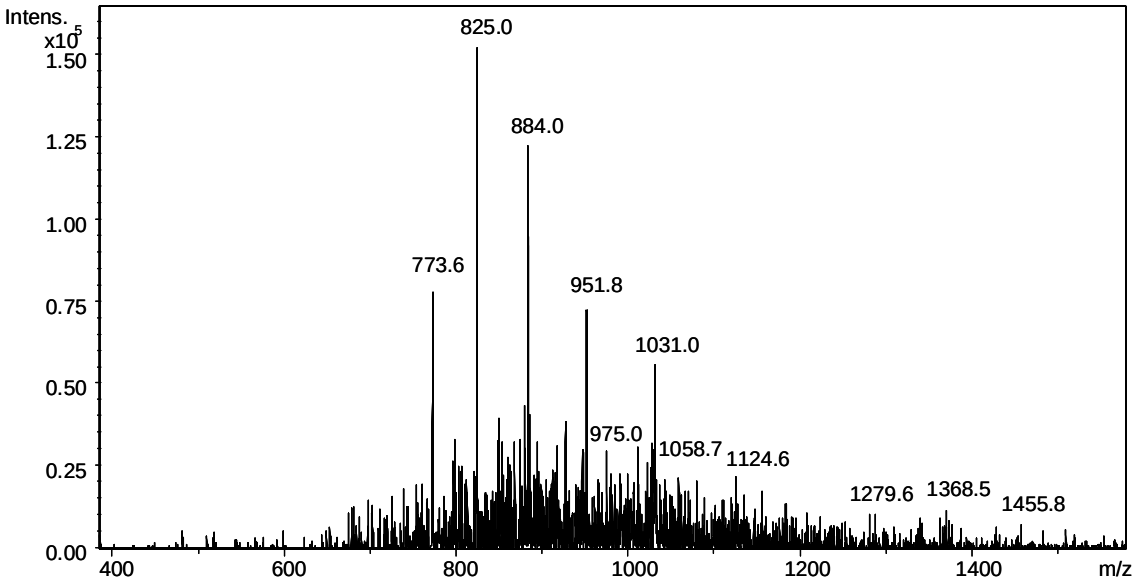


Figure 107: ETD fragmentation of the 16+ charged precursor at 773.6m/z from cytochrome c. Note that the precursor is still present, and that the spectrum remains dominated by 'charge reduced' peaks, but the intensity of the precursor has decreased from $8x10^6$ to $8x10^4$ counts.

Due to the increased sensitivity and higher efficiency of the ion trap as compared to FT instruments, it speculated that sequence data for small intact proteins could be obtained on a chromatographic time-scale.

Udeshi et al. [265] demonstrate ETD/PTR analysis of ribosomal proteins on a chromatographic time-scale using their modified LTQ instrument. For their analysis, a 90 minute gradient was applied to an unspecified (probably C4, based on the broad width of chromatographic peaks) column. MS was collected for 3 seconds per scan, followed by ETD/PTR on the six most abundant precursors for 3 seconds. Analysis of the fragmentation spectra was performed by denovo sequencing and manual assignment of ion series. Successful identification was made of 46 of the 55 proteins in the Escherishia coli ribosome. It is important to note that with a search space of only 55 proteins for analysis is three orders of magnitude lower than for a complex lysate, and this technique remains unfeasible for general proteomics analysis.

Injection of 2 picomoles of cytochrome c onto a monolithic column provided a sharp peak of focussed protein (see Figure 108). Current software does not allow automated ETD coupled to PTR, so the retention time of the protein was noted, and a second separation was performed. ETD and PTR were initiated at the appropriate time and a fragmentation spectrum was obtained. The technique was then applied to a sample generated from a SILAC labelled lysate. The sample was initially separated by ion exchange and 20µl of the 200 obtained from each fraction was injected onto the ion trap (see Figure 109). A

candidate small protein, part of a clearly distinguishable pair, was chosen for online fragmentation, the retention time of the protein's elution was noted, and a fragmentation spectrum was generated from the 11+ peak. Although a number of fragments were generated (see Figure 110) and 33 peptide masses were able to be deconvoluted, insufficient data was obtained to produce a match in Prosight PTM against the human database.

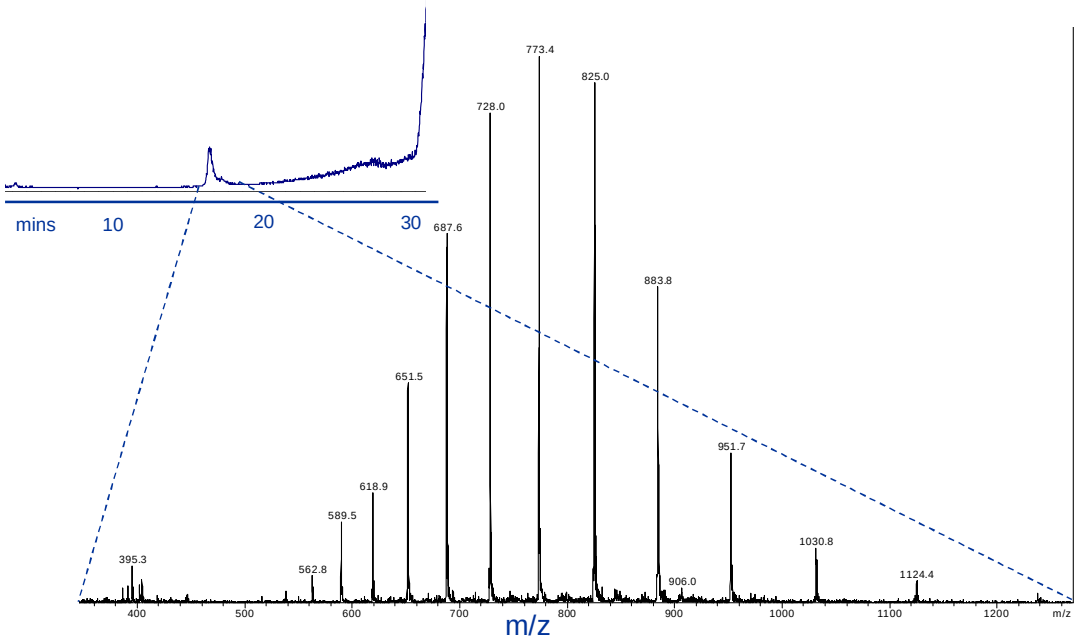


Figure 108: On-line spectrum of 2pmol cytchrome c, focussed by PS-DVB monolithic LC. Inset at the top left is the total ion chromatogram corresponding to the separation.

Attempts were made to achieve analysis of Histone H1 by ETD/PTR using the HCT ion trap, but the heterogeneity of the protein resulted in poor stability of the multiple overlapping charge envelopes. In addition, the mass of H1 is ~20 kDa, meaning that it is far outside the 9000 Da full coverage envelope, and thus poor data was returned upon ETD fragmentation and PTR (data not shown). Histone H1 is constitutively phosphorylated, in addition to possessing multiple acetylation and methylation modifications. A paper by Garcia [257] describes the analysis of PTMs in histone H1, applying a C8 column to the separation of the different isoforms. A gradient was applied to the loaded sample from 5 to 70% v/v acetonitrile, followed by analysis by reinfusion of dried samples onto an FT ICR instrument. H1 is an interesting target for separation prior to analysis, and further work may be performed using multidimensional separations to effectively reduce the complexity of the sample prior to analysis.

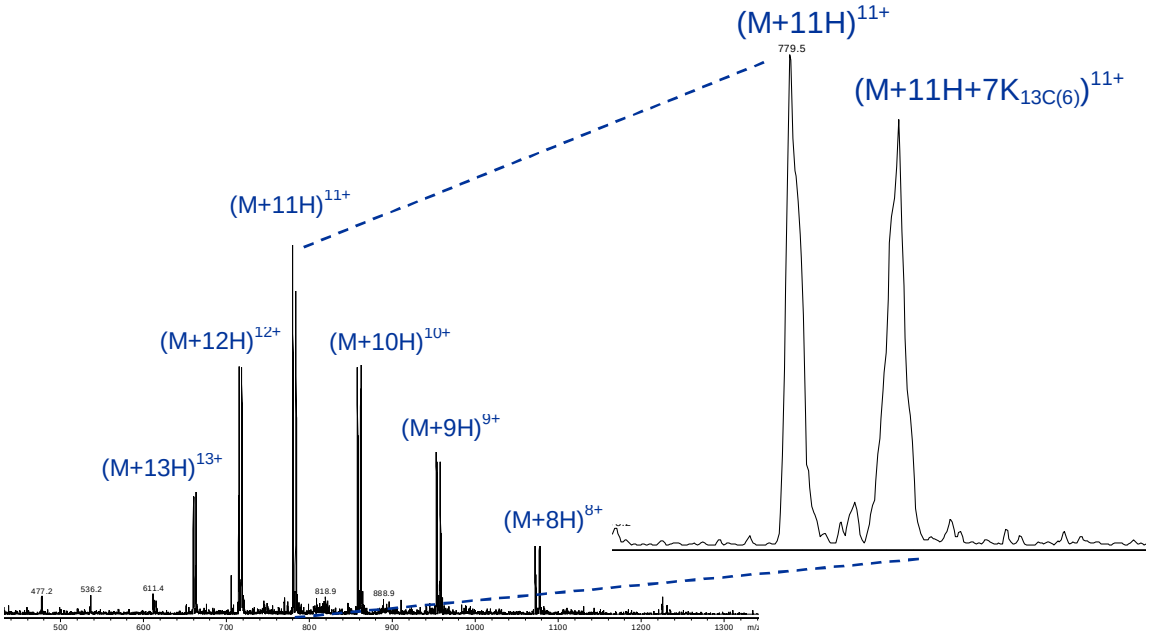


Figure 109: On line analysis of a SILAC labeled human protein from the washthrough fraction of an ion exchange separation of K562 cells. 5 µL of the 200 µL total fraction was injected onto a monolithic PS-DVB reversed phase column. An XIC of the protein's elution is shown in the inset in Figure 110.

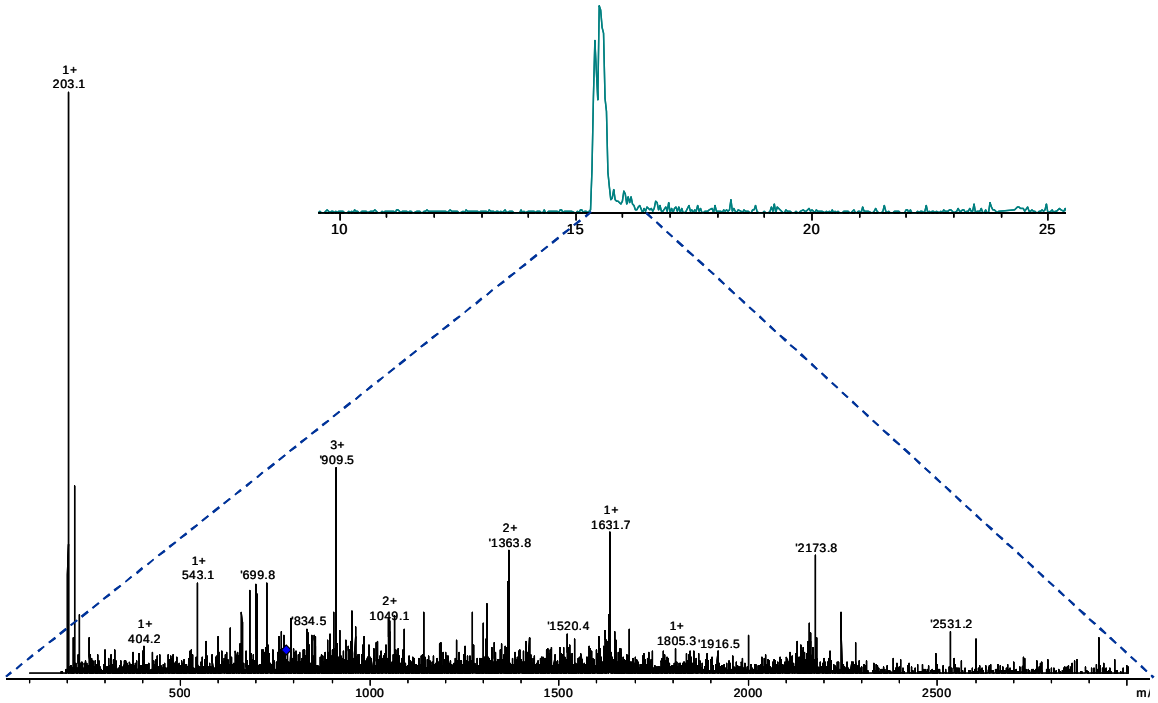


Figure 110: On line fragmentation of an unknown SILAC labelled human protein. Note was taken of the elution profile of the protein in an initial MS only separation, and fragmentation was initiated at the appropriate time in a second separation. Although a number of highly intense fragment ions were generated, the majority of peaks could not be assigned with high significance to a specific protein. The intense 203.1 m/z peak corresponds to the PTR reagent, which was not specifically excluded from the spectrum.

