Development of Proteomic Techniques for Biomarker Discovery

Heather Allingham

Supervisor: Dr A R Pitt

University of Glasgow

Student Number: 0508874a

Abstract

The main aim of my research, presented here, is to develop proteomic research techniques, for their use in biomarker discovery and identification. This is broken down into three main chapters:

- Biomarker Identification in Stroke Brain guided by MALDI-imaging.
- Evaluation of the Effectiveness of Heat Treatment for Prevention of Proteomic Sample Degradation using Label Free Relative Quantitation.
- Discovery and Identification of Biomarkers for Hypertension

Within these chapters special attention is paid to sample preparation, development and assessment of new methods for biomarkers discovery and identification, the importance of experimental design and the application of relevant and useful statistical methods to enable the mining of useful information from rich datasets. The biological changes in stroke induced mouse brain tissue are studied, the prevention of degradation to tissue samples by a novel heat treatment method and changes to plasma samples from hypertensive, wild type and a congenic strain of rat are also studied. The outcomes of this work are multiple, namely:

- The identification of a possible marker for stroke from mouse brain tissue
- The effect of a new heat treatment device on proteomic data obtained from mouse brain tissue
- The novel application of a statistical analysis to a new type of dataset (LC-MS and label free quantitation of biological samples)
- The identification of possible biomarkers for hypertension in rat plasma using this method

Table of Contents

List of Tables

List of Figures

Acknowledgements

I would like to thank all the people who have helped me during the time I spent on experimentation in the laboratory and during the writing of this thesis. In particular, I would like to thank my supervisor, Dr Andy Pitt for his help and guidance over the last few years. Additionally, I would like to thank Dr Richard Goodwin for guidance throughout my research and particularly for reviewing and advising on this thesis.

I would also like to thank the BBSRC, EPSRC and MRC for funding me throughout this research and for funding in general of the Doctoral Training Centre in Proteomic Technologies and RASOR, without which I would not have been able to complete this research.

I would like to thank all those that work in the Sir Henry Wellcome Functional Genomics Facility in Glasgow University for all their help, assistance and patience. In addition I would like to thank all those within RASOR for their help and encouragement. In particular I must thank Dr Richard Burchmore, Dr Karl Burgess, Dr 'Jay' Jayawardena, Dr Sarah Cummings and Dr Susan Horne for all of their assistance in the laboratory.

I must also thank Dr Hilary Carswell, Dr Martin McBride and the Cardiovascular Research Group at Glasgow University for providing samples and, equally importantly, guidance and advice during this research. I would also like to thank Dr John McClure for invaluable advice and help with statistical analysis undertaken in this thesis. I would also like to thank the team at Denator for the loan of their instrument which was vital to part of this research.

In addition I would like to thank all the members of the Doctoral Training Centre for their help, support, encouragement (scientific and otherwise) and friendship. In particular I would like to thank Kat Thomson, Kit-Yee Tan, Scott Heron, Emma Carrick and Michael Lang for their moral support, fun and friendship over the period of this research and hopefully for years to come.

I would also like to thank all my non-scientific friends and family for their unfailing support and interest in my research, in particular my parents, my sisters, Helen and Emma, and my boyfriend Grant for putting up with me at my worst and encouraging me to finish when my resolve was wavering.

Author’s Declaration

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

Heather Allingham

1 Introduction

Proteomics is a field that has emerged in the post-genomic era (1). The sequencing of the human genome revealed that we have around 40 000 genes (2). Subsequently this number was reduced to 20 000 to 25 000 (3). Each human gene may encode several different proteins, resulting in a huge combination of different proteins in the human body. This means that studying genes alone cannot solve the many different mysteries of disease. If DNA is the blueprint of life then proteins are the building blocks from which life is constructed, and understanding their functions, expression and interactions is vital in the quest for information regarding life and disease.

Proteomics is essentially the study of the complete set of proteins in a cell, organ or organism (4), and the phrase was first coined in 1995 by Wasinger et al (5). While the genome is static, the proteome is dynamic and ever changing, either as a reaction to environment or due to development. Understanding these dynamic processes will be important in understanding mechanisms of disease, and one of the most important applications of proteomics is the discovery and identification of biomarkers for disease (6). Additionally, proteomics encompasses expression patterns, interactions and functional states of proteins present in a cell, organism or organ. Problematically the increased importance of proteomics has not been accompanied with an improvement in analytical tools and methods, in particular if compared to the tools available for the study of genomics such as sequencing techniques and expression microarrays. There are many challenges and issues to overcome in the study of proteomics. This review of the literature concentrates on the technologies and tools used in the research presented here, including liquid chromatography, mass spectrometry and label free relative quantitation, as well as alternative proteomic methods, such as differential gel electrophoresis, capillary electrophoresis and isotopic labelling for quantitation. The theory behind these technologies is also covered, as well as the thought process behind biomarker discovery and identification. The diseases studied and more specific techniques used are covered in the short introduction at the start of individual chapters.

1.1 Biomarker Discovery

Diagnosis is one of the most important stages in treating disease, and is the first line of defence available to a clinician. Sometimes diagnosis will be visual, or may be the result of, for instance, a biopsy. Early diagnosis and selection of appropriate treatment can be

critical, and mean the difference between a failing or successful treatment for certain diseases. There are many treatments which have been developed, that are not always successful due to late diagnosis. It is thought that biomarkers in body fluids may be one of the solutions to this problem.

A biomarker is a something which can be measured and used to diagnose or classify disease and also to measure progress and therapeutic response of disease. Body fluids pass through tissues, including diseased tissues, and collect biomolecules, including peptides and proteins, that can indicate the status of health. If these biomolecules can indicate the presence of a disease they are a type of biomarkers. The ideal marker, in the context of this work, is a biomolecule that is present only in diseased patients and absent in healthy controls, or vice versa. In reality a biomarker will most likely be detected as a change in abundance of the marker or relative abundance between markers. Additionally, markers can also occur as more than one biomolecule, for instance a disease could be characterised by the presence of a set of markers. This means that it is important to characterise as many biomolecules as possible, with high accuracy and resolution (7). Biomarkers can be detected in different body fluids such as blood, urine, cerebrospinal fluid and saliva. Biofluids such as plasma and urine, do not have an associated genome or transcriptome, meaning that gene expression measurement is not possible (8). This is why proteomics, as well as metabolomics, is such an important field for biomarker discovery in biofluids (9).

Urine has been an attractive fluid for the study of biomarkers as it is non-invasive to obtain in large quantities and is less complex than blood or plasma, due to the much lower protein concentration (10). Repeated sampling from the same individual is also commonly achievable. A further attractive quality of urine for biomarker discovery is that it contains proteins and peptides in their native state that are of a relatively low molecular mass and are highly soluble. This means they can be analysed in their natural state with little manipulation (11). Urine is also a relatively stable fluid, and the urinary proteome has been shown to be stable at 4 ºC for up to 3 days or at room temperature for up to 6 hours. It is thought that this is due to its storage for several hours in the bladder, meaning that proteolytic degradation by endogenous proteases is complete by time of voiding (11). Urine can be especially useful for diagnosis of kidney disease (12). However, urine does have certain disadvantages as a fluid for biomarker discovery. Urine samples have a wide variation in protein concentration, due to fluid intake of the individual. This can be normalised using creatinine concentration (13) or other peptides present in urine (14) but this does add an additional complication to the analysis. Urine samples also have inconsistent pH levels that can alter protease activity, leading to greater sample variability

(15). There are also clear differences between early and mid-stream urine samples (16), meaning that sample collection must be well controlled (17).

Plasma, or serum, is commonly studied with regards to cardiovascular disease (18-23) and cancer (24-28), as the biomolecules found in blood directly reflect the health of an individual and can indicate the presence and stage of disease. Serum represents the soluble fraction of blood that remains following clotting and is therefore simpler than plasma, which contains clotting factors. Plasma is obtained from blood following the addition of an anti-coagulant, such as heparin or EDTA, in the presence of an inert powder, such as glass beads. The advantage of using serum over plasma is its reduced complexity and lower protein concentration, however the clotting process is not uniform, unlike the preparation of plasma, and interesting proteomic and metabolomic features may be lost with the insoluble blood clot (8). Additionally, an array of proteases are activated when clotting occurs, generating degradation products, meaning that the protein profile of serum is quite different from that of plasma. As a consequence of this the Human Proteome Consortium has recommended that plasma is the fluid of choice, as opposed to serum, for proteomic studies (29). Plasma potentially contains elements of all proteins in the body meaning plasma has a high potential for biomarker discovery. Plasma is studied in the research presented here, specifically in Chapter 4, “Discovery and Identification of Biomarkers for Hypertension”.

Cerobrospinalfluid is another alternative for biomarker discovery, particularly with regards to neurodegenerative diseases (30).