A general observation while performing post-reversed phase fractionation was that some species, while exhibiting superlative signal-to-noise in the chromatographic separation would, when reinfused, be so low in intensity that they were almost indistinguishable from system noise, or not visible at all. Some loss of intensity was expected, because comparison of the apex of a chromatographic peak with a solution containing a uniform

concentration of the same species will almost always lead to a deficiency in the concentration of the latter, due to dilution, however, small contaminant species at a low intensity co-eluting with the protein peaks were often enhanced in intensity when re-infused, and some protein species showed equal concentrations in comparison to chromatographic peaks.

It is therefore likely that precipitation, degradation or sequestration of the protein by the surface of the microplate wells is occurring, and the physicochemical properties of some proteins result in their resistance to this process, allowing further analysis. Freeze-thawing of samples is likely to result in degradation, and it is assumed to be for this reason that the fractionations performed by Dr. Kelleher's group [187] are maintained at a constant temperature of 4ºC. However, another factor is likely to be the concentration of organic solvent in the fractions. Polson [266] describes the efficacy of acetonitrile as a precipitant for proteins as being 50% of protein even at 33% acetonitrile, and since protein separations are performed with monoliths with an acetonitrile concentration ranging from 15 to 70%, the combination of leaving the collected fractions in a chilled environment, and the acid present in the samples (which can also be used as a precipitant) is likely to have a significant effect on the solubility of proteins. This is unlikely to be a problem during chromatography of the proteins, as they are loaded from a large volume in aqueous solvent, and after elution, transfer to the mass spectrometer is performed in a matter of seconds at 60ºC.

Further research must be performed to determine if these effects are causing the loss of signal intensity, and to provide steps to ameliorate the process.

7.5. Conclusions:

FT-ICR MS can be used to perform fragmentation analysis of intact proteins. This is enhanced by the ability to separate proteins in high resolution while they remain in solution – something unique to this methodology. Although the scan rate and efficiency of FT-ICR MS are insufficient to analyse proteins on-line, post column splitting provides targets for fraction re-infusion, and longer-duration analysis. Multidimensional separation of proteins using PS-DVB based columns coupled to FT-ICR MS is sufficient to allow analysis of the smaller proteins from a complex mixture such as the lysate of the parasite *Leishmania donovani*.Additionally, PS-DVB monolithic columns allow for fast and efficient separation of histones, with which partial separation of methylation and acetylation states can be

observed. Coupled to high resolution FT-ICR instruments, isotopic resolution and hence accurate mass of proteins can be obtained from on-line separation and MS detection while collected fractions can be infused for further analysis including fragmentation to provide sequence data.

ETD coupled with PTR implemented on an inexpensive, low resolution ion trap instrument allows the analysis of intact small proteins. Although further work is necessary to optimise conditions and sample preparation, the combination of high resolution separation with inexpensive top down analysis instrumentation becomes a powerful alternative to the traditional 'bottom-up' proteomics analyses.

8. Chapter 8 – General Discussion

The general aim of this research was to determine whether new column technologies for intact protein separation could be implemented as a general proteomic technology.
To be accepted as a general method for proteomics, the technique developed must show significant improvements over existing methodology, be complementarity to existing methods, or facilitate analyses that were previously not possible.

For this reason, it was important to ensure that the complete analysis time was kept to a minimum, that attempts were made to automate as much of the process as possible, and that methods were compatible with standard workflows such as quantitation and complex samples.

The findings from the project may be summarised as follows:

- High resolution, multidimensional separation of intact proteins can be performed by both automated and semi-automated processes.

- Sample preparation conditions are equivalent to that of two dimensional gel electrophoresis.

- Optimised lysis, resuspension, fractionation and digest conditions were developed, and are responsible for large improvements in the protein identification rate.

- Quantitation of samples has been shown to work with iTRAQ, ICAT, SILAC and label free methodologies, and evaluation has been performed on other labeling techniques that proved less successful.

- A complete experiment, including two protein separation dimensions, and one peptide LC/MS dimension may be performed in one week.

- The technique is applicable to top-down analysis using both traditional, high resolution FT-ICR MS and new methods such as ETD/PTR.

When discussing the methodologies and data presented in this thesis, it is important to examine them in the context of biology, and specifically biochemistry. As biochemists, our

desire is to understand the totality of the processes and mechanisms of living creatures, because understanding provides us with the means to modify and repair such processes. Analysis of an organism as a whole provides a very crude picture of the processes involved; the heart, for example, according to Aristotle, was the seat of reason, rather than the brain. Until Galen performed his pioneering dissections of mammalian bodies, therefore, a number of different theories were entertained as to the functions of various body parts. Thus, to understand a process or a mechanism, a reductionist approach, where components are dissected or separated so that they can be examined individually, is essential. On the molecular level, this is also clearly true, as Michaelis and Menten's pioneering work on enzyme kinetics was performed by the analysis of reaction rates of a single enzyme (the breakdown of sucrose to glucose and fructose by invertase) [267] and the application of different concentrations of purified substrate to analyse its activity. Once components are dissected, they can be more easily identified from the plethora of other components of similar properties, and assigned a name and function. Currently and historically, therefore, to perform any sort of biological analysis a form of separation is required.

In the context of protein chemistry, a variety of chromatographic and electrophoretic approaches can be used to separate different proteins. For purification, the most successful of these are affinity columns, but as affinity columns depend upon specific interaction, the generation of an appropriate capture molecule increases the time required to perform a purification significantly. A single separation method possessing resolution sufficient to separate every protein in a proteome distinctly is not yet available, because, unlike DNA, proteins have a variety of physiochemical properties commensurate with their myriad of functions within (and without) a living organism, and proteins with divergent functions may have the same properties. In addition, each gene in a genome persists in the same number of copies as other genes and there is no ability to amplify low abundance proteins for analysis. It is fortunate, therefore, that mass spectrometry is available to provide an accurate final dimension of high resolution separation and to enable rapid identification of protein species.

For relatively low complexity protein samples (those that have been produced from already fractionated samples, for example) a single dimension of separation may be sufficient, but for higher complexity samples such as cellular lysates, or whole organism 'grind and find' experiments, the probability of two or more species with the same mass and response to the initial separation mechanism becomes very high. Each time an additional separation

technique that separates by another physiochemical property is added to an experiment, the deconvolution of species is improved by the resolution of the new separation technology, and it is for this reason that orthogonal separations are available to allow multidimensional separation of proteins. However, each additional manipulation results in loss of sample and therefore quantitation becomes vital to effective analysis. Two dimensions of separation are commonly in use, due to the exponential increase in analysis time resulting from additional dimensions.

The techniques providing major competition to multidimensional LC of proteins bear reiteration, and are:

- Multidimensional LC of peptides
- SDS PAGE followed by digestion and LC of peptides
- 2D PAGE of proteins
- Other methods for multidimensional LC of proteins e.g. the PF2D system (Beckman)

8.1. Multidimensional LC of peptides

Multidimensional LC of peptides was pioneered by Yates [21, 24], and uses the same chromatography chemistry as the primary method described in this thesis, except that conventional resins are employed rather than polystyrene-based ones. As a methodology it has clearly been successful, with a huge number of experiments published using the technique. It does, however, have several issues, not least of which is that initial digestion of the proteins involved results in mixing of truncation or otherwise modified products with their parent proteins, and spreads the peptides from high abundance proteins across the gamut of the separation, resulting in ion suppression. For complex samples, however, it is a rapid method of separation, ranging from 12 hours [268], through 24 hours [269], to 2.8 days [270] for a complete experiment. In fact, as the final peptide separation often performed in 2D intact protein separation is digestion and peptide separation, peptide 2DLC is one dimension fewer, leading to a much lower resolution but significantly faster separation. As described in Chapter 3, 2D intact protein LC is a complementary approach to 2D peptide LC. In the experiments described, protein identification numbers were roughly similar, but the many improvements made to intact protein separation methodology since that comparison was performed should now allow a more favourable comparison.

8.2. 1D PAGE followed by LC/MS

This technique, otherwise known as GeLC, is normally a hybrid protein-peptide method in two dimensions, thus it has the benefit that protein isoforms can be separated by mass. In addition, it possesses rapidity on a similar time-scale to peptide 2DLC. For top-down analysis, as exemplified by Kelleher, preparatory continuous elution PAGE has been used to provide liquid fractions for further protein separation by C4 columns [177, 259].

Again, this is one of the more popular methods for proteomic analysis [271-273] and has been previously coupled to PS-DVB monoliths for fast peptide separation [215, 216]. Its popularity stems from its great simplicity, especially that most biological researchers are familiar with the preparation and running of SDS gels. Additionally, the compatibility of the separation method with SDS allows resuspension and separation of more hydrophobic proteins than other methods. In the comparison described in chapter 3 1D PAGE was again shown to be a complimentary method to 2D intact protein LC, and the number of identifications was not significantly different.

8.3. 2D PAGE

The highest resolution protein separation technology currently available for general samples remains 2D PAGE, and this is almost entirely due to the fact that it is analogue in both dimensions. Any separation technology relying on fractionation will be restricted in resolution by the number of first dimension fractions taken. To ameliorate the loss of resolution caused by a low number of fractions, a fraction number significantly higher than the expected resolution of the peaks could be collected, but this would proportionally increase the time taken for subsequent analysis, a drawback that 2D PAGE conveniently avoids. In addition, 2D PAGE can be parallelised very easily with the use of multi-gel tanks, and due to the increasing requirement for statistical rigour in proteomics experiments, the ability to perform multiple separations in a single gel tank is a significant benefit, which cannot be replicated with LC except by the use of multiple HPLC systems. However, matching gel replicates for quantitative analysis, even when the separation is performed in parallel in the same tank, requires complex software and manual intervention, which would suggest the use of a protein labelling technology to allow comparisons within a single gel. The commercially successful DiGE labelling technique relies on fluorescent

dyes to observe quantitative differences on the same gel. The final protein analysis, however, relies on matching a fluorescent dye image to a differently stained (usually Sypro Ruby or Orange) protein gel image, which is significantly different, and again requires significant manual intervention. In addition, multiple protein identifications are usually obtained from a single spot and it is therefore erroneous to assume a particular identified protein is responsible for the quantitative difference. A hybrid approach, which would involve minimal labelling of the gel with CyDyes in conjunction with isotopic labeling, would allow confirmation of the differential protein. This approach would couple the great sensitivity and accurate quantitation provided by the CyDyes with the ability to unambiguously identify the protein responsible for that quantitation. It would be important, however, to normalise the data using a number of spots of equal fluorescence with both dyes, in case one isotopic label does not label as efficiently as the other.

In regions where a large number of protein spots bleed together (in the middle of the gel, for example, there is often a morass of charge trains from multiple isoforms of proteins of similar mass and pI) fluorescence may not be sufficient to provide accurate quantitation, and hybrid fluorescence and isotopic labelling may be very beneficial. In fact, it is tempting to simply analyse the entire gel by isotopic quantitation: a reduction in observed dynamic range is inevitable, but normalisation can be performed across the whole experiment, and quantitative differences not observed using Cydyes are likely to be detected. The most sensible way to analyse the entire gel would be to section it, in other words to perform a fractionation, and thus the high resolution of the gel is reduced to the resolution of the number of slices taken. Rather than performing this type of experiment using 2DGE, it is therefore tempting to use the intact protein separation method described in this thesis, because, due to the improved automation available with LC methods, intact protein separation using LC can easily provide equivalent resolution to a pixellated 2D gel with greater convenience. In fact, with the use of 8plex iTRAQ reagents, 4 replicates of control and test samples can be separated simultaneously under identical conditions, saving a great deal of time.

8.4. PF2D

The most similar alternative to the technique described in this thesis is the PF2D system. The PF2D system has been the subject of a number of papers (for example: [40, 170, 172]), and, as described previously, relies on chromatofocussing, followed by reversed phase separation on a non-porous silica column. Apart from the comparatively poor

resolution of the first dimension, and the short lifetime of the columns (5 separations for the ion exchange column), the PF2D provides a good platform as a dedicated system for intact protein separations by LC. However 2D intact LC using PS-DVB columns provides substantially greater robustness with no observable column deterioration resulting from three years of near-daily use. In addition, the methodology can be applied to any type of HPLC system capable of providing the flow rates required for each dimension (100-300 μL/min and 10-30 μL/min depending on user preference), with the added benefit that the same system can be reconfigured for other uses if necessary.

8.5. Overall analysis of protein 2D separation

In balance, in its current state, two dimensional separation of intact proteins by LC is a complementary method of separation when compared to other common methods. The resolution improvement due to the PS-DVB columns is substantial, but due to the necessity of fraction collection and the sample manipulation performed at each stage, it is suggested for use in specific situations, where lower complexity samples with substantial post translational modifications require separation. This is particularly applicable for proteins that are to be analysed by top-down, as extraction of substantial concentrations of protein from 2D gels is not trivial. Using pISEP as an alternative first dimension allows the restriction of the range for separation to a specific set of pIs, which, with a quoted resolution of 0.01pH units, would allow focussed separations of closely related species.

The technique is more limited than expected for very complex mixtures, as to achieve high resolution separation requires the collection of several hundred fractions, which must then be analysed serially by LC/MS since each fraction will still contain multiple proteins, especially in the central portion of the separation. Fast peptide separations, as discussed in chapter 5, can speed up the process of analysis by two (for direct injection) to four (for parallel LC) times, but more than one week per experiment is still necessary. One situation where separation of complex mixtures would be very desireable is for top-down analysis, as the lower minimum protein requirement (as compared to similar methodologies) and high resolution of both dimensions are a major benefit. Time taken to analyse the proteins is considerably decreased, as there is no digestion followed by third dimension peptide separation. With current instrumentation, an ion trap equipped with ETD and PTR is the most appropriate candidate for top-down in chromatographic time, simply due to the poor efficiency of ECD and the general insensitivity of FT-ICR instruments, although the comparatively poor resolution of ion traps poses major difficulties for analysis of any

proteins larger than 10 kDa. In contrast, the on-line top down analyses performed using FT-ICR instruments, as described by Kelleher [258] and Smith [274], rely on comparatively large amounts of sample compensating for poor chromatography and a low scan rate, and even under these conditions, the number of proteins identified is very small in comparison to bottom-up approaches. In the near future, the addition of ETD sources to hybrid FT instruments, such as the orbitrap or the LTQ-FT, or high resolution ToF instruments, may allow effective top down to be performed on a chromatographic timescale with high resolution and high fragmentation efficiency, but this is not currently a reality.

During this research in general, replicates have been performed for chromatographic purposes, and for specific analyses, such as the comparison of phosphorylated and dephosphorylated ovalbumin. Replication of the separation of complex mixtures was not intended to be a goal of the project, and, indeed, was unfeasible, due to the extended durations required to completely analyse each sample. The methodology of intact protein separation was in continuous development throughout the project, and each further analysis, even on the same organism, was not directly comparable to others, due to improving lysis conditions, labelling, automation, fraction collection, resuspension, digestion and LC/MS analysis. However, since the aim of the project was the development of a proteomic methodology, each experiment can be considered a replicate of the experimental procedure itself. The multitude of 2D intact protein separations presented in this thesis, from single protein species, such as histones, to multiply isotopically labelled mammalian lysates, demonstrates rigorous testing of the method under as many conditions as were available.

For future experiments, rather than applying the full two dimensional protein separation technology to complex samples, a simpler approach with a single dimension of intact protein separation, followed by in solution digest and peptide analysis, is likely to be a useful tool for researchers. The most appropriate first dimension protein separation (whether ion exchange, pISEP, high-pH monolithic reversed phase, or standard acidic monolithic reversed phase) may be applied to a particular type of sample, based on the properties of the sample and the experimental design. For example, albumin is heterogeneous, and persists in multiple charge states which are normally distributed throughout the fractions on ion exchange runs. However, using monolithic reversed phase columns, albumin is confined to a single large peak, as the hydrophobicity of albumin is less affected by its heterogeneity than its overall charge. For plasma analysis, a simple

separation by RP monolith, followed by tryptic digestion and peptide analysis by conventional C18 column may provide better results than a peptide 2DLC or GeLC experiment. For phosphorylation analyses, high resolution fractionation of ion exchange or pISEP separations, followed by digestion and neutral loss scans to detect phosphopeptides may show differences as isoforms change from fraction to fraction.

Additionally, the techniques for intact protein LC quantitation may be applied to any of the above methods of protein separation, and, in general, an iTRAQ experiment incorporating multiple replicates into a single run is suggested as an excellent method for generating high-quality quantitative data, and will be especially useful in situations where isoform separation is incorporated, such as with pISEP or ion exchange. Indeed, the application of iTRAQ to the type of low-complexity sample described as being the ideal sample type for two dimensional LC of proteins is of great value, as the relative abundances of different isoforms can be elucidated.

In summary, two dimensional intact protein separation by LC using PS-DVB columns is not, perhaps, the new ‘gold standard’ of protein separation, however, there is considerable room for improvement in all of the processes which are incorporated into the method. It is heartening, in particular, to remember that two-dimensional gel electrophoresis has been in continual development for thirty years, and improvements are still being made in terms of range, reproducibility and functionality [19, 20].

9. References

1 Mundy, C. (2001) The human genome project: a historical perspective. Pharmacogenomics **2**, 37-49

2 Hochstrasser, D. F. (1998) Proteome in perspective. Clinical Chemistry and Laboratory Medicine **36**, 825-836

3 Sengupta, N. and Seto, E. (2004) Regulation of histone deacetylase activities. Journal of Cellular Biochemistry **93**, 57-67

4 Villar-Garea, A. and Imhof, A. (2006) The analysis of histone modifications. Biochimica Et Biophysica Acta-Proteins and Proteomics **1764**, 1932-1939

5 Haslam, S. M., North, S. J. and Dell, A. (2006) Mass spectrometric analysis of N- and O-glycosylation of tissues and cells. Current Opinion in Structural Biology **16**, 584-591

6 Spickett, C. M., Pitt, A. R., Morrice, N. and Kolch, W. (2006) Proteomic analysis of phosphorylation, oxidation and nitrosylation in signal transduction. Biochimica Et Biophysica Acta-Proteins and Proteomics **1764**, 1823-1841

7 Chen, F., David, D., Ferrari, A. and Gotz, J. (2004) Posttranslational modifications of tau - Role in human tauopathies and modeling in transgenic animals. Current Drug Targets **5**, 503-515

8 Hsu, W., Rosenquist, G. L., Ansari, A. A. and Gershwin, M. E. (2005) Autoimmunity and tyrosine sulfation. Autoimmunity Reviews **4**, 429-435

9 Bachmair, A., Finley, D. and Varshavsky, A. (1986) Invivo Half-Life of a Protein Is a Function of Its Amino-Terminal Residue. Science **234**, 179-186

10 Ghaemmaghami, S., Huh, W., Bower, K., Howson, R. W., Belle, A., Dephoure, N., O'Shea, E. K. and Weissman, J. S. (2003) Global analysis of protein expression in yeast. Nature **425**, 737-741

11 Sharp, P. M. and Li, W. H. (1987) The Codon Adaptation Index - a Measure of Directional Synonymous Codon Usage Bias, and Its Potential Applications. Nucleic Acids Research **15**, 1281-1295

12 Bantscheff, M., Schirle, M., Sweetman, G., Rick, J. and Kuster, B. (2007) Quantitative mass spectrometry in proteomics: a critical review. Analytical and Bioanalytical Chemistry **389**, 1017-1031

13 Andersen, J. S., Wilkinson, C. J., Mayor, T., Mortensen, P., Nigg, E. A. and Mann, M. (2003) Proteomic characterization of the human centrosome by protein correlation profiling. Nature **426**, 570-574

14 Broadhead, R., Dawe, H. R., Farr, H., Griffiths, S., Hart, S. R., Portman, N., Shaw, M. K., Ginger, M. L., Gaskell, S. J., McKean, P. G. and Gull, K. (2006) Flagellar motility is required for the viability of the bloodstream trypanosome. Nature **440**, 224-227

15 Angers, S., Li, T., Yi, X. H., MacCoss, M. J., Moon, R. T. and Zheng, N. (2006) Molecular architecture and assembly of the DDB1-CUL4A ubiquitin ligase machinery. Nature **443**, 590-593

16 Cottrell, J. S. (1994) Protein Identification by Peptide Mass Fingerprinting. Peptide Research **7**, 115-&

17 O'Farrell, P. H. (1975) High resolution two-dimensional electrophoresis of proteins. J. Biol. Chem. **250**, 4007-4021

18 Gorg, A., Boguth, G., Obermaier, C., Posch, A. and Weiss, W. (1995) 2-Dimensional Polyacrylamide-Gel Electrophoresis with Immobilized Ph Gradients in the First Dimension (Ipg-Dalt) - the State-of-the-Art and the Controversy of Vertical Versus Horizontal Systems. ELECTROPHORESIS **16**, 1079-1086

19 Gorg, A., Weiss, W. and Dunn, M. J. (2004) Current two-dimensional electrophoresis technology for proteomics. PROTEOMICS **4**, 3665-3685