Tissue samples are also studied in this work in Chapter 2, “Biomarker Identification in Stroke Brain Guided by MALDI-imaging” and in Chapter 3, “Assessment of Denator Heat Treatment for Prevention of Proteomic Sample Degradation using Label Free Relative Quantitation”.

The local concentration of any biomarker will be high in the vicinity of the diseased tissue itself, particularly with regards to diseases such as cancer or stroke, where a clear site of disease is present. Fine-needle aspiration biopsy is a method used to obtain samples from, for example, cancerous tumours (31). An obvious disadvantage of this is the invasive nature of this procedure. Also, there may be many different classes of cell types present in biopsies of this type, hence the development of laser microdissection (LMD) techniques. LMD is used to rapidly separate populations of cells from tissue, which has been thinly sliced using a cryostat device, for further analysis (32). However, LMD can subject

samples to subjective sample processing; both at the selection of relevant cell types and also during the dissection process itself. It should be taken into consideration that this can have an impact on the quality and results of subsequent studies (33). Formalin-fixed, paraffin-embedded (FFPE) tissue is a commonly encountered clinical sample from biopsy samples or samples of a surgical origin and can be a valuable resource. Unfortunately the protein cross-linking, due to the formalin fixing, makes protein profiling by western blot and protein microarray analysis difficult (34;35). Currently, immunohistochemistry is the most routinely used method for analysis of FFPE tissue, which is a guided technique, whereby a particular protein is stained using tagged antibodies. There remains a lack of data about protein expression in normal and cancerous cells, due to the difficulties associated with dealing with these samples (36;37). It has been shown to be possible to disrupt the cross-linked proteins in FFPE tissue for protein identification using mass spectrometry (36), meaning it may be possible to unlock the data potential from these types of samples.

Biological samples and proteins, protein complexes and biomarkers in particular are often very complex. This means that an effective, unbiased proteomic profiling approach, without any pre-selection, may require a multi-discipline approach to accomplish protein identification and quantitation. This means that there are several optimisation steps in method and technology development, including sample collection, storage and preparation, protein and peptide separation, protein and peptide identification and quantitation, including statistical analysis, and finally independent confirmation and validation. These steps are discussed further in following sections.

Clinically relevant biomarkers (i.e. markers that can be measured in a clinical sample, such as plasma or urine) can provide a “snap shot” view of a specific stage of disease. Therefore it is important to identify specific plasma proteins without complication by the presence of abundant plasma proteins. This is discussed in more detail in section 1.10 and additionally in Chapter 4, “Discover and Identification of Biomarkers for Hypertension”. Valuable, usable biomarkers are specific, sensitive and reproducible. It is worth noting that the majority of FDA approved biomarkers for cancer are single proteins found in serum, CA19-9 for pancreatic cancer, CA125 for ovarian cancer and PSA for prostate cancer (38;39).

In my work, presented here, an unguided proteomic method is developed and used for biomarker discovery, assuming no prior knowledge of disease. This means that not only can possible biomarkers be unearthed, but possibly mechanisms of disease can be

understood and previously unknown functions of disease discovered. This method has been utilised by Kiga *et al* (40) for the identification of stroke plasma biomarkers in spontaneously hypertensive stroke prone rats, which is the same genetic model used in chapter 4 of the work presented here. Kiga *et al* used an unguided biomarker discovery method and successfully identified haptoglobin as a biomarker candidate for stroke. This identification was validated using traditional biochemical methods. Similarly, Silbger *et al* (41) have used an unguided biomarker discovery method for human myocardial infarction plasma samples. The combination of mass spectrometry with bioinformatic methods enabled the identification of potential biomarkers. The use of these methods is discussed in more detail in the introduction to chapter 4 "Discovery and Identification of Biomarkers for Hypertension".

1.2 Sample Collection and Storage

As sample collection and storage is the initial step in any study it is vital that it is performed correctly. It has been shown that sampling procedures have the greatest effect on proteomic data from plasma samples, whereas handling and storage have relatively small effects (42). However, it is still important that standardised protocols for sample handling and storage are developed and adhered to, if results are ever to become comparable between laboratories (43). Studies have also shown that the standardisation of sample tubes is important for reliable plasma analysis (44). It should be considered that commercially available blood collection tubes contain many components, including silicones, PEGs and polymers, which may interfere with mass spectrometric analysis (45). Additionally, the sample tubes used will influence adsorption of plasma protein onto the inner surfaces of the tubes. For instance, significant differences have been seen in the mass spectrometric data from serum samples collected in glass tubes containing no anticoagulant in comparison with plastic serum separator tubes (46).

A further consideration is the anti-coagulant used, as this can also influence the proteomic data obtained from the plasma samples (47). Platelets tend to be more stable in citrate anti-coagulants, but this is often used in liquid form which dilutes the plasma samples. Heparin is used to enhance the activity of antithrombin III, but it also binds to other proteins (48). EDTA can lead to platelet clumping and aggregation, meaning that plasma protein profiles from samples treated with EDTA are most different from those obtained from samples treated with citrate or heparin (42;49).

Plasma samples are also affected by storage conditions, with minimal changes detected within 6 hours at room temperature, but noticeable changes detected after 8 hours at room temperature, with even more pronounced changes after 24 hours (50). For plasma or serum stores at 4ºC minimal changes were seen after 24 hours, but more significant changes were seen after 48 and 96 hours (51). Samples stored at -20 ºC, -80 ºC or in liquid nitrogen show no significant changes, even with long term storage (43;51), however freeze/thaw cycles can change sample composition, due to protein precipitation, aggregation and surface adsorption (46;47).

As there are many steps in sample collection and storage that can influence sample composition and proteomic profiles it is vital that these steps be kept constant across a study, if data is to be in any way comparable. Ideally, universally used standardised protocols would be accepted, creating comparable datasets between laboratories.

The sample collection and preparation methods used in my research will be discussed further in the introduction to each chapter.

1.3 Mass Spectrometry

Recent advances in mass spectrometry (MS) have been one of the major driving forces behind progress in proteomics (52). A mass spectrometer is able to determine the molecular weight of chemical compounds by ionising, separating, and measuring ions according to their mass-to-charge ratio (m/z). The ions are generated in the ionisation source by inducing either a positive or negative charge. Once the ions are formed in the gas phase they can be directed into a mass analyser, separated according to mass and finally detected. The result of ionisation, ion separation, and detection is a mass spectrum that can provide molecular weight or even structural information.

In the research presented here, an electrospray quadrupole time-of-flight mass spectrometer is the main instrument used. This is the main focus of the introduction to mass spectrometry discussed here.

The mass-to-charge ratio of the ions can be determined in several ways, including the exact time taken for the ion to reach the detector (used in time-of-flight mass analysers, TOF), or by trapping the ions in an electric or magnetic field and sequentially filtering them out based on size (with the smallest first) by scanning the rf-voltage making ion trajectories sequentially unstable (used in trap analysers). This can then be used to identify an

unknown sample, by the mass of the molecule and of the fragments created. The six primary components of any mass spectrometer are the vacuum system, sample introduction device, ionisation source, mass analyser, ion detector and data collection and storage.

Until the early 1980's mass spectrometry was not routinely applied to proteins, as proteins were too fragile and of too large a size to withstand the processes of ionisation. Developments of 'soft' ionisation techniques (53), such as electrospray mass spectrometry (ESI-MS) and matrix assisted laser desorption ionisation mass spectrometry (MALDI-MS) have overcome these problems, allowing the detailed analysis of proteins. Mass spectrometry has now become an important tool for characterisation and identification of proteins and peptides from biological samples and advances in this area have enabled the identification of proteins with an unprecedented speed and sensitivity.

1.3.1 Time-of-flight Mass Spectrometry

The significance of TOF mass analysers has grown over the last 20 years, especially in the biomedical fields. The TOF mass analyser was originally described in 1946 by Stephen (54). By 1951 Goudsmit *et al* (55) were using the TOF mass spectrometry to measure the time of flight of ions for a number of complete revolutions in a magnetic field allowing, in particular, the mass of heavier elements to be determined.

In a TOF analyser, ions are separated by differences in their velocities as they move in a straight path toward a collector in order of increasing mass-to-charge ratio. The first commercial instrument was marketed by the Bendix Corporation in 1955, and was based on a design by Wiley and McLaren (56).

Put in the simplest terms, time-of-flight analysis works by accelerating a set of ions to a detector with the same amount of energy. As the ions have the same energy but a different mass, the ions reach the detector at different times. The lower mass ions reach the detector first because of their greater velocity and the higher mass ions take longer.