20 O'Farrell, P. (1975) High Resolution Two-Dimensional Electrophoresis of Proteins. Journal of Biological Chemistry **250**, 4007-4021
21 Link, A. J., Eng, J., Schieltz, D. M., Carmack, E., Mize, G. J., Morris, D. R., Garvik, B. M. and Yates, J. R. (1999) Direct analysis of protein complexes using mass spectrometry. Nature Biotechnology **17**, 676-682
22 Lundell, N. M., K (1992) Two-dimensional liquid chromatography of peptides: An optimization strategy. Chromatographia **34**, 369-375
23 Takahashi, N., Ishioka, N., Takahashi, Y. and Putnam, F. W. (1985) Automated Tandem High-Performance Liquid-Chromatographic System for Separation of Extremely Complex Peptide Mixtures. Journal of Chromatography **326**, 407-418
24 Washburn, M. P., Wolters, D. and Yates, J. R. (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nature Biotechnology **19**, 242-247
25 Giribaldi, M., Perugini, L., Sauvage, F. X. and Schubert, A. (2007) Analysis of protein changes during grape berry ripening by 2-DE and MALDI-TOF. PROTEOMICS **7**, 3154-3170
26 Kowalczewska, M., Fenollar, F., Lafitte, D. and Raoult, D. (2006) Identification of candidate antigen in Whipple's disease using a serological proteomic approach. PROTEOMICS **6**, 3294-3305
27 Souza, R. A., Henriques, C., Alves-Ferreira, M., Mendonca-Lima, L. and Degrave, W. M. (2007) Investigation of a protein expression profile by high-resolution bidimensional electrophoresis of Trypanosoma cruzi epimastigotes. Analytical Biochemistry **365**, 144-146
28 Rodriguez-Pineiro, A. M., de la Cadena, M. P., Lopez-Saco, A. and Rodriguez-Berrocal, F. J. (2006) Differential Expression of Serum Clusterin Isoforms in Colorectal Cancer. Mol Cell Proteomics **5**, 1647-1657
29 Beranova-Giorgianni, S. (2003) Proteome analysis by two-dimensional gel electrophoresis and mass spectrometry: strengths and limitations. Trac-Trends in Analytical Chemistry **22**, 273-+
30 Garbis, S., Lubec, G. and Fountoulakis, M. (2005) Limitations of current proteomics technologies. Journal of Chromatography A **1077**, 1-18
31 Rabilloud, T. (2002) Two-dimensional gel electrophoresis in proteomics: Old, old fashioned, but it still climbs up the mountains. Proteomics **2**, 3-10
32 Santoni, V., Molloy, M. and Rabilloud, T. (2000) Membrane proteins and proteomics: Un amour impossible? Electrophoresis **21**, 1054-1070
33 Ong, S. E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A. and Mann, M. (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Molecular & Cellular Proteomics **1**, 376-386
34 Han, D. K., Eng, J., Zhou, H. L. and Aebersold, R. (2001) Quantitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry. Nature Biotechnology **19**, 946-951
35 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. and Pappin, D. J. (2004) Multiplexed Protein Quantitation in Saccharomyces cerevisiae Using Amine-reactive Isobaric Tagging Reagents. Mol Cell Proteomics **3**, 1154-1169
36 Schnolzer, M., Jedrzejewski, P. and Lehmann, W. D. (1996) Protease-catalyzed incorporation of O-18 into peptide fragments and its application for protein sequencing by electrospray and matrix-assisted laser desorption/ionization mass spectrometry. ELECTROPHORESIS **17**, 945-953
37 Reh, E., Hahn, B. and Lamotte, S. (2006) Evaluation of stationary phases for 2-dimensional HPLC of proteins Part 1. Validation of commercial RP-columns.

Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences **844**, 204-212

38 Barré, O., Solioz, M (2006) Improved protocol for chromatofocusing on the ProteomeLab PF2D. PROTEOMICS **6**, 5096-5098

39 Cussac, D., Pichereaux, C., Colomba, A., Capilla, F., Pont, F., Gaits-Iacovoni, F., Lamant, L., Espinos, E., Burlet-Schiltz, O., Monsarrat, B. , Delsol, G., Payrastre, B. (2006) Proteomic analysis of anaplastic lymphoma cell lines: Identification of potential tumour markers. PROTEOMICS **6**, 3210-3222

40 Ying, W., Jiang, Y., Guo, L., Hao, Y., Zhang, Y., Wu, S., Zhong, F., Wang, J., Shi, R., Li, D., Wan, P., Li, X., Wei, H., Li, J., Wang, Z., Xue, X., Cai, Y., Zhu, Y., Qian, X. and He, F. (2006) A Dataset of Human Fetal Liver Proteome Identified by Subcellular Fractionation and Multiple Protein Separation and Identification Technology. Mol Cell Proteomics **5**, 1703-1707

41 Wang, W., Guo, T., Song, T., Lee, C.S., Balgley, B.M. (2007) Comprehensive yeast proteome analysis using a capillary isoelectric focusing-based multidimensional separation platform coupled with ESI-MS/MS. PROTEOMICS **7**, 1178-1187

42 Harwood, M. M., Christians, E. S., Fazal, M. A. and Dovichi, N. J. (2006) Single-cell protein analysis of a single mouse embryo by two-dimensional capillary electrophoresis. Journal of Chromatography A **1130**, 190-194

43 Zhang, M. Q. and El Rassi, Z. (2006) Two-dimensional microcolumn separation platform for proteomics consisting of on-line coupled capillary isoelectric focusing and capillary electrochromatography. 1. Evaluation of the capillary-based two-dimensional platform with proteins, peptides, and human serum. Journal of Proteome Research **5**, 2001-2008

44 Yoo, C., Zhao, J., Pal, M., Hersberger, K., Huber, C. G., Simeone, D. M., Beer, D. G. and Lubman, D. M. (2006) Automated integration of monolith-based protein separation with on-plate digestion for mass spectrometric analysis of esophageal adenocarcinorna human epithelial samples. Electrophoresis **27**, 3643-3651

45 Opiteck, G. J., Ramirez, S. M., Jorgenson, J. W. and Moseley, M. A. (1998) Comprehensive two-dimensional high-performance liquid chromatography for the isolation of overexpressed proteins and proteome mapping. Analytical Biochemistry **258**, 349-361

46 Simpson, D. C., Ahn, S., Pasa-Tolic, L., Bogdanov, B., Mottaz, H. M., Vilkov, A. N., Anderson, G. A., Lipton, M. S. and Smith, R. D. (2006) Using size exclusion chromatography-RPLC and RPLC-CIEF as two-dimensional separation strategies for protein profiling. Electrophoresis **27**, 2722-2733

47 Sharma, S., Simpson, D. C., Tolic, N., Jaitly, N., Mayampurath, A. M., Smith, R. D. and Pasa-Tolic, L. (2007) Proteomic Profiling of Intact Proteins Using WAX-RPLC 2-D Separations and FTICR Mass Spectrometry. J. Proteome Res. **6**, 602-610

48 Pappin, D. J. C., Hojrup, P. and Bleasby, A. J. (1993) Rapid Identification of Proteins by Peptide-Mass Fingerprinting. Current Biology **3**, 327-332

49 Meyer, V. R. (1992) Tswett,Michael Columns - Facts and Speculations. Chromatographia **34**, 342-346

50 Scott, R. P. W. (1992) Modern Liquid-Chromatography. Chemical Society Reviews **21**, 137-145

51 Liu, H. B., Lin, D. Y. and Yates, J. R. (2002) Multidimensional separations for protein/peptide analysis in the post-genomic era. Biotechniques **32**, 898-+

52 Oberacher, H., Premstaller, A. and Huber, C. G. (2004) Characterization of some physical and chromatographic properties of monolithic poly(styrene-co-divinylbenzene) columns. Journal of Chromatography **1030**, 201-208

53 Claessens, H. A. and van Straten, M. A. (2004) Review on the chemical and thermal stability of stationary phases for reversed-phase liquid chromatography. Journal of Chromatography A **1060**, 23-41

54 Boehm, R. E. and Martire, D. E. (1994) Theory of Liquid-Chromatographic Retention and Solute-Transfer Thermodynamics Using the Bethe-Guggenheim Quasi-Chemical Approach. Journal of Physical Chemistry **98**, 1317-1327

55 Dorsey, J. G., Cooper, W. T., Siles, B. A., Foley, J. P. and Barth, H. G. (1996) Liquid chromatography: Theory and methodology. Analytical Chemistry **68**, R515-R568

56 Geng, X. B. and Regnier, F. E. (2003) An integrated theory of adsorption and partition mechanism and each contribution to solute retention in reversed phase liquid chromatography. Chinese Journal of Chemistry **21**, 311-319

57 Miyabe, K. and Guiochon, G. (2002) New model of surface diffusion in reversed-phase liquid chromatography. Journal of Chromatography A **961**, 23-33

58 Thevenonemeric, G., Tchapla, A. and Martin, M. (1991) Role of Pi-Pi Interactions in Reversed-Phase Liquid-Chromatography. Journal of Chromatography **550**, 267-283

59 Vailaya, A. and Horvath, C. (1998) Retention in reversed-phase chromatography: partition or adsorption? Journal of Chromatography A **829**, 1-27

60 Martin, A. J. and Synge, R. L. (1941) A new form of chromatogram employing two liquid phases: A theory of chromatography. 2. Application to the micro-determination of the higher monoamino-acids in proteins. Biochem. J. **35**, 1358-1368

61 van Deemter, J. J., Zuiderweg, F. J. and Klinkenberg, A. (1956) Longitudinal diffusion and resistance to mass transfer as causes of nonideality in chromatography. Chemical Engineering Science **5**, 271-289

62 Dao T.-T. Nguyen, D. G. S. R. J.-L. V. (2008) Validation of an ultra-fast UPLC-UV method for the separation of antituberculosis tablets. Journal of Separation Science **9999**, NA

63 Banks, J. F. (1997) Protein analysis by packed capillary liquid chromatography with electrospray ionization and time-of-flight mass spectrometry detection. Journal of Chromatography A **786**, 67-73

64 Rieux, L., Niederlander, H., Verpoorte, E. and Bischoff, R. (2005) Silica monolithic columns: Synthesis, characterisation and applications to the analysis of biological molecules. Journal of Separation Science **28**, 1628-1641

65 Zhuravlev, N. D., Siepmann, J. I. and Schure, M. R. (2001) Surface coverages of bonded-phase ligands on silica: A computational study. Analytical Chemistry **73**, 4006-4011

66 Kennedy, J. F., Rivera, Z. S. and White, C. A. (1989) The Use of Hplc in Biotechnology. Journal of Biotechnology **9**, 83-106

67 Podgornik, A., Barut, M., Strancar, A., Josic, D. and Koloini, T. (2000) Construction of large volume monolithic columns. Analytical Chemistry **72**, 5693-5699

68 Kaji, H., Nakanishi, K. and Soga, N. (1995) Formation of Porous Gel Morphology by Phase-Separation in Gelling Alkoxy-Derived Silica - Phenomenological Study. Journal of Non-Crystalline Solids **185**, 18-30

69 Tholey, A., Toll, H. and Huber, C. G. (2005) Separation and Detection of Phosphorylated and Nonphosphorylated Peptides in Liquid Chromatography-Mass Spectrometry Using Monolithic Columns and Acidic or Alkaline Mobile Phases. Anal. Chem. **77**, 4618-4625

70 Premstaller, A., Oberacher, H. and Huber, C. G. (2000) High-Performance Liquid Chromatography-Electrospray Ionization Mass Spectrometry of Single- and Double-Stranded Nucleic Acids Using Monolithic Capillary Columns. Anal. Chem. **72**, 4386-4393

71 Legido-Quigley, C., Marlin, N. and Smith, N. W. (2004) Comparison of styrene-divinylbenzene-based monoliths and Vydac nano-liquid chromatography columns for protein analysis. Journal of Chromatography A **1030**, 195-200
72 Ivanov, A. R., Zang, L. and Karger, B. L. (2003) Low-attomole electrospray ionization MS and MS/MS analysis of protein tryptic digests using 20-mu m-i.d. polystyrene-divinylbenzene monolithic capillary columns. Analytical Chemistry **75**, 5306-5316
73 Vailaya, A. (2005) Fundamentals of reversed phase chromatography: Thermodynamic and exothermodynamic treatment. Journal of Liquid Chromatography & Related Technologies **28**, 965-1054
74 Okada, T. (1995) Chromatographic-Separation of Ions - a Review on Separation Mechanisms (Review). Bunseki Kagaku **44**, 579-601
75 Wong, J. W., Albright, R. L. and Wang, N. H. L. (1991) Immobilized Metal-Ion Affinity-Chromatography (Imac) Chemistry and Bioseparation Applications. Separation and Purification Methods **20**, 49-106
76 Larsen, M. R., Thingholm, T. E., Jensen, O. N., Roepstorff, P. and Jorgensen, T. J. D. (2005) Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns. Molecular & Cellular Proteomics **4**, 873-886
77 Samuelsson, G. (1999) What's happening?: Protein A columns: current concepts and recent advances. Transfusion Science **21**, 215-217
78 Kobata, A. and Endo, T. (1992) Immobilized Lectin Columns - Useful Tools for the Fractionation and Structural-Analysis of Oligosaccharides. Journal of Chromatography **597**, 111-122
79 Leickt, L., Grubb, A. and Ohlson, S. (1998) Affinity screening for weak monoclonal antibodies. Journal of Immunological Methods **220**, 19-24
80 Annesley, T. M. (2003) Ion Suppression in Mass Spectrometry. Clin Chem **49**, 1041-1044
81 Paull, B. and Nesterenko, P. N. (2005) Novel ion chromatographic stationary phases for the analysis of complex matrices. Analyst **130**, 134-146
82 Dole, M., Mack, L. L. and Hines, R. L. (1968) Molecular Beams of Macroions. Journal of Chemical Physics **49**, 2240-&
83 Tanaka, K. W., H.; Ido Y.; Akita, S.; Yoshida, Y.; Tamio, Y.; Matsuo, Y.T. (1988) Protein and polymer analyses up to <I>m/z</I> 100 000 by laser ionization time-of-flight mass spectrometry. Rapid Communications in Mass Spectrometry **2**, 151-153
84 Aberth, W. H. and Sperry, R. R. (1974) Use of a Microchannel Plate for Ion-Beam Tuning of Analytical Mass Spectrometers. Review of Scientific Instruments **45**, 128-129
85 Mantus, D. S., Valaskovic, G. A. and Morrison, G. H. (1991) High mass resolution secondary ion mass spectrometry via simultaneous detection with a charge-coupled device. Anal. Chem. **63**, 788-792
86 Fenn, J. B. (2002) Electrospray ionization mass spectrometry: How it all began. J Biomol Tech **13**, 101-118
87 Bogdan Bogdanov, R. D. S. (2005) Proteomics by FTICR mass spectrometry: Top down and bottom up. Mass Spectrometry Reviews **24**, 168-200
88 Weyer, K., Kloppel, K. D. and Vonbunau, G. (1983) Secondary-Ion Mass-Spectrometry (Sims) of Organic-Compounds .2. Mass-Spectra of Peptides. International Journal of Mass Spectrometry and Ion Processes **51**, 235-243
89 Eckart, K. and Schwarz, H. (1987) Sequencing of Tentoxin by Using Fast-Atom-Bombardment (Fab) High-Resolution (Hr) Tandem Mass-Spectrometry (Msms) - Scope and Limitation of a Novel Strategy. Helvetica Chimica Acta **70**, 489-498

90 Barber, M., Bordoli, R. S., Sedgwick, R. D. and Tyler, A. N. (1981) Fast Atom Bombardment of Solids as an Ion-Source in Mass-Spectrometry. Nature **293**, 270-275
91 Chang, W. C., Huang, L. C. L., Wang, Y. S., Peng, W. P., Chang, H. C., Hsu, N. Y., Yang, W. B. and Chen, C. H. (2007) Matrix-assisted laser desorption/ionization (MALDI) mechanism revisited. Analytica Chimica Acta **582**, 1-9
92 Paul, W. and Steinwedel, H. (1953) *Ein Neues Massenspektrometer Ohne Magnetfeld. Zeitschrift Fur Naturforschung Section a-a Journal of Physical Sciences **8**, 448-450
93 Goudsmit, S. A. (1948) A Time-of-Flight Mass Spectrometer. Physical Review **74**, 622-623
94 Comisaro.Mb and Marshall, A. G. (1974) Fourier-Transform Ion-Cyclotron Resonance Spectroscopy. Chemical Physics Letters **25**, 282-283
95 Hu, Q. Z., Noll, R. J., Li, H. Y., Makarov, A., Hardman, M. and Cooks, R. G. (2005) The Orbitrap: a new mass spectrometer. Journal of Mass Spectrometry **40**, 430-443
96 Steel, S. H., M (1998) Understanding the Quadrupole Mass Filter through Computer Simulation. J. Chem. Educ. **75**, 1049
97 Church, D. A. (1969) Storage-Ring Ion Trap Derived from the Linear Quadrupole Radio-Frequency Mass Filter. Journal of Applied Physics **40**, 3127-3134
98 Mamyrin, B. A., Karataev, V. I., Shmikk, D. V. and Zagulin, V. A. (1973) Mass-Reflectron a New Nonmagnetic Time-of-Flight High-Resolution Mass-Spectrometer. Zhurnal Eksperimentalnoi I Teoreticheskoi Fiziki **64**, 82-89
99 Larsen, B. S. M., C. N. (1998) Mass Spectrometry of Biological Materials. Marcel Dekker, New York
100 Marshall, A. G., Hendrickson, C. L. and Jackson, G. S. (1998) Fourier transform ion cyclotron resonance mass spectrometry: A primer. Mass Spectrometry Reviews **17**, 1-35
101 Kingdon, K. H. (1923) A Method for the Neutralization of Electron Space Charge by Positive Ionization at Very Low Gas Pressures. Physical Review **21**, 408
102 Chien, B. M., Michael, S. M. and Lubman, D. M. (1994) The design and performance of an ion trap storage--reflectron time-of-flight mass spectrometer. International Journal of Mass Spectrometry and Ion Processes **131**, 149-179
103 Michael, S. M., Chien, B. M. and Lubman, D. M. (1993) Detection of Electrospray-Ionization Using a Quadrupole Ion-Trap Storage Reflectron Time-of-Flight Mass-Spectrometer. Analytical Chemistry **65**, 2614-2620
104 Bogdanov, B. and Smith, R. D. (2005) Proteomics by FTICR mass spectrometry: Top down and bottom up. Mass Spectrometry Reviews **24**, 168-200
105 Rudnick, P. A., Wang, Y. J., Evans, E., Lee, C. S. and Balgley, B. M. (2005) Large scale analysis of MASCOT results using a mass accuracy-based THreshold (MATH) effectively improves data interpretation. Journal of Proteome Research **4**, 1353-1360
106 Shadforth, I. T., K.; Crowther, D.; Bessant, C (2005) Determination of partial amino acid composition from tandem mass spectra for use in peptide identification strategies. PROTEOMICS **5**, 1787-1796
107 Johnson, R. S., Martin, S. A., Biemann, K., Stults, J. T. and Watson, J. T. (1987) Novel Fragmentation Process of Peptides by Collision-Induced Decomposition in a Tandem Mass-Spectrometer - Differentiation of Leucine and Isoleucine. Analytical Chemistry **59**, 2621-2625
108 Zubarev, R. A., Horn, D. M., Fridriksson, E. K., Kelleher, N. L., Kruger, N. A., Lewis, M. A., Carpenter, B. K. and McLafferty, F. W. (2000) Electron capture dissociation for structural characterization of multiply charged protein cations. Analytical Chemistry **72**, 563-573

109 Zubarev, R. A., Kelleher, N. L. and McLafferty, F. W. (1998) Electron capture dissociation of multiply charged protein cations. A nonergodic process. Journal of the American Chemical Society **120**, 3265-3266
110 Syka, J. E. P., Coon, J. J., Schroeder, M. J., Shabanowitz, J. and Hunt, D. F. (2004) Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. Proc. Natl. Acad. Sci. U. S. A. **101**, 9528-9533
111 Little, D. P., Speir, J. P., Senko, M. W., Oconnor, P. B. and McLafferty, F. W. (1994) Infrared Multiphoton Dissociation of Large Multiply-Charged Ions for Biomolecule Sequencing. Analytical Chemistry **66**, 2809-2815
112 Spengler, B., Luetzenkirchen, F., Metzger, S., Chaurand, P., Kaufmann, R., Jeffery, W., Bartlet-Jones, M. and Pappin, D. J. C. (1997) Peptide sequencing of charged derivatives by postsource decay MALDI mass spectrometry. International Journal of Mass Spectrometry **169**, 127-140
113 Lennon, J. J. and Walsh, K. A. (1999) Locating and identifying posttranslational modifications by in-source decay during MALDI-TOF mass spectrometry. Protein Science **8**, 2487-2493
114 Roepstorff, P. and Fohlman, J. (1984) Proposal for a Common Nomenclature for Sequence Ions in Mass-Spectra of Peptides. Biomedical Mass Spectrometry **11**, 601-601
115 Ambihapathy, K., Yalcin, T., Leung, H. W. and Harrison, A. G. (1997) Pathways to immonium ions in the fragmentation of protonated peptides. Journal of Mass Spectrometry **32**, 209-215
116 Khalsa-Moyers, G. and McDonald, W. H. (2006) Developments in mass spectrometry for the analysis of complex protein mixtures. Brief Funct Genomic Proteomic **5**, 98-111
117 Syrstad, E. A., Turec and ek, F. (2005) Toward a general mechanism of electron capture dissociation. J. Am. Soc. Mass Spectrom. **16**, 208-224
118 Anusiewicz, W., Berdys-Kochanska, J. and Simons, J. (2005) Electron attachment step in electron capture dissociation (ECD) and electron transfer dissociation (ETD). Journal of Physical Chemistry A **109**, 5801-5813
119 Coon, J. J., Ueberheide, B., Syka, J. E. P., Dryhurst, D. D., Ausio, J., Shabanowitz, J. and Hunt, D. F. (2005) Protein identification using sequential ion/ion reactions and tandem mass spectrometry. Proc. Natl. Acad. Sci. U. S. A. **102**, 9463-9468
120 Hartmer, R. K., D.A.; Gebhardt, C.R.; Ledertheil, T.; Brekenfeld, A. (2008) Multiple Ion/Ion Reactions in the 3-D Ion Trap:
Selective Reagent Anion Production for ETD and PTR from a Single Compound. International Journal of Mass Spectrometry
121 Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. and Walter, P. (2002) Molecular Biology of the Cell, Fourth Edition. Garland
122 Shaw, M. M. and Riederer, B. M. (2003) Sample preparation for two-dimensional gel electrophoresis. Proteomics **3**, 1408-1417
123 Neugebauer, J. M. (1990) Detergents - an Overview. Methods in Enzymology **182**, 239-253
124 Jorgensen, W. L., Duffy, E. M. and Tiradorives, J. (1993) Computational Investigations of Protein Denaturation - Apomyoglobin and Chaotrope-Arene Interactions. Philosophical Transactions of the Royal Society of London Series a-Mathematical Physical and Engineering Sciences **345**, 87-96
125 Facincani, A. P., Ferro, J. A., Pizauro, J. M., Pereira, H. A., Lemos, E. G. D., do Prado, A. L. and Ferro, M. I. T. (2003) Carbohydrate metabolism of Xylella fastidiosa: Detection of glycolytic and pentose phosphate pathway enzymes and cloning and expression of the enolase gene. Genetics and Molecular Biology **26**, 203-211
126 Gould, T. A., Watson, W. T., Choi, K. H., Schweizer, H. P. and Churchill, M. E. A. (2004) Crystallization of Pseudomonas aeruginosa AHL synthase Lasl using beta-

turn crystal engineering. Acta Crystallographica Section D-Biological Crystallography **60**, 518-520

127 Muise, A. M. and Ro, H. S. (1999) Enzymic characterization of a novel member of the regulatory B-like carboxypeptidase with transcriptional repression function: stimulation of enzymic activity by its target DNA. Biochem. J. **343**, 341-345

128 Van Burik, J. A. H., Schreckhise, R. W., White, T. C., Bowden, R. A. and Myerson, D. (1998) Comparison of six extraction techniques for isolation of DNA from filamentous fungi. Medical Mycology **36**, 299-303

129 Borthwick, K. A. J., Coakley, W. T., McDonnell, M. B., Nowotny, H., Benes, E. and Groschl, M. (2005) Development of a novel compact sonicator for cell disruption. Journal of Microbiological Methods **60**, 207-216

130 Jiang, L., He, L. and Fountoulakis, M. (2004) Comparison of protein precipitation methods for sample preparation prior to proteomic analysis. Journal of Chromatography A **1023**, 317-320

131 Arakawa, T., Kita, Y. and Timasheff, S. N. (2007) Protein precipitation and denaturation by dimethyl sulfoxide. Biophysical Chemistry **131**, 62-70

132 Shih, Y. C., Prausnitz, J. M. and Blanch, H. W. (1992) Some Characteristics of Protein Precipitation by Salts. Biotechnology and Bioengineering **40**, 1155-1164