The arrival time of an ion at the detector is dependent upon the mass, charge, and kinetic energy of the ion. The TOF analyser consists of a field free drift region, D, which is held at constant electrical potential, and a source region, S, where a gradient in electrical potential exists, V_s. Ions which are present in the short source region are accelerated into the drift region by the action of the electric field, E_s. Analysis is based on the principle that ions of

different m/z values will have been accelerated to the same kinetic energy, E_K, in the source:

$$E_K = zeE_S \qquad (1.3.1)$$

where e is the charge on an electron in coulombs and z is the number of charges. However, ions of different mass will exhibit different velocities (v):

$$v = (2zeE_S / m)^{1/2} \qquad (1.3.2)$$

where m is the mass of the ion. The time, t, taken for ions to traverse the length, d, of the drift region is also dependent on the mass of the ions:

$$t = \left(m / 2zeE_S \right)^{1/2} d \qquad (1.3.3)$$

Thus, the time that is required for each ion to traverse the drift region is different: high mass ions take longer to reach the detector than low mass ions. The basic formula by which the time spectrum can be converted to a mass spectrum is given in the equation 1.3.4:

$$m/z = 2eE_S \left(t/d \right)^2 \qquad (1.3.4)$$

A major limitation in achieving higher resolution in TOF mass analysers is the consequence of the distribution in time, in space and in kinetic energy of the initial ion packet. It must be assumed for the purposes of calculation that all the ions will be accelerated from the same position and have the same E_K. However, this is not the case as there are a number of sources producing inaccuracy associated with ion formation and acceleration (57).

1.3.2 Quadrupole Mass Spectrometry

The quadrupole mass filter was first developed in the 1950s by Paul et al (58). Paul shared the Nobel Prize in Physics in 1989 for his work on ion trapping. In a quadrupole device, a

quadrupolar electrical field (comprising radiofrequency and direct-current components) is used to separate ions (59;60). Another instrument that Paul developed was the quadrupole ion trap, which can trap and mass-analyse ions using a three-dimensional quadrupolar radiofrequency electric field. The first commercial ion trap system was introduced in 1983 by Finnigan MAT (San Jose, CAL, USA.), originally as a gas chromatography detector. More recently, ion trap instruments serve not only as gas chromatography detectors but also as liquid chromatography detectors and stand-alone mass spectrometers (61).

Quadrupoles are made up of four parallel rods with a direct current (DC) voltage and a superimposed radio-frequency (RF) potential. Consequently, adjusting the RF potential and the level of the DC field component can effectively scan a mass range. Quadruples are tolerant of relatively poor vacuums and this makes them well-suited to electrospray ionisation, as the ions are produced under atmospheric pressure conditions. Quadrupoles are now capable of routinely analysing up to a m/z of 3000. This is useful because electrospray ionisation of proteins and other biomolecules commonly produces a charge distribution below a m/z of 3000. Additionally, the relatively low cost of quadrupole mass spectrometers makes them attractive as electrospray analysers.

Another type of mass spectrometer used in my research is the ion trap. The physics behind the quadrupole mass analyser and the quadrupole ion trap mass analyser is very similar. In an ion trap the ions are trapped in a radio frequency quadrupole field. One method of using an ion trap for mass spectrometry is to generate ions externally before injecting them into the trapping volume. The ions are then ejected and detected as the radio frequency field is scanned. It is also possible to isolate one ion species by ejecting all others from the trap. The isolated ions can subsequently be fragmented by collisional activation and the fragments detected to generate a fragmentation spectrum. The primary advantage of quadrupole ion traps is that multiple collision-induced dissociation experiments can be performed without having multiple analysers. Other important advantages include its relatively small size and the ability to trap and accumulate ions to increase the signal-to-noise ratio of a measurement. Quadrupole ion traps have been utilised in many applications, mainly for the analysis of peptides and small molecules.

MS can be used for peptide mass fingerprinting (PMF) (62). This is a technique that is used to identify proteins by matching the masses of the fragments to the theoretical peptide masses obtained from a protein database. Specific proteases, such as trypsin, cleave proteins at certain points in the amino acid chain. This means that it is possible to theoretically digest a protein, and predict the peptides that will be produced. As the masses

of amino acids are accurately known, a list of predicted peptide masses can be produced. This peptide mass information can be used to produce a unique ‘peptide mass fingerprint’ for the protein, which can then be used for fast identification of unknown proteins in a sample. Different databases have been developed which contain vast amounts of protein data and predicted peptide masses. These can be used to search mass spectrometric data against, and therefore identify proteins quickly and uniquely. A statistical scoring system is used to determine the accuracy of identifications.

1.3.3 Tandem Mass Spectrometry

In tandem mass spectrometry (MS-MS) the ions initially produced are then fragmented and analysed further. Tandem in space instruments, such as the Q-TOF, requires more than one analyser. The first is required for analyte isolation. The ions then pass through a collision cell where they are fragmented with an inert gas such as argon or nitrogen, and a second analyser is used for detection of the fragment mass to charge ratios. Quadrupole mass filters work well for MS-MS analysis as they can be used to allow only ions in a pre-defined mass range to pass through the quadrupole, enhancing the sensitivity of the technique.

Additionally, MS-MS can be used in a targeted manner. For instance, if certain features are of a particular interest, their masses can be chosen for preferential fragmentation, in order to attempt to ensure identifications are made. This method is used in my research presented here, namely in chapters 3 and 4.

One of the major problems with biological samples, such as blood or urine, is that they can be extremely complex. This means that a separation technique or pre-fractionation method is needed prior to analysis by mass spectrometry. Commonly technologies such as high performance liquid chromatography (HPLC) and capillary electrophoresis (CE) are used to separate samples. This has resulted in particular interest in the development of methods for interfacing liquid separation techniques directly to mass spectrometry. This introduces the complication of attempting to maintain a high vacuum within the mass spectrometer, with various flow rate interfaces. Capillary electrophoresis uses flow rates of around 100 nL/min whereas HPLC can use flow rates of over 1 mL/min. An additional consideration is the need to vaporise the eluent prior to ionisation.

1.3.4 Ionisation for Mass Spectrometry

There are many different techniques used for ion production in mass spectrometry. These include amongst others: electrospray ionisation (ESI) (63-66), micro-electrospray ionisation (micro-ESI) (67;68), nano-electrospray ionisation (nESI) (69;70), fast atom bombardment (FAB) (71), atmospheric pressure chemical ionisation (APCI) (72), plasma desorption (73), laser ablation/desorption ionisation (LA, LDI) (74;75), and the workhorse of many biological studies, matrix assisted laser desorption ionisation (MALDI) (76).

MALDI mass spectrometry is a form of laser desorption mass spectrometry, where sample molecules are laser-desorbed from a solid or liquid matrix containing a highly UV-absorbing substance. The matrix protects the biomolecules from being destroyed by the laser beam and also facilitates vaporisation and ionisation, making it a widely used technicue in biological analysis. Additional advantages of MALDI mass spectrometry are that it is a sensitive technique with low sample consumption and is fairly quick and simple to use.

Electrospray ionisation mass spectrometry (ESI-MS) techniques have had an influential impact on the ability to use mass spectrometry for the study of large biomolecules. Electrospray has become the most widely used LC-MS interface technique in modern mass spectrometry as it is a versatile mode for the production of gas phase analyte ions from solutions and is almost ideal for biological analytes. Chapman may have used some of the initial elements of ESI with work performed in 1937 (77). However, it was research by Dole and colleagues in the 1960s that is usually credited as the birth of electrospray ionisation (78). It was not until the mid 1980s that its use as an interface for mass spectrometry was demonstrated by Fenn, revealing the power of the technique (79;80) and winning Fenn the Nobel prize for Chemistry in 2002.

In ESI-MS, highly charged droplets dispersed from a capillary in an electric field are evaporated and the resulting ions are drawn into a mass spectrometry inlet. The charged droplet travels through the ambient air for a few milliseconds (81). During this time the radius of the droplet decreases due to solvent evaporation and the droplet approaches the point where the forces of surface tension and the Coulomb repulsion counterbalance each other. It then suffers what is commonly referred to as ‘Coulomb explosion’ or ‘Rayleigh fission’, after Lord Rayleigh (1882).

It has been shown that, under the influence of aerodynamic forces corresponding to ESI-MS conditions, droplet fission occurs already at charge states of about 80% of the Rayleigh limit, where the forces of surface tension and the Coulomb repulsion counterbalance each other. All experimental investigations have shown that there was no ‘explosion’ of the droplet, but that the droplet surface formed a cone-like shape, from which a number, approximately 20, of smaller droplets are ejected (82;83). The primary droplet loses about 15% of its charge and about 2% of its mass (84). From the primary droplet there are two or three successive uneven fissions until the charge state of the droplet becomes too low for a further fission. Following the evaporation of the remaining solvent the contents, analyte and electrolyte molecules and some additional ions, end up as a lowly charged residue. The daughter offspring droplets, however, undergo further fissions following the method of evaporation and fission just described.

Finally, the evaporation of the remaining solvent leads to a ‘charge condensation’ on the analyte molecule and hence to a gas-phase analyte ion. It is thought that the resultant analyte molecules transferred to the gas-phase are not ‘dry’ gas-phase ions. This is due to a gradual transition from a very small charged droplet via a charged solvent-analyte cluster to a gas-phase molecule to which a number of charges, are attached (85). This is particularly believed to be the case for large biomolecules (86).