133 Sivaraman, T., Kumar, T. K. S., Jayaraman, G. and Yu, C. (1997) The mechanism of 2,2,2-trichloroacetic acid-induced protein precipitation. Journal of Protein Chemistry **16**, 291-297

134 McCarthy, J., Hopwood, F., Oxley, D., Laver, M., Castagna, A., Righetti, P. G., Williams, K. and Herbert, B. (2003) Carbamylation of proteins in 2-D electrophoresis - Myth or reality? Journal of Proteome Research **2**, 239-242

135 Kanamori, T. and Hayakawa, T. (1987) Ion-Exchange Chromatography of Ionic Detergent-Solubilized Proteins - Application to Purification of Rat Parotid-Gland Phosphoproteins Including Ribosomal Protein-S6. Analytical Biochemistry **167**, 372-380

136 Midura, R. J. and Yanagishita, M. (1995) Chaotropic Solvents Increase the Critical Micellar Concentrations of Detergents. Analytical Biochemistry **228**, 318-322

137 Samyn, B., Sergeant, K., Castanheira, P., Faro, C. and Van Beeumen, J. (2005) A new method for C-terminal sequence analysis in the proteomic era. Nature Methods **2**, 193-200

138 Rahali, V. and Gueguen, J. (1999) Chemical cleavage of bovine beta-lactoglobulin by BNPS-skatole for preparative purposes: Comparative study of hydrolytic procedures and peptide characterization. Journal of Protein Chemistry **18**, 1-12

139 Horrobin, D. F. (2003) Modern biomedical research: an internally self-consistent universe with little contact with medical reality? Nature Reviews Drug Discovery **2**, 151-154

140 EyreWalker, A. (1996) Synonymous codon bias is related to gene length in Escherichia coli: Selection for translational accuracy? Molecular Biology and Evolution **13**, 864-872

141 Gygi, S. P., Rist, B., Griffin, T. J., Eng, J. and Aebersold, R. (2002) Proteome analysis of low-abundance proteins using multidimensional chromatography and isotope-coded affinity tags. Journal of Proteome Research **1**, 47-54

142 Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H. and Aebersold, R. (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nature Biotechnology **17**, 994-999

143 Miseta, A. and Csutora, P. (2000) Relationship between the occurrence of cysteine in proteins and the complexity of organisms. Molecular Biology and Evolution **17**, 1232-1239

144 Yao, X. D., Afonso, C. and Fenselau, C. (2003) Dissection of proteolytic O-18 labeling: Endoprotease-catalyzed O-16-to-O-18 exchange of truncated peptide substrates. Journal of Proteome Research **2**, 147-152

145 Johnson, K. L. and Muddiman, D. C. (2004) A method for calculating 16o/18o peptide ion ratios for the relative quantification of proteomes. J. Am. Soc. Mass Spectrom. **15**, 437-445
146 Aggarwal, K., Choe, L. H. and Lee, K. H. (2005) Quantitative analysis of protein expression using amine-specific isobaric tags in Escherichia coli cells expressing rhsA elements. Proteomics **5**, 2297-2308
147 Ross, P. L., Huang, Y. L. 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. and Pappin, D. J. (2004) Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. Molecular & Cellular Proteomics **3**, 1154-1169
148 Ong, S.-E. and Mann, M. (2007) A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC). Nat. Protocols **1**, 2650-2660
149 Palagi, P. M., Walther, D., Quadroni, M., Catherinet, S., Burgess, J., Zimmermann-Ivol, C. G., Sanchez, J. C., Binz, P. A., Hochstrasser, D. F. and Appel, R. D. (2005) MSight: An image analysis software for liquid chromatography-mass spectrometry. Proteomics **5**, 2381-2384
150 Mattoo, S. and Cherry, J. D. (2005) Molecular pathogenesis, epidemiology, and clinical manifestations of respiratory infections due to Bordetella pertussis and other Bordetella subspecies. Clinical Microbiology Reviews **18**, 326-+
151 Preston, A. (2005) Bordetella pertussis: the intersection of genomics and pathobiology. Canadian Medical Association Journal **173**, 55-62
152 Natera, S., Machuca, C., Padron-Nieves, M., Romero, A., Diaz, E. and Ponte-Sucre, A. (2007) Leishmania spp.: proficiency of drug-resistant parasites. International Journal of Antimicrobial Agents **29**, 637-642
153 Bates, P. A. (2007) Transmission of Leishmania metacyclic promastigotes by phlebotomine sand flies. International Journal for Parasitology **37**, 1097-1106
154 Amato, V. S., Tuon, F. F., Bacha, H. A., Neto, V. A. and Nicodemo, A. C. (2008) Mucosal leishmaniasis - Current scenario and prospects for treatment. Acta Tropica **105**, 1-9
155 Croft, S. L., Sundar, S. and Fairlamb, A. H. (2006) Drug resistance in leishmaniasis. Clinical Microbiology Reviews **19**, 111-+
156 Alitalo, R. (1990) Induced-Differentiation of K562 Leukemia-Cells - a Model for Studies of Gene-Expression in Early Megakaryoblasts. Leukemia Research **14**, 501-&
157 Nowell, P. C. and Hungerford, D. A. (1960) Minute Chromosome in Human Chronic Granulocytic Leukemia. Science **132**, 1497-1497
158 Laurent, E., Talpaz, M., Kantarjian, H. and Kurzrock, R. (2001) The BCR Gene and Philadelphia Chromosome-positive Leukemogenesis. Cancer Res **61**, 2343-2355
159 Finn, A. J., Feng, G. and Pendergast, A. M. (2003) Postsynaptic requirement for Abl kinases in assembly of the neuromuscular junction. Nat Neurosci **6**, 717-723
160 Maru, Y. and Witte, O. N. (1991) The BCR gene encodes a novel serine/threonine kinase activity within a single exon. Cell **67**, 459-468
161 Liu, X. S. and Jiang, J. (2006) Enhancement of imatinib-induced apoptosis by matrine in bcr/abl-positive leukemia K562 cells. Pharmaceutical Biology **44**, 287-291
162 Adler, E. M., Gough, N. R. and Blundon, J. A. (2006) Differentiation of PC12 Cells. Sci. STKE **2006**, tr9-
163 Brightman, F. A. and Fell, D. A. (2000) Differential feedback regulation of the MAPK cascade underlies the quantitative differences in EGF and NGF signalling in PC12 cells. Febs Letters **482**, 169-174
164 Apffel, J. A., Alfredson, T. V. and Majors, R. E. (1981) Automated Online Multidimensional High-Performance Liquid-Chromatographic Techniques for the

Cleanup and Analysis of Water-Soluble Samples. Journal of Chromatography **206**, 43-57

165 Opiteck, G. J., Jorgenson, J. W. and Anderegg, R. J. (1997) Two-dimensional SEC/RPLC coupled to mass spectrometry for the analysis of peptides. Analytical Chemistry **69**, 2283-2291

166 Cass, C. L., Johnson, J. R., Califf, L. L., Xu, T., Hernandez, H. J., Stadecker, M. J., Yates, J. R. and Williams, D. L. (2007) Proteomic analysis of Schistosoma mansoni egg secretions. Molecular and Biochemical Parasitology **155**, 84-93

167 Alger, H. M., Sayed, A. A., Stadecker, M. J. and Williams, D. L. (2002) Molecular and enzymatic characterisation of Schistosoma mansoni thioredoxin. International Journal for Parasitology **32**, 1285-1292

168 Luu, V. D., Brems, S., Hoheisel, J. D., Burchmore, R., Guilbride, D. L. and Clayton, C. (2006) Functional analysis of Trypanosoma brucei PUF1. Molecular and Biochemical Parasitology **150**, 340-349

169 McLean, L., Young, I. S., Doherty, M. K., Robertson, D. H. L., Cossins, A. R., Gracey, A. Y., Beynon, R. J. and Whitfield, P. D. (2007) Global cooling: Cold acclimation and the expression of soluble proteins in carp skeletal muscle. Proteomics **7**, 2667-2681

170 Chahal, F. C., Entwistle, J., Glover, N. and MacDonald, G. C. (2006) A targeted proteomic approach for the identification of tumor-associated membrane antigens using the ProteomeLab (TM) PF-2D in tandem with mass spectrometry. Biochemical and Biophysical Research Communications **348**, 1055-1062

171 Ying, W. T., Jiang, Y., Guo, L. H., Hao, Y. W., Zhang, Y. J., Wu, S. F., Zhong, F., Wang, J. L., Shi, R., Li, D., Wan, P., Li, X. H., Wei, H. D., Li, J. Q., Wang, Z. S., Xue, X. F., Cai, Y., Zhu, Y. P., Qian, X. H. and He, F. C. (2006) A dataset of human fetal liver proteome identified by subcellular fractionation and multiple protein separation and identification technology. Molecular & Cellular Proteomics **5**, 1703-1707

172 Zhao, J., Zhu, K., Lubman, D. M., Miller, F. R., Shekhar, M. P. V., Gerard, B. and Barder, T. J. (2006) Proteomic analysis of estrogen response of premalignant human breast cells using a 2-D liquid separation/mass mapping technique. Proteomics **6**, 3847-3861

173 Canas, B., Pineiro, C., Calvo, E., Lopez-Ferrer, D. and Gallardo, J. M. (2007) Trends in sample preparation for classical and second generation proteomics. Journal of Chromatography A **1153**, 235-258

174 Chattopadhyay, A. and Harikumar, K. G. (1996) Dependence of critical micelle concentration of a zwitterionic detergent on ionic strength: Implications in receptor solubilization. Febs Letters **391**, 199-202

175 Watarai, H., Inagaki, Y., Kubota, N., Fuju, K., Nagafune, J., Yamaguchi, Y. and Kadoya, T. (2000) Proteomic approach to the identification of cell membrane proteins. Electrophoresis **21**, 460-464

176 Kreunin, P., Urquidi, V., Lubman, D. M. and Goodison, S. (2004) Identification of metastasis-associated proteins in a human tumor metastasis model using the mass-mapping technique. Proteomics **4**, 2754-2765

177 Meng, F. Y., Cargile, B. J., Patrie, S. M., Johnson, J. R., McLoughlin, S. M. and Kelleher, N. L. (2002) Processing complex mixtures of intact proteins for direct analysis by mass spectrometry. Analytical Chemistry **74**, 2923-2929

178 Huntington, J. A. and Stein, P. E. (2001) Structure and properties of ovalbumin. Journal of Chromatography B: Biomedical Sciences and Applications **756**, 189-198

179 Vidakovics, M. L., Paba, J., Lamberti, Y., Ricart, C. A., de Sousa, M. V. and Rodriguez, M. E. (2007) Profiling the Bordetella pertussis proteome during iron starvation. Journal of Proteome Research **6**, 2518-2528

180 Warringer, J. and Blomberg, A. (2006) Evolutionary constraints on yeast protein size. Bmc Evolutionary Biology **6**

181 Adessi, C., Miege, C., Albrieux, C., Rabilloud, T. (1997) Two-dimensional electrophoresis of membrane proteins: A current challenge for immobilized pH gradients. ELECTROPHORESIS **18**, 127-135
182 Pasquali, C., Fialka, I., Huber, L.A. (1997) Preparative two-dimensional gel electrophoresis of membrane proteins. ELECTROPHORESIS **18**, 2573-2581
183 Hofmann, K., Stoffel, W. (1993) TMbase - A database of membrane spanning protein segments. Biol. Chem. Hoppe-Seyler **374**, 166
184 Leifso, K., Cohen-Freue, G., Dogra, N., Murray, A. and McMaster, W. R. (2007) Genomic and proteomic expression analysis of Leishmania promastigote and amastigote life stages: The Leishmania genome is constitutively expressed. Molecular and Biochemical Parasitology **152**, 35-46
185 Wu, G., Culley, D. E. and Zhang, W. W. (2005) Predicted highly expressed genes in the genomes of Streptomyces coelicolor and Streptomyces avermitilis and the implications for their metabolism. Microbiology-Sgm **151**, 2175-2187
186 Bridges, D. J., Pitt, A. R., Hanrahan, O., Brennan, K., Voorheis, H. P., Herzyk, P., de Koning, H. P. and Burchmore, R. J. S. (2008) Characterisation of the plasma membrane subproteome of bloodstream form Trypanosoma brucei. Proteomics **8**, 83-99
187 Parks, B. A., Jiang, L., Thomas, P. M., Wenger, C. D., Roth, M. J., Boyne, M. T., Burke, P. V., Kwast, K. E. and Kelleher, N. L. (2007) Top-Down Proteomics on a Chromatographic Time Scale Using Linear Ion Trap Fourier Transform Hybrid Mass Spectrometers. Anal. Chem. **79**, 7984-7991
188 Simpson, D., Ahn, S., Pasa-Tolic, L., Bogdanov, B., Mottaz, H.M., Vilkov, A.N., Anderson, G.A., Lipton, M.S., Smith, R.D. (2006) Using size exclusion chromatography-RPLC and RPLC-CIEF as two-dimensional separation strategies for protein profiling. ELECTROPHORESIS **27**, 2722-2733
189 McDonald, T., Sheng, S., Stanley, B., Chen, D., Ko, Y., Cole, R. N., Pedersen, P. and Van Eyk, J. E. (2006) Expanding the Subproteome of the Inner Mitochondria Using Protein Separation Technologies: One- and Two-dimensional Liquid Chromatography and Two-dimensional Gel Electrophoresis. Mol Cell Proteomics **5**, 2392-2411
190 Sheng, S., Chen, D. and Van Eyk, J. E. (2006) Multidimensional Liquid Chromatography Separation of Intact Proteins by Chromatographic Focusing and Reversed Phase of the Human Serum Proteome: Optimization and Protein Database. Mol Cell Proteomics **5**, 26-34
191 Giddings, J. (1981) Advances in Chromatography. CRC Press, London
192 W. Hearn, M. T., Grego, B. and Hancock, W. S. (1979) High-performance liquid chromatography of amino acids, peptide and proteins : XX. Investigation of the effect of pH and ion-pair formation on the retention of peptides on chemically-bonded hydrocarbonaceous stationary phases. Journal of Chromatography **185**, 429-444
193 Young, P. M. and Wheat, T. E. (1990) Optimization of high-performance liquid chromatographic peptide separations with alternative mobile and stationary phases. Journal of Chromatography A **512**, 273-281
194 Toll, H., Oberacher, H., Swart, R. and Huber, C. G. (2005) Separation, detection, and identification of peptides by ion-pair reversed-phase high-performance liquid chromatography-electrospray ionization mass spectrometry at high and low pH. Journal of Chromatography A **1079**, 274-286
195 Lubman, D. M., Kachman, M. T., Wang, H. X., Gong, S. Y., Yan, F., Hamler, R. L., O'Neil, K. A., Zhu, K., Buchanan, N. S. and Barder, T. J. (2002) Two-dimensional liquid separations-mass mapping of proteins from human cancer cell lysates. Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences **782**, 183-196

196 Sluyterman, L. A. A. and Elgersma, O. (1978) Chromatofocusing - Isoelectric-Focusing on Ion-Exchange Columns .1. General Principles. Journal of Chromatography **150**, 17-30
197 Bates, R. C., Kang, X. Z. and Frey, D. D. (2000) High-performance chromatofocusing using linear and concave pH gradients formed with simple buffer mixtures I. Effect of buffer composition on the gradient shape. Journal of Chromatography A **890**, 25-36
198 Sluyterman, L. A. A. and Wijdenes, J. (1978) Chromatofocusing - Isoelectric-Focusing on Ion-Exchange Columns .2. Experimental-Verification. Journal of Chromatography **150**, 31-44
199 Hearn, M. T. W. and Lyttle, D. J. (1981) Buffer-Focusing Chromatography Using Multicomponent Electrolyte Elution Systems. Journal of Chromatography **218**, 483-495
200 Wagner, G. and Regnier, F. E. (1982) Rapid Chromatofocusing of Proteins. Analytical Biochemistry **126**, 37-43
201 Tsonev, L. (2005) CryoBioPhysica, Inc.
202 Liu, Y. S. and Anderson, D. J. (1997) Gradient chromatofocusing high-performance liquid chromatography .2. Theoretical aspects. Journal of Chromatography A **762**, 47-54
203 Liu, Y. S. and Anderson, D. J. (1997) Gradient chromatofocusing high-performance liquid chromatography .1. Practical aspects. Journal of Chromatography A **762**, 207-217
204 Pepaj, M., Lundanes, E. and Greibrokk, T. (2007) Separation of apolipoprotein A-I from human plasma by on-line two dimensional liquid chromatography. Journal of Liquid Chromatography & Related Technologies **30**, 1879-1894
205 McCalley, D. V. (2004) Effect of buffer on peak shape of peptides in reversed-phase high performance liquid chromatography. Journal of Chromatography A **1038**, 77-84
206 Knighton, D. R., Harding, D. R. K., Napier, J. R. and Hancock, W. S. (1985) Purification of Synthetic Lipid Associating Peptides and the Monitoring of the Deformylation of Nin-Formyltryptophan by Reversed-Phase High-Performance Liquid-Chromatography. Journal of Chromatography **347**, 237-248
207 Garcia, M. C. (2005) The effect of the mobile phase additives on sensitivity in the analysis of peptides and proteins by high-performance liquid chromatography-electrospray mass spectrometry. Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences **825**, 111-123
208 Mizobe, M., Kondo, F., Kumamoto, K., Kanda, Y. and Seguchi, H. (1997) High-performance liquid chromatographic analysis of bilirubin and biliverdin from jaundiced broilers. Journal of Veterinary Medical Science **59**, 677-680
209 de Andrade, A. S. R., Santoro, M. M., de Melo, M. N. and Mares-Guia, M. (1998) Leishmania (Leishmania) amazonensis: Purification and enzymatic characterization of a soluble serine oligopeptidase from promastigotes. Experimental Parasitology **89**, 153-160
210 Yoo, C., Patwa, T. H., Kreunin, P., Miller, F. R., Huber, C. G., Nesvizhskii, A. I. and Lubman, D. M. (2007) Comprehensive analysis of proteins of pH fractionated samples using monolithic LC/MS/MS, intact MW measurement and MALDI-QIT-TOF MS. Journal of Mass Spectrometry **42**, 312-334
211 Haynes, P. A., Fripp, N. and Aebersold, R. (1998) Identification of gel-separated proteins by liquid chromatography electrospray tandem mass spectrometry: Comparison of methods and their limitations. Electrophoresis **19**, 939-945
212 Patterson, S. D., Thomas, D. and Bradshaw, R. A. (1996) Application of combined mass spectrometry and partial amino acid sequence to the identification of gel-separated proteins. Electrophoresis **17**, 877-891

213 Stone, K. L., Deangelis, R., LoPresti, M., Jones, J., Papov, V.V., Williams, K.R. (1998) Use of liquid chromatography-electrospray ionization-tandem mass spectrometry (LC-ESI-MS/MS) for routine identification of enzymatically digested proteins separated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis. Electrophoresis **19**, 1046-1052

214 Issaq, H. J., Fox, Stephen D., Mahadevan, Megha, Conrads, Thomas P., Veenstra, Timothy D. (2003) Effect of Experimental Parameters on the HPLC Separation of Peptides and Proteins. Journal of Liquid Chromatography & Related Technologies **26**, 2255 - 2283

215 Batycka, M., Inglis, N. F., Cook, K., Adam, A., Fraser-Pitt, D., Smith, D. G. E., Main, L., Lubben, A. and Kessler, B. M. (2006) Ultra-fast tandem mass spectrometry scanning combined with monolithic column liquid chromatography increases throughput in proteomic analysis. Rapid Communications in Mass Spectrometry **20**, 2074-2080

216 Roe, A. J., Tysall, L., Dransfield, T., Wang, D., Fraser-Pitt, D., Mahajan, A., Constandinou, C., Inglis, N., Downing, A., Talbot, R., Smith, D. G. E. and Gally, D. L. (2007) Analysis of the expression, regulation and export of NleA-E in Escherichia coli O157 : H7. Microbiology-Sgm **153**, 1350-1360

217 Premstaller, A., Oberacher, H., Walcher, W., Timperio, A. M., Zolla, L., Chervet, J. P., Cavusoglu, N., van Dorsselaer, A. and Huber, C. G. (2001) High-Performance Liquid Chromatography-Electrospray Ionization Mass Spectrometry Using Monolithic Capillary Columns for Proteomic Studies. Anal. Chem. **73**, 2390-2396

218 Gauthier, A., Puente, J. L. and Finlay, B. B. (2003) Secretin of the enteropathogenic Escherichia coli type III secretion system requires components of the type III apparatus for assembly and localization. Infection and Immunity **71**, 3310-3319

219 Smith, P. K., Krohn, R. I., Hermanson, G. T., Mallia, A. K., Gartner, F. H., Provenzano, M. D., Fujimoto, E. K., Goeke, N. M., Olson, B. J. and Klenk, D. C. (1985) Measurement of protein using bicinchoninic acid. Analytical Biochemistry **150**, 76-85

220 Huber, C. G., Kleindienst, G. and Bonn, G. K. (1997) Application of micropellicular poly-styrene/divinylbenzene stationary phases for high-performance reversed-phase liquid chromotography electrospray mass spectrometry of proteins and peptides. Chromatographia **44**, 438-448

221 Walcher, W., Oberacher, H., Troiani, S., Holzl, G., Oefner, P., Zolla, L. and Huber, C. G. (2002) Monolithic capillary columns for liquid chromatography-electrospray ionization mass spectrometry in proteomic and genomic research. Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences **782**, 111-125

222 Walcher, W., Toll, H., Ingendoh, A. and Huber, C. G. (2004) Operational variables in high-performance liquid chromatography-electrospray ionization mass spectrometry of peptides and proteins using poly(styrene-divinylbenzene) monoliths. Journal of Chromatography A **1053**, 107-117

223 Huber, C. G. and Premstaller, A. (1999) Evaluation of volatile eluents and electrolytes for high-performance liquid chromatography-electrospray ionization mass spectrometry and capillary electrophoresis-electrospray ionization mass spectrometry of proteins - I. Liquid chromatography. Journal of Chromatography A **849**, 161-173

224 Liu, H. J., Berger, S. J., Chakraborty, A. B., Plumb, R. S. and Cohen, S. A. (2002) Multidimensional chromatography coupled to electrospray ionization time-of-flight mass spectrometry as an alternative to two-dimensional gels for the identification and analysis of complex mixtures of intact proteins. Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences **782**, 267-289