Impurities may also be attached if they are present in the droplet. However, it has been reported that when proteins or peptides are the analyte, the impurity must be present in orders of magnitude higher concentration before complete suppression of the analyte ion is obtained in the mass spectrum. A possible explanation for this fact is that an influence of the analyte surface activity may lead to enrichment of analyte molecules in the offspring droplets, which are then preferentially released as ions, while the non-surface active impurities are more likely to remain in the lowly charged residue (87).

Following repeated cycles of solvent evaporation and coulombic fissions by the droplets the gas-phase ions generated are guided to the counter electrode. The process by which gas-phase ions are generated is still under considerable debate with two main theories proposed; namely the ion evaporation model and the charge residue model.

The charge residue model relies on the formation of extremely small droplets, radius of approximately 1 nm, which contain only a single ion. Solvent evaporation from the droplet will lead to the conversion of the droplet to a gas phase ion. Such a model was assumed to be the case by Dole *et al* (78) who first investigated gas-phase ion production by

electrospray. This charge residue model for gas-phase ion generation is now also known as the single ion droplet theory (SIDT).

The ion evaporation model assumes that before the droplet reaches this ultimate stage prior to gas-phase ion generation, the field on the droplets surface becomes strong enough to overcome solvation forces and lifts a solute ion from the droplet surface into the gas-phase. This model was first introduced by Iribarne and Thomson (88) in 1976. Typically the droplets from which ion emission/evaporation becomes competitive with Rayleigh fission have a radius of $R \approx 8$ nm and $N \approx 80$ charges (88). It is proposed that under such conditions the droplet no longer undergoes further fission but emits gas-phase ions. Also as the charge, N, decreases emission can continue as the droplet radius, R, also decreases from further solvent evaporation. This model therefore allows the gas-phase ion production without the formation of extremely small droplets, $R \approx 1$ nm, which contain a single ion. Also such emission can occur even when the droplet contains other solutes such as charge-paired-electrolytes (89).

At present the general consensus is that neither the charge residue model nor the ion evaporation model can account for all the experimental observations, and it has therefore been speculated that the two models describe different aspects of the observed process of gas-phase ion formation (90).

In the mid 1990s Wilm and Mann described a revolutionary development to electrospray ionisation, which they described as microelectrospray (91). The idea they used was to apply a narrow, fused-silica capillary as an ESI emitter, instead of the routinely used, relatively wide, steel or fused silica capillaries. This reduced the internal diameter of the tip from 100 μm to 20 μm, leading to a decrease in the initial size of liquid droplets. In turn this reduced sample consumption and increased the ionisation yield, resulting in higher sensitivity of the system. Around the same time, in 1994, Caprioli *et al* reported a modified electrospray ion source, where the fused-silica capillary filled with the reverse phase material was used as the emitter needle (92). The same group then reported the sensitivity of the nanospray device, demonstrating a successful analysis of a peptide at a concentration of 500 zmol/L (93).

There were further improvements to the nanospray system by Davis *et al* (94) and Gatlin *et al* (95). These were based on the integration of column packing with the emitter tip, resulting in a system free from the after-column void volumes. Reduction of these void

volumes to a minimum plays an important role in micro and nanoscale liquid chromatography.

Furthermore, following the development of nanospray electrospray ionisation (nESI) greater insight has been gained due to the decreased size of the primary droplet. In conventional ESI, with μL/min forced-flow rates, the droplets formed are in the μm range. However, for nESI, where the ESI capillary outlet is replaced with a tapered glass pipette with a small orifice, within the region of 5 to 15 μm, the resulting flow rate is approximately 50 nL/min. In this case the droplet diameter has been estimated at 180 nm (91;96).

The low surface activity of analytes is therefore less problematic when dealing with nanospray, as gas-phase ions can be produced directly from primary droplets, unlike ESI for which gas-phase ions are produced almost exclusively from offspring and satellite droplets. The ionisation efficiency is also greatly increased as there is sufficient time for the primary and residue droplet ions to be entirely transferred to the gas-phase, due to their very high surface-to-volume ratio.

There is an increased tolerance to salt contamination when using nESI. This can be explained by the theory that the formation of droplets with a diameter one order of magnitude smaller means that fewer fission events are necessary before ions are released, and each additional fission event results in an increase in salt concentration.

However, one of the challenges with the development of nESI was that the system could not be connected to the typical, analytical HPLC system due to the much higher flow rate of the mobile phase. This will be discussed further in the following section.

To summarise, the ion evaporation model is believed to be the process for gas-phase ion formation for small ions (e.g. salt ions); however, the complications of experimental study at such small droplet size and charge prevent sufficient empirical results. For macro ions (for instance proteins larger than 6500 Da) the charge residue model is more likely (84;96).

1.4 Separation Techniques

1.4.1 Liquid Chromatography

Liquid chromatography (LC) is an analytical technique, used for separating mixtures of many types. In this work it will be applied to body fluids, which can essentially be thought of as complex mixtures of biomolecules. Liquid chromatography involves passing a sample, in a solution known as the ‘mobile phase’ over a solid ‘stationary phase’, which is packed into a column (97). The stationary phase retards the movement of the sample selectively. Each sample component has a characteristic time of movement through the column, known as the ‘retention time’. In High Performance Liquid Chromatography (HPLC) the analyte is forced through a column, containing the stationary phase, at a high pressure. Normal phase liquid chromatography was the first HPLC method, which uses a polar stationary phase, and a non-polar mobile phase. It is not commonly used now as water can alter the hydration state of the chromatographic media, causing problems with reproducibility of retention times. Reverse phase (RP) HPLC is now more commonly used.

RP-HPLC uses a non-polar stationary phase, such as a silica surface modified with long chain hydrocarbons, and a polar mobile phase, such as water. The separation is based on the hydrophobicity of a sample. The sample components will bind selectively to the stationary phase and their movement is slowed through the column. Strongly binding components will move slowly and weakly binding components will move quickly. The mobile phase is initially a high water content solution. This allows strongly binging components to bind to the stationary phase, and weakly binding components to be separated. The organic content of the mobile phase is gradually increased, which slowly displaces bound components off the column. This ‘gradient’ technique increases the resolution, and lowers the peak broadening effects.

It is well recognised that peak broadening originates from three main sources:

- Multiple path of an analyte through the column packing
- Molecular diffusion
- Effect of mass transfer between phases

In 1956 Van Deemter introduced an equation (98) which combined all three of these sources. It represented them as the dependence of the theoretical plate height (HETP) on the mobile phase linear velocity.

The plate model supposes that the chromatographic column contains a large number of layers, called *theoretical plates*. There are separate equilibriums of the sample between the stationary and mobile phase, which occur on each plate. The analyte moves down the column by transfer of the mobile phase from one plate to the next. It is important to remember that the plates do not actually exist, but are a theoretical model which is used to understand the processes that occur in the column. The plates can also be used as a way to measure column efficiency, either by stating the number of plates in a column, N (the more plates; the more efficient), or by stating the plate height, HETP (the smaller the plates; the more efficient).

If the length of the column is L, then:

$$HETP = L / N \qquad (1.4.1.1)$$

The number of theoretical plates that a real column possesses can be calculated using the following equation:

$$N = \frac{5.55 t_r^2}{w_{1/2}^2} \qquad (1.4.1.2)$$

Where t_r is the retention time of the analyte and $w_{1/2}$ is the half-height peak width of the peak in question. As can be seen from this equation, columns have different numbers of theoretical plates for different analytes.

The van Deemter equation was originally written for gas chromatography, but HPLC has the same physical processes and the equation also fits. This is explained as follows:

The analyte will take multiple paths through the column. Band broadening is caused by the differing flow velocities through the column:

$$H_p = 2\lambda d_p \qquad (1.4.1.3)$$

Where H_p is the HETP arising from the variation in the zone flow velocity, d_p is the average particle diameter and λ is a constant, which is close to 1. This shows that H_p may be reduced (increasing efficiency) by reducing the particle diameter. λ depends on the distribution of the particle size.

Molecules will disperse due to diffusion. Longitudinal diffusion (along the column) leads to band broadening. This is described as follows:

$$H_d = 2\frac{\gamma D_m}{v} \qquad (1.4.1.4)$$

H_d is the HETP arising from diffusion, D_m is the analyte diffusion coefficient in the mobile phase, γ is the factor related to the diffusion restriction by column packing and v is the flow velocity. It can be seen from this equation that the higher the eluant velocity, the lower the diffusion effect is on the band broadening. Molecular diffusion in the liquid phase is much lower than in the gas phase, and this effect is close to negligible at standard HPLC flow rates.

Mass transfer is the third parameter of interest. This is the combination of adsorption kinetics and mass transfer inside the particles. Adsorption kinetic is almost negligible compared to diffusion inside the particles and the band broadening due to this effect is expressed as follows:

$$H_m = \omega\frac{d_p^2}{D_m}v \qquad (1.4.1.5)$$

Where d_p is the particle diameter, D_m is the diffusion coefficient of the analyte in the mobile phase, ω is the coefficient determined by the pore size distribution, shape and also particle size distribution and v is the flow velocity. It can be seen from this equation that the HETP has a linear dependence on the flow rate, as the slower the velocity, the more uniformly analyte molecules penetrate inside the particle and the less the effect of different penetration on the efficiency. Conversely, at faster flow rates the elution distance between molecules with different penetration depths will be high.

Each term described above has an effect on the band broadening, so the sum of them all gives the total column plate height:

$$H = 2\lambda d_p + 2\frac{\gamma D_m}{v} + \frac{\omega d_p^2}{D_m} v \qquad (1.4.1.6)$$

The liquid chromatograph can be interfaced with different detectors for analysis of the sample. Commonly used detection systems are UV-Visible absorption (99;100), fluorescence (101) and mass spectrometry for a more detailed analysis of the sample (102-104). HPLC is probably the most commonly used separation technique, in conjunction with mass spectrometry and it is the method of choice in this work.

As previously mentioned, with the development of nESI came the need for liquid chromatography systems with lower flow rates, and this is the type of instrument employed in my research.

Initially, post column splitters were used to reduce flow rates. However, this reduced the sensitivity of the analysis to such a degree to render it useless. The only practical solution was to design capillary, high performance liquid chromatography. In theory, this is a simple technique. In essence the capillary HPLC system is purchased, connected to a nano-ESI-MS instrument and sample is analysed. In reality this type of system can have many challenges associated with it (105).

Typical nano-LC systems usually have flow rates ranging from 1μL/min to 10nL/min and use very narrow RP columns (often 75 μm internal diameter). The columns used are often made of fused silica, and sometimes carbon, titanium or stainless steel. The stationary phase is usually made up from beads with diameters of 5 μm, or smaller. An alternative to the traditional columns are monolithic column, which are prepared using a polymerized, porous filling. Such tiny columns are easily blocked, with even the submicroscopic solids causing problems. An additional issue is that there is no possibility to back flush these columns, so once blocked they are permanently damaged.

Further issues can be encountered if the column is overloaded. This should not permanently damage the column, however, several blank runs may need to be used to ‘clean’ the column and prevent contamination in the following samples. Additionally, there can be many problems encountered with leakages and void volumes in these tiny systems.

LC-MS can be used in many different ways for proteomic analysis. One type of LC-MS-MS analysis that is used in my research, in chapter 2, is known as MudPIT

(multidimensional protein identification technology) (106). This involves a 2-dimensional separation and is often used to identify proteins and peptides from complex biological mixtures. It can be visualised in the following way, with two chromatography steps, as shown in figure 1.1:

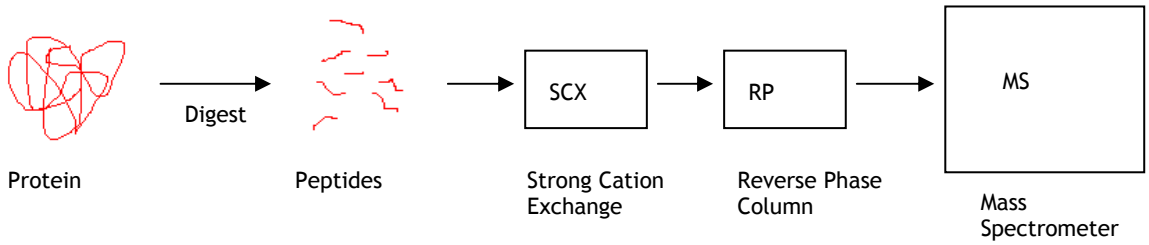


Figure 1.1: Schematic demonstrating the process of MudPIT (multidimensional protein identification technology). It is shown here that the protein samples are digested into peptides. Following that there are two separation steps using a Strong Cation Exchange Column and also a Reverse Phase Column.

This method is utilised in Chapter 2 in the research presented here.

1.4.2 Capillary Electrophoresis

Capillary electrophoresis (CE) has become an important bioanalytical separation technique, for proteins and peptides (107). Some of the advantages of CE include that it is of a high speed and resolution, it is sensitive and requires only a small amount of reagent, and is easily automated for high throughput analysis. Proteins create certain challenges from a separations point of view (108), as they can possess a wide range of physicochemical properties.

Capillary electrophoresis is an automated analytical technique that separates sample components by applying a voltage across a capillary filled with a buffer solution. The ends of the capillary are dipped into reservoirs of the buffer, and electrodes made of an inert material are dipped into the buffer reservoirs to complete the electrical circuit. The buffer solution will conduct current through the inside of the capillary. The sample is injected into one end of the capillary and is separated according to its size and charge. Migration through the column is dependent on the electroosmotic flow. In silica capillaries, when a buffer solution above around pH 3 is used, the surface silanol groups are ionised forming a fixed anion layer on the surface of the capillary. Cations from the buffer solution then form a mobile layer next to the fixed anion layer. This layer of cations is pulled towards the cathode, pulling the buffer solution and sample with it. This is demonstrated in figure 1.2.

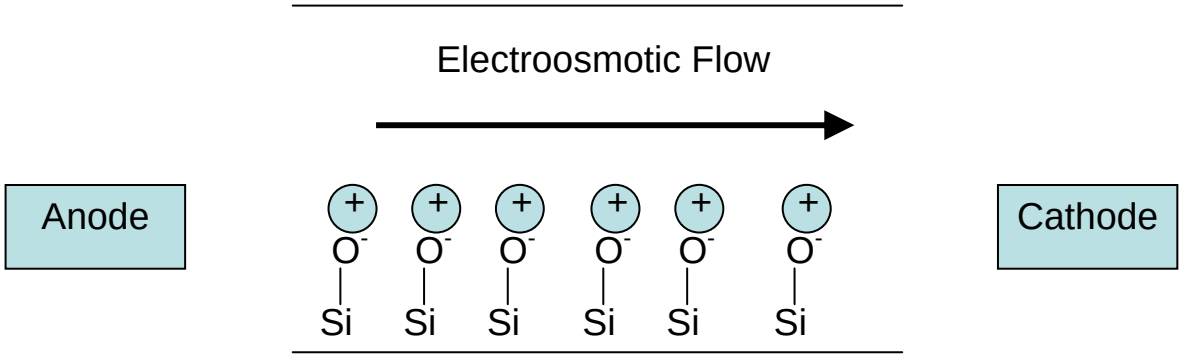


Figure 1.2 Electroosmotic Flow

This means that there is little drag from the edges of the capillary giving a very high-resolution separation. The flow profile in capillary electrophoresis is flat, as opposed to the laminar flow profile seen in pressure driven systems such as chromatography. Electroosmotic flow does therefore not contribute to band broadening, making CE a higher-resolution technique.

Common modes of detection after separation by CE include UV-Visible absorbance, laser induced fluorescence and mass spectrometry (109). UV-Visible absorbance is probably the most commonly used detection method for CE and also the most simple, but it is the least sensitive detection method. Detection is often measured at 214 nm where absorption is proportional to any peptide bonds that may be present. Different types of detection cells are available; including the Z shaped cell and the bubble cell. These cells help to increase the sensitivity of the detection.

Laser induced fluorescence (LIF) detection is more sensitive than UV-Visible absorbance, but is slightly more complicated to implement (110). However it is simpler to use than on-line mass spectrometric detection. Since proteins do not produce much natural fluorescence, a derivatisation step is often needed, although it is possible to use native fluorescence from aromatic residues (such as tyrosine or tryptophan) for detection. LIF as a quantitative method is discussed in more detail in section 1.6.

Mass spectrometry is the detection technique that can give most information about a separated proteomic sample, but interfacing between CE and MS can be challenging. CE can be coupled online to an electrospray mass spectrometer (111), or to a spotter for MALDI mass spectrometry analysis. The drawbacks of coupling CE to online mass spectrometry are that a relatively complicated interface is required. Sheath flow interfaces are the most popular. This interface implements a sheath liquid that flows around the end

of the capillary, therefore completing the electrical circuit and allowing the electrophoretic separation to take place.

1.4.3 Biomarker Discovery using CE-MS and LC-MS

CE-MS has been successfully used to identify biomarkers for diabetec renal disease by Mischak *et al* (112). They showed that CE-MS is a viable technique for inspection of polypeptides in urine samples, with respect to identifying potential biomarkers for disease. A 'normal' urinary polypeptide pattern was established and it was shown that for patients with diabetic renal disease, and healthy controls, specific patterns of polypeptides could be seen. This allows diagnosis of diabetic renal disease from urine samples, and also the possibility of distinguishing between different types of renal disease with a high specificity.