225 Toll, H., Wintringer, R., Schweiger-Hufnagel, U. and Huber, C. G. (2005) Comparing monolithic and microparticular capillary columns for the separation and analysis of peptide mixtures by liquid chromatography-mass spectrometry. Journal of Separation Science **28**, 1666-1674
226 Premstaller, A., Oefner, P. J., Oberacher, H. and Huber, C. G. (2002) Capillary array high-performance liquid chromatography of nucleic acids and proteins. Analytical Chemistry **74**, 4688-4693
227 Tan, A. T. and Woodwort.Rc (1968) Absorbance Changes in 275 Mmu Region Associated with Exposure of Buried Tryptophanyl Residues. Biochemical and Biophysical Research Communications **32**, 739-&
228 Shaw, J., Rowlinson, R., Nickson, J., Stone, T., Sweet, A., Williams, K. and Tonge, R. (2003) Evaluation of saturation labelling two-dimensional difference gel electrophoresis fluorescent dyes. Proteomics **3**, 1181-1195
229 Hsu, J. L., Huang, S. Y., Chow, N. H. and Chen, S. H. (2003) Stable-isotope dimethyl labeling for quantitative proteomics. Analytical Chemistry **75**, 6843-6852
230 de Oliveira, A. H. C., Ruiz, J. C., Cruz, A. K., Greene, L. J., Rosa, J. C. and Ward, R. J. (2006) Expression in E-coli and purification of the nucleoside diphosphate kinase b from Leishmania major. Protein Expression and Purification **49**, 244-250
231 Basselin, M., Denise, H., Coombs, G. H. and Barrett, M. P. (2002) Resistance to pentamidine in Leishmania mexicana involves exclusion of the drug from the mitochondrion. Antimicrobial Agents and Chemotherapy **46**, 3731-3738
232 Searle, S. and Smith, D. F. (1993) Leishmania-Major - Characterization and Expression of a Cytoplasmic Stress-Related Protein. Experimental Parasitology **77**, 43-52
233 Schmidt, A., Kellermann, J. and Lottspeich, F. (2005) A novel strategy for quantitative proteornics using isotope-coded protein labels. Proteomics **5**, 4-15
234 Wiese, S. (2007) Protein labeling by iTRAQ: A new tool for quantitative mass spectrometry in proteome research (vol 7, pg 340, 2007). Proteomics **7**, 1004-1004
235 Unwin, R. D., Smith, D. L., Blinco, D., Wilson, C. L., Miller, C. J., Evans, C. A., Jaworska, E., Baldwin, S. A., Barnes, K., Pierce, A., Spooncer, E. and Whetton, A. D. (2006) Quantitative proteomics reveals posttranslational control as a regulatory factor in primary hematopoietic stem cells. Blood **107**, 4687-4694
236 Jardim, A., Hanson, S., Ullman, B., McCubbin, W. D., Kay, C. M. and Olafson, R. W. (1995) Cloning and Structure-Function Analysis of the Leishmania-Donovani Kinetoplastid, Membrane Protein-11. Biochemical Journal **305**, 315-320
237 Rosenzweig, D., Smith, D., Opperdoes, F., Stern, S., Olafson, R. W. and Zilberstein, D. (2008) Retooling Leishmania metabolism: from sand fly gut to human macrophage. Faseb Journal **22**, 590-602
238 Kaimul, A. M., Nakamura, H., Masutani, H. and Yodoi, J. (2007) Thioredoxin and thioredoxin-binding protein-2 in cancer and metabolic syndrome. Free Radical Biology and Medicine **43**, 861-868
239 Li, L. M., Okura, M. and Imamoto, A. (2002) Focal adhesions require catalytic activity of Src family kinases to mediate integrin-matrix adhesion. Molecular and Cellular Biology **22**, 1203-1217
240 Cammas, A., Pileur, F., Bonnal, S., Lewis, S. M., Leveque, N., Holcik, M. and Vagner, S. (2007) Cytoplasmic relocalization of heterogeneous nuclear ribonucleoprotein A1 controls translation initiation of specific mRNAs. Molecular Biology of the Cell **18**, 5048-5059
241 Lewis, S. M., Veyrier, A., Ungureanu, N. H., Bonnal, S., Vagner, S. and Holcik, M. (2007) Subcellular relocalization of a trans-acting factor regulates XIAP IRES-dependent translation. Molecular Biology of the Cell **18**, 1302-1311
242 Nacken, W. and Kerkhoff, C. (2007) The hetero-oligomeric complex of the S100A8/S100A9 protein is extremely protease resistant. Febs Letters **581**, 5127-5130

243 Ikemoto, M., Murayama, H., Itoh, H., Totani, M. and Fujita, M. (2007) Intrinsic function of S100A8/A9 complex as an anti-inflammatory protein in liver injury induced by lipopolysaccharide in rats. Clinica Chimica Acta **376**, 197-204
244 Laird, A. D. and Shalloway, D. (1997) Oncoprotein signalling and mitosis. Cellular Signalling **9**, 249-255
245 Calandra, T. and Roger, T. (2003) Macrophage migration inhibitory factor: A regulator of innate immunity. Nature Reviews Immunology **3**, 791-800
246 Gunning, P., O'Neill, G. and Hardeman, E. (2008) Tropomyosin-based regulation of the actin cytoskeleton in time and space. Physiological Reviews **88**, 1-35
247 Simon, J. P. and Aunis, D. (1989) Biochemistry of the Chromogranin-a Protein Family. Biochemical Journal **262**, 1-13
248 Haglund, K. and Dikic, I. (2005) Ubiquitylation and cell signaling. Embo Journal **24**, 3353-3359
249 Hsu, J. L., Huang, S. Y. and Chen, S. H. (2006) Dimethyl multiplexed labeling combined with microcolumn separation and MS analysis for time course study in proteomics. Electrophoresis **27**, 3652-3660
250 Tribl, F. L., C; Dombert, T.; Langenfeld, E.; Piechura, H.; Warscheid, B.; Meyer, H.E.; Marcus, K (2008) Towards multidimensional liquid chromatography separation of proteins using fluorescence and isotope-coded protein labelling for quantitative proteomics. PROTEOMICS **8**, 1204-1211
251 Albeck, J. G., MacBeath, G., White, F. M., Sorger, P. K., Lauffenburger, D. A. and Gaudet, S. (2006) Collecting and organizing systematic sets of protein data. Nat Rev Mol Cell Biol **7**, 803-812
252 Sanders, W. S., Bridges, S. M., McCarthy, F. M., Nanduri, B. and Burgess, S. C. (2007) Prediction of peptides observable by mass spectrometry applied at the experimental set level. BMC Bioinformatics **8 Suppl 7**, S23
253 Sobott, F., McCammon, M. G., Hernandez, H. and Robinson, C. V. (2005) The flight of macromolecular complexes in a mass spectrometer. Philosophical Transactions of the Royal Society of London Series a-Mathematical Physical and Engineering Sciences **363**, 379-389
254 Thomas, C. E., Kelleher, N. L. and Mizzen, C. A. (2006) Mass spectrometric characterization of human histone H3: A bird's eye view. Journal of Proteome Research **5**, 240-247
255 Siuti, N., Roth, M. J., Mizzen, C. A., Kelleher, N. L. and Pesavento, J. J. (2006) Gene-specific characterization of human histone H2B by electron capture dissociation. Journal of Proteome Research **5**, 233-239
256 Pesavento, J. J., Garcia, B. A., Streeky, J. A., Kelleher, N. L. and Mizzen, C. A. (2007) Mild Performic Acid Oxidation Enhances Chromatographic and Top Down Mass Spectrometric Analyses of Histones. Mol Cell Proteomics **6**, 1510-1526
257 Garcia, B. A., Joshi, S., Thomas, C. E., Chitta, R. K., Diaz, R. L., Busby, S. A., Andrews, P. C., Loo, R. R. O., Shabanowitz, J., Kelleher, N. L., Mizzen, C. A., Allis, C. D. and Hunt, D. F. (2006) Comprehensive phosphoprotein analysis of linker histone H1 from Tetrahymena thermophila. Molecular & Cellular Proteomics **5**, 1593-1609
258 Patrie, S. M., Ferguson, J. T., Robinson, D. E., Whipple, D., Rother, M., Metcalf, W. W. and Kelleher, N. L. (2006) Top down mass spectrometry of < 60-kDa proteins from Methanosarcina acetivorans using quadrupole FTMS with automated octopole collisionally activated dissociation. Molecular & Cellular Proteomics **5**, 14-25
259 Meng, F. Y., Du, Y., Miller, L. M., Patrie, S. M., Robinson, D. E. and Kelleher, N. L. (2004) Molecular-level description of proteins from Saccharomyces cerevisiae using quadrupole FT hybrid mass spectrometry for top down proteomics. Analytical Chemistry **76**, 2852-2858

260 LeDuc, R. D., Taylor, G. K., Kim, Y. B., Januszyk, T. E., Bynum, L. H., Sola, J. V., Garavelli, J. S. and Kelleher, N. L. (2004) ProSight PTM: an integrated environment for protein identification and characterization by top-down mass spectrometry. Nucleic Acids Research **32**, W340-W345

261 Boyne, M. T., Pesavento, J. J., Mizzen, C. A. and Kelleher, N. L. (2006) Precise characterization of human histories in the H2A gene family by top down mass spectrometry. Journal of Proteome Research **5**, 248-253

262 Xia, Y., Thomson, B. A. and McLuckey, S. A. (2007) Bidirectional Ion Transfer between Quadrupole Arrays: MSn Ion/Ion Reaction Experiments on a Quadrupole/Time-of-Flight Tandem Mass Spectrometer. Anal. Chem.

263 Xia, Y., Chrisman, P. A., Erickson, D. E., Liu, J., Liang, X. R., Londry, F. A., Yang, M. J. and McLuckey, S. A. (2006) Implementation of ion/ion reactions in a quadrupole/time-of-flight tandem mass spectrometer. Analytical Chemistry **78**, 4146-4154

264 Kaplan, D. A., Hartmer, R., Speir, J. P., Stoermer, C., Gumerov, D., Easterling, M. L., Brekenfeld, A., Kim, T., Laukien, F. and Park, M. A. (2008) Electron transfer dissociation in the hexapole collision cell of a hybrid quadrupole-hexapole Fourier transform ion cyclotron resonance mass spectrometer. Rapid Communications in Mass Spectrometry **22**, 271-278

265 Udeshi, N. D., Shabanowitz, J., Hunt, D. F. and Rose, K. L. (2007) Analysis of proteins and peptides on a chromatographic timescale by electron-transfer dissociation MS. Febs Journal **274**, 6269-6276

266 Polson, C., Sarkar, P., Incledon, B., Raguvaran, V. and Grant, R. (2003) Optimization of protein precipitation based upon effectiveness of protein removal and ionization effect in liquid chromatography-tandem mass spectrometry. Journal of Chromatography B **785**, 263-275

267 Michaelis, L. and Menten, M. L. (1913) The kenetics of the inversion effect. Biochemische Zeitschrift **49**, 333-369

268 Usaite, R., Wohlschlegel, J., Venable, J. D., Park, S. K., Nielsen, J., Olsson, L. and Yates, J. R. (2008) 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

269 Nagalla, S. R., Canick, J. A., Jacob, T., Schneider, K. A., Reddy, A. P., Thomas, A., Dasari, S., Lu, X. F., Lapidus, J. A., Lambert-Messerlian, G. M., Gravett, M. G., Roberts, C. T., Luthy, D., Malone, F. D. and D'Alton, M. E. (2007) Proteomic analysis of maternal serum in Down syndrome: Identification of novel protein biomarkers. Journal of Proteome Research **6**, 1245-1257

270 Hilder, T. L., Malone, M. H., Bencharit, S., Colicelli, J., Haystead, T. A., Johnson, G. L. and Wu, C. C. (2007) Proteomic identification of the cerebral cavernous malformation signaling complex. Journal of Proteome Research **6**, 4343-4355

271 He, T., Roelofsen, H., Alvarez-Llamas, G., de Vries, M., Venema, K., Welling, G. W. and Vonk, R. J. (2007) Differential analysis of protein expression of Bifidobacterium grown on different carbohydrates. Journal of Microbiological Methods **69**, 364-370

272 Schmidt, F., Hustoft, H. K., Strozynski, M., Dimmler, C., Rude, T. and Thiede, B. (2007) Quantitative proteome analysis of cisplatin-induced apoptotic Jurkat T cells by stable isotope labeling with amino acids in cell culture,SDS-PAGE, and LC-MALDI-TOF/TOF MS. Electrophoresis **28**, 4359-4368

273 Shaw, J. L. V., Smith, C. R. and Diamandis, E. P. (2007) Proteomic analysis of human cervico-vaginal fluid. Journal of Proteome Research **6**, 2859-2865

274 Sharma, S., Simpson, D. C., Tolic, N., Jaitly, N., Mayampurath, A. M., Smith, R. D. and Pasa-Tolic, L. (2007) Proteomic profiling of intact proteins using WAX-

RPLC 2-D separations and FTICR mass spectrometry. Journal of Proteome Research **6**, 602-610