Wittke *et al* (113) have also demonstrated that it is possible to use CE-MS to distinguish between patients who have received a successful kidney transplant, and those whose bodies are rejecting a transplanted kidney. Diagnosing these conditions by analysis of urine, rather than the more traditional diagnosis method of kidney biopsy, has the advantage of being less invasive. It also makes it possible to make earlier diagnosis, allowing prompt remedial treatment. In addition, Kolch *et al* (114) have also reported on polypeptide patterns that have been identified in other fluids such as plasma and cerebrospinal fluid.

Meier *et al* (115) have also used capillary electrophoresis coupled to an electrospray time-of-flight mass spectrometer (CE-ESI-TOF-MS) to identify patterns of polypeptides in urine. Patterns were identified for the recognition of diabetic nephropathy. They were able to detect, on average, 767 polypeptides in an individual sample. Overall 5076 different polypeptides were detected, in the mass range 800 Da – 66.5kDa. Each one was characterised by its mass and its CE retention time. This study showed that the urinary proteome contains a greater range of polypeptides than previously recognised. A 'normal polypeptide pattern' was also established, from healthy patients urine samples.

Ramstrom *et al* (116) have published work concerning analysis of human body fluids using LC and CE coupled to a high resolution mass spectrometry technique, known as Fourier Transform Ion Cyclotron Resonance (FT-ICR) mass spectrometry. LC and CE are easily interfaced with electrospray mass spectrometry (ESI), and this is a well-documented technique. FT-ICR was used because it is a high-resolution technique. Tryptic digests of

different body fluids were used including saliva, urine, plasma and cerebrospinal fluid. They found that the LC approach gives more information, and accepts a higher sample load, whereas the CE experiments are faster to run. It was also shown that it is possible to quantify the peptides seen in body fluid samples using labelling methods. It was concluded that neither LC nor CE is a 'better' separation technique, but they are complementary and the best results were obtained when these techniques were used in conjunction with one another. However, it must be noted that both technologies may not always be available to the researcher, and liquid chromatography remains the more commonly used technique, and additionally is the technique used throughout the work described in following chapters.

1.5 Quantitation using 2D Gel Electrophoresis

2D gel electrophoresis is a widely used and powerful technique for analysis of protein abundance (117), and has for many years been thought of as the 'gold standard' for protein separation. 2D electrophoresis separates proteins by their isoelectric points and in the second dimension by their size and charge. Unfortunately 2D gel electrophoresis is expensive, time consuming and labour intensive. It requires many gels to be run, as only one sample can be analysed on each gel, and comparisons made. Gels must be stained and scanned in order to visualise the protein sample. There is also a certain lack of reproducibility between gels. This can cause problems with distinguishing between variation in the gels or biological change between samples.

1.5.1 Differential Gel Electrophoresis

Fluorescence detection has a high sensitivity and a wide dynamic range, and is an ideal detection method for quantification of proteins and peptides in small biological samples. Fluorescence detection has previously been widely applied to 2 dimensional polyacrylamide gel electrophoresis (2D-PAGE) in proteomics (117). When fluorescence detection is applied in this way it results in a technique known as differential gel electrophoresis (DIGE). Fluorescent dyes (Cy dyes DIGE fluorophores) are used to label different protein sample sets, and an internal standard is labelled in the same way. These samples are then mixed and run at the same time, on the same 2D gel. This allows spot matching and minimises gel-to-gel variation. The fluorescent dyes are size and charge matched, and are spectrally resolvable. The dyes modify the lysine residues in a peptide or protein, via an amide linkage. The same protein labelled with the different Cy dyes will migrate to nearly the same point on the 2D gel. The gel can then be scanned at different

laser wavelengths, to separately visualise the differentially labelled protein samples. As the dyes are linear, sensitive and have a wide dynamic range this makes the DIGE technique quantitative. The sensitivity of the Cy dyes can be down to 0.1ng of protein (118), meaning that this technique is more sensitive than silver staining (which has a sensitivity of 1ng(119)). DIGE has the advantage over traditional 2D gels that the gel-to-gel variability is eliminated and an internal standard can be run on each gel, along with the samples.

There are currently 2 different classes of CyDye labels available (117), minimal labelling dyes and saturation labelling dyes. There are 3 minimal labelling dyes, enabling analysis of three samples on the same gel, although typically the third dye is used for a standard allowing two samples to be analysed on each gel. The minimal labelling dyes are capable of detecting differences between protein abundance when 50 μg of total protein, or 1 μg of individual protein, is present. There are only two saturation-labelling dyes, but they are capable of detecting differences on gels when only 5 μg of total protein, or 100 pg of individual protein is present.

In minimal labelling the ratio of dye to protein is kept low, so that only around 3-5% of the total protein in a sample is labelled. This is so that protein molecules visualized on a gel will be those labelled with mainly single dye molecules. Minimal labelling dyes label the lysine residues in a protein, via an amide linkage. This approach is possible due to the fact that most proteins have a relatively high lysine content. If all lysine residues in a protein sample were to be labelled, this would result in a large consumption of expensive dye. This could also lead to insolubility problems. One significant problem with minimal dye labelling is that the small proportion of labelled protein molecules will migrate to a slightly different point on the 2D gel to the unlabelled protein (120). This can cause problems if spot picking is required for post-gel mass spectrometric analysis, as the majority of the protein will not be at the points indicated by the dyes, meaning that additional post-electrophoretic staining are required to locate the majority of the spot. As low amounts of protein loading is required for DIGE a sensitive stain, such as SYPRO Ruby is required. This adds extra expense, time and experimental complication (121).

Saturation labelling is used to label all available cysteine residues in a protein. This means that a high dye-to-protein ratio must be used. Proteins generally have a relatively low cysteine content, meaning that it is reasonably efficient to label all cysteines that are present. The saturation CyDyes (see figures 1.3, 1.4) contain a maleimide reactive group which forms a covalent bond with the thiol group present in cysteine residues, via a thioether linkage. The saturation dyes have a neutral charge and are matched in molecular

weight, so all proteins labelled with different dyes will migrate to the same spot on a gel. The saturation dyes are normally used when only a very small sample is available.

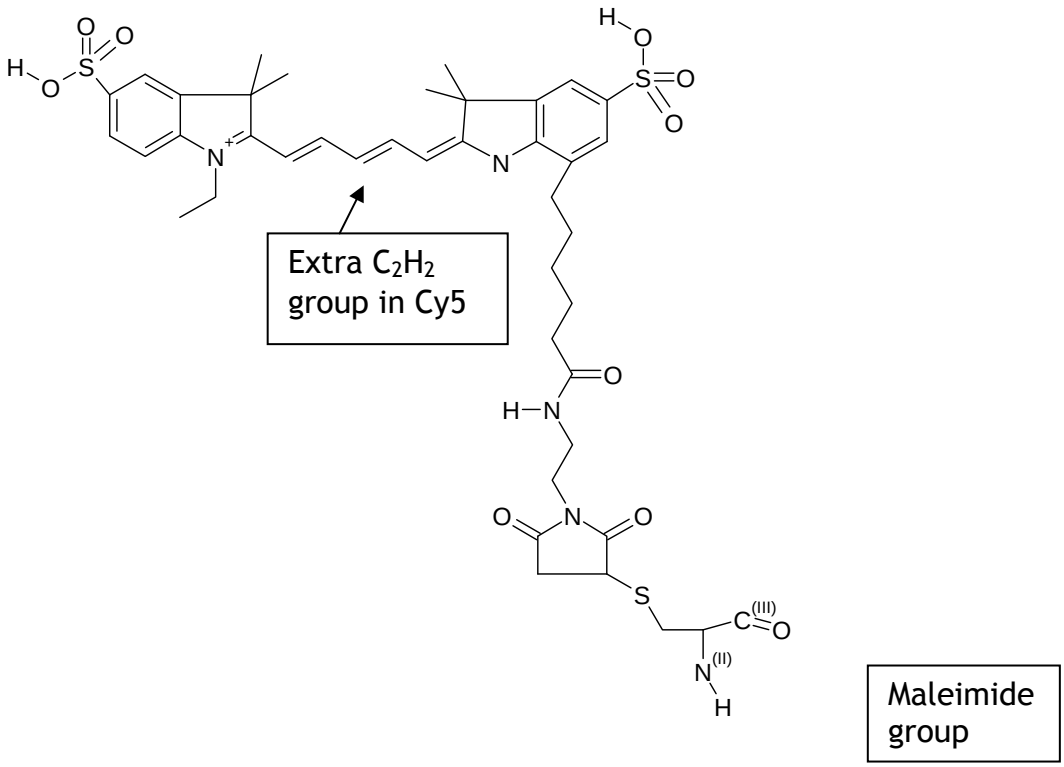


Figure 1.3 Cy5 Saturation dye. Structures obtained from Amersham.com

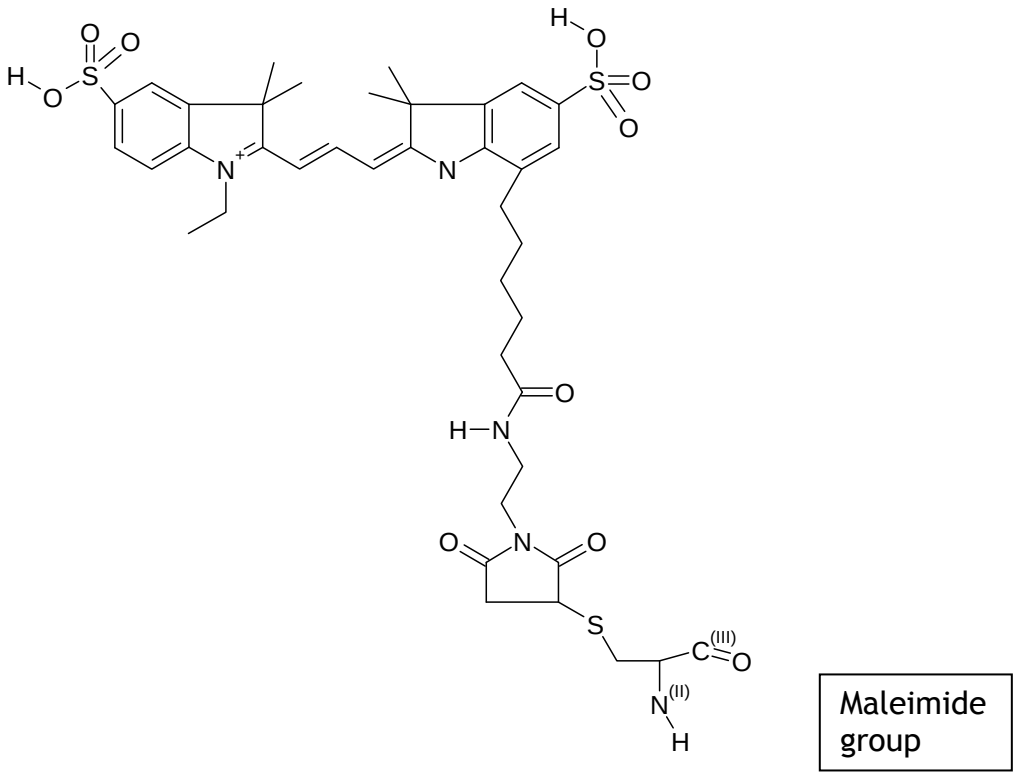


Figure 1.4 Cy3 Saturation dye. Structures obtained from Amersham.com

2D gel electrophoresis and DIGE still have limitations. Proteins that are present at a low relative abundance in a sample may not be visualised, and pre-fractionation and enrichment technologies may still be needed. Large and hydrophobic proteins may also not enter the gel. DIGE also has the problem that if a protein contains a large number of lysine residues it may be labelled more efficiently compared to a protein will less lysine residues. Another point to consider is that if a protein contains no cysteine residues saturation labelling will not be possible.

1.5.2 Alexa Molecule Internal Standard (ALIS)

An alternative method to DIGE for quantitation on gel has been developed (121), using an internal standard method. It is as sensitive as the DIGE minimal labelling method, but use of post-electrophoretic stains as apposed to pre-labelling removes the problem of the labelled proteins migrating differentially. This method works on the principle that if all proteins are absorbed into the IPG strip with equal efficiency, then the incorporation of an internal protein standard that has been labelled with a fluorescent Alexa molecule (ALIS) can provide a way of quantifying the experimental variability. The sample is spiked with the standard prior to 2D gel electrophoresis separation. After the separation the gel is stained with a fluorescent protein stain, which is sufficiently spectrally separate from the ALIS. The ALIS and the sample protein can be visualised and quantified independently. The spot volumes of the total protein can then be normalised with the ALIS output. This means that variability between gels can be taken into account in the quantification. As the internal standard proteins are labelled there will be a shift on the gel in comparison to the unlabelled proteins, but this does not influence the normalisation, as it is not a direct ratiometric normalisation that is used, as in DIGE, but a global ratiometric normalisation method. This is when a median of a selection of spot volumes is used for normalisation. Fluorescent stains that can be used with this method include SYPRO Ruby, RuTBS, Sulfo-rhodamine G and Deep Purple. One of the main advantages of ALIS to DIGE is that it has a much lower cost, as less expensive dye is required.

1.6 Quantitation with CE-LIF

Fluorescence detection has been used routinely for quantitation in DIGE, however fluorescence detection has not been so widely used with separation technologies such as high performance liquid chromatography (HPLC) and capillary electrophoresis (CE). It is possible to label proteins and peptides with a fluorescent dye, in the same way as proteins

are labelled for DIGE (117). These labelled proteins can then be separated, detected using laser induced fluorescence (LIF) detection and identified with mass spectrometry. LIF detection is more straightforward to implement than mass spectrometry as a detection method, but is more complicated than a UV absorbance detection system, as a derivatisation step is usually needed. However, LIF detection has the advantage over UV absorbance that it is much more sensitive (109). A LIF detector along with a CE (or possibly an LC) separation can be used to detect very low amounts of fluorescent analytes, even possibly down to the single molecule level (122-125).

The theory of LIF is relatively simple (126). Light from a laser of specific wavelength is shone onto a sample, containing a known fluorophore. The laser light excites the fluorophore to a higher energy level. It then relaxes, emitting fluorescent light, of a longer wavelength. This light can then be measured, using a photomultiplier tube, or other detector.

CE-LIF is already used as a powerful and sensitive qualitative tool for separation and detection of proteins and peptides (127), and a CE-LIF system can be relatively straight forward to build, see figure 1.5. Biological molecules such as amino acids and proteins can display native fluorescence. For peptides and proteins the native fluorescence of tyrosine and tryptophan residues can be used. However this native fluorescence is often low, and an expensive laser may be required in order to excite it in the UV region of the spectrum. This means that labelling will usually be necessary. Some of the common fluorescent tags used for labelling proteins and peptides are succinimidyl esters (128), isothiocyanates (129) and sulphonyl chlorides (130). Dyes such as fluorescein isothiocyante are commonly used as they have a relatively low cost, and can be easily excited with a common Argon laser. In the analysis of labelled molecules the detection of these molecules is likely to limited by the chemistry of the labelling reaction (such as incomplete labelling), and not by the LIF detection system (127).

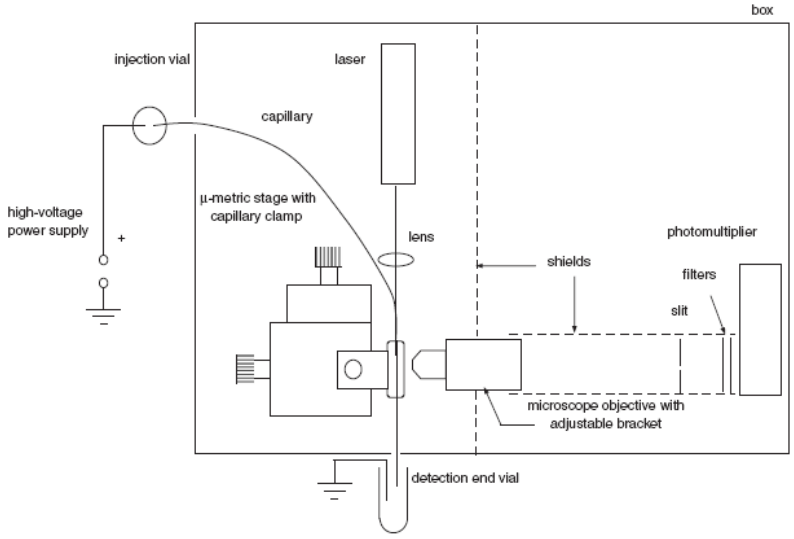


Figure 1.5 Schematic showing home built CE-LIF system (127)

One of the major limitations of LIF detection is that the laser wavelengths required may not be available, or may be very expensive, restricting their widespread use. Continuous wave lasers were traditionally used for CE-LIF, but alternatives have been found such as semiconductor lasers. A further alternative for excitation is an optical fibre and a blue light emitting diode (LED) (131). The LED can be used as an excitation source, and the optical fibre is used to transmit the light to the sample detection window. LEDs can be purchased with increased power and with a large range of emission wavelengths. They are of a small size and straightforward to use. The LED excitation system also has the major advantage of being much cheaper and simpler than a laser excitation source. LEDs are also convenient and attractive in regards to miniaturisation of the detection system. However, unfortunately all the light coming from an LED cannot be used for excitation. This is due to reflectance and scattering increasing background noise and reducing the efficiency of the LED excitation. Use of an optical fibre helps to reduce reflectance and light scattering, and therefore make the system more flexible, and easier to use.

CE-LIF has been shown to have many biological applications, for instance it has been demonstrated (132) that CE-LIF can be used for quantitative nuclear and cytoplasmic localisation of antisense oligonucleotides, by labelling with a fluorescent molecule (fluorescein). CE-LIF has also been used for detection and semi-quantitative determination of GFAP (glial fibrillary acidic protein) mRNA (133), in mouse brain. It was shown that CE-LIF was a sensitive technique, but that care must be taken to remove any possible interfering substances prior to analysis.

1.7 Stable Isotopic labelling for Quantitation

Widely used approaches for quantitative proteomics may involve protein separation by 2D gel electrophoresis (134), as mentioned previously in section 1.5. These techniques are time consuming, labour intensive and expensive. Also, as again mentioned previously, gel analysis is not useful for large, hydrophobic or extremely basic or acidic proteins. The gel techniques are also biased towards the most abundant proteins present, meaning lower abundant proteins are often not identified. This has encouraged the development of gel free and mass spectrometric based proteomic technologies, in order to obtain more information and decrease analysis times.

Stable isotope labelling has become a popular method for quantitation (135). This involves proteins or peptides, in two, or more, different samples, being differentially labelled using stable isotope tags. The tags will produce specific, and known, mass shifts in the mass spectra for the labelled peptides. These mass shifts in the spectra can then be used as internal standards for relative quantitation. Different methods for stable isotope labelling have been developed, and some of these will be reviewed here.

1.7.1 Isotope Coded Affinity Tags

One of the commonly used stable isotope labelling methods is based on isotope coded affinity tags (ICAT) (136). These consist of three parts, a specific chemical reactivity, an isotopically coded linker and an affinity tag, see figure 1.6:

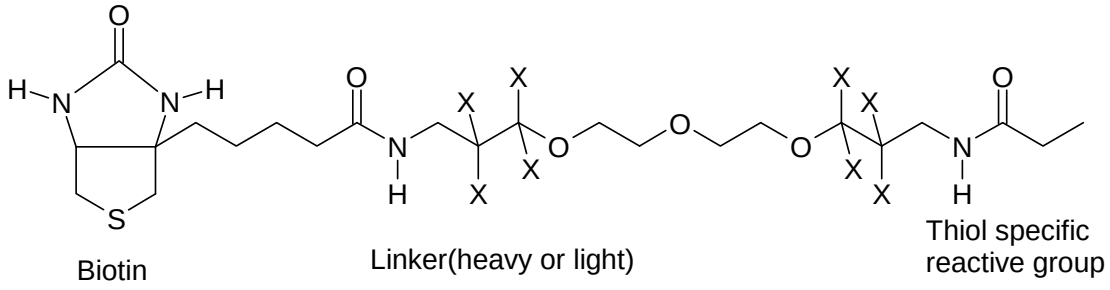


Figure 1.6 ICAT reagent (137)

In the light form ‘X’ is hydrogen and in the heavy form ‘X’ is deuterium. The biotin is used to enrich for peptides labelled with ICAT reagent and the thiol specific reactive group is used to attach the tag to cysteine residues.

Two sets of samples are derivatised, one with the heavy reagent and one with the light reagent. The samples are then combined and cleaved with an enzyme to generate peptides, some of which will be tagged. The tagged peptides (these will be any peptides containing cysteine residues) are then isolated using an avidin affinity chromatography column. The isolated peptides are then separated using HPLC and analysed by tandem mass spectrometry. The quantity and the sequence information of the proteins can then be determined using MS-MS. To do this the mass spectrometer operates in a dual mode. It changes between measuring the relative quantities of peptides eluting and recording the sequence information. The peptides are quantified by measuring the relative peak areas for the pairs of identical peptides that are tagged with the light or heavy forms of the ICAT reagent. These differ in mass-to-charge ratio by the difference in mass between the light and heavy reagents. Sequence information from the tagged peptides is generated by selecting ions of a particular mass-to-charge ratio, for fragmentation. These peptides are fragmented and the masses of the fragments can be compared with masses from sequence databases, in order to identify the protein that the tagged peptide is from.

One of the main principles behind the ICAT approach to quantitative proteomics is that a short sequence of amino acids from a particular protein (this can be between 5 and 25 amino acids) contains enough information to uniquely identify that protein. Another important principle is that pairs of peptides tagged with the light and heavy ICAT reagents are chemically identical. This is why they can be used as internal standards for quantification.

The ratio between the intensities of the heavier and lighter peaks can provide an accurate measurement of the relative abundance of the peptides in the original samples. This is because the MS intensity will be the same, regardless of the isotopic composition. Selective enrichment for peptides containing cysteine residues also reduces the complexity of the sample.

It has also been shown (137) that ICAT coupled with three dimensional chromatography (cation exchange, biotin affinity, reverse-phase) of an enzymatic digestion of the tagged protein sample is a useful technique for analysis of low abundance proteins in proteomic samples. This is very useful as low abundance proteins are notoriously difficult to analyse using 2D gel electrophoresis coupled to mass spectrometry, as it simply lacks the sensitivity required. It must be remembered that the labelling reaction is concentration dependent (138), and also that denaturing agents will affect the labelling procedure, for instance urea improves the labelling (seemingly by making the sulfhydryl groups more

accessible) and SDS compromises the reaction. Minimal reaction time should also be used in order to avoid any side reactions, and the reaction can be quenched when complete.

Despite the many advantages of ICAT there are certain disadvantages associated with this technique (139). Even though the mass difference from labelling with deuterium is small, there is still a shift in the mobility of the deuterium labelled peptides in size or mass dependent separation technologies. For instance in RP-HPLC the heavy peptide may migrate more rapidly than the light peptide, and can elute as a separate fraction. Therefore in order to accurately quantify the heavy and light peptide pair, both peptides must elute completely, allowing integration of the ion current for each peptide. This means that the sequencing of the peptides must be sacrificed in order to get accurate relative quantification. If the peptide pair do not co-elute, they do not act as true standards. This can reduce the accuracy of the quantification. For instance one peptide of the pair may elute with a different peptide that could suppress its ionisation.

Problems with ICAT can also occur due to the fact that electrospray ionisation mass spectrometry can produce peptides with different charge states. This means that the mass difference between peptide pairs will be altered depending on the charge state, for instance for doubly charged ions the mass difference will be halved, in comparison to singly charged ions, meaning software algorithms are needed to compensate for this.

Since the ICAT reagents tag the thiol group in cysteine residues, peptides that do not contain cysteine cannot be analysed. This can be an important issue for consideration, as cysteine is a relatively rare amino acid. It has been described how an SH group can be incorporated into peptides lacking cysteine, in order to make tagging possible (140). Since every tryptic peptide will contain an N-terminal amino group, a simple method for converting peptide amino groups into sulfhydryl groups was designed. This means that the same ICAT reagents can be used for peptides that lack or contain cysteine.

A second generation of ICAT reagents has been developed, known as cleavable ICAT reagents (cICAT) (141;142). The cICAT reagents consist of a protein reactive group, a linker region and a retrieval tag. Nine atoms of either ^{12}C or ^{13}C are incorporated into the linker part of the molecule, for differentiation. The samples are treated in the same way as the original ICAT reagents, but the biotin affinity tag can be cleaved off after enrichment but prior to mass spectrometric analysis. MS is then used to screen for paired isotopic peaks that are different in weight by 9 Da. As the biotin tag can complicate fragmentation patterns, cleaving it prior to analysis avoids this issue.

It is also possible to label an internal standard with a cICAT reagent (141), allowing absolute quantification to become possible. However, it must be noted that absolute peptide concentration must be extrapolated to absolute protein concentration, and this is reliant on factors such as digestion efficiency and sample handling. Therefore samples must be processed extremely carefully and reproducibly in order to achieve accurate absolute quantification. The peptide to be used for an internal standard must also be selected carefully.

A further ICAT technique that has been developed is the use of visible ICAT reagents (VICAT) (140). VICAT reagents contain a visible tag, which allows the position of the tag and therefore the tagged peptides to be monitored during separation. They also contain a photocleavable linker, meaning that the majority of the tag can be removed prior to mass spectrometric analysis, avoiding problems with fragmentation. They also contain an isotope tag, but instead of the traditional deuterium atoms they use ^{13}C and ^{15}N, avoiding any problems caused by non-comigration of paired peptides. The VICAT reagents are designed so that tagged peptides can be initially separated on a gel strip by isoelectric focussing. The visible tag allows the position of the tagged peptide to be determined, this region can then be excised and the peptides analysed by RP-HPLC-ESI-MS-MS.

1.7.2 Tandem Mass Tagging

A different approach to quantitative proteomics, that overcomes some of the problems associated with ICAT is tandem mass tagging (TMT) (139). These tags are designed so that the same peptides labelled with different tandem mass tags comigrate in all separations. The tandem mass tags are peptides made up of a tagging group joined to a sensitisation group and a mass normalisation group. Each tag in a pair has the same overall mass and atomic composition. A second-generation tag has also been developed that contains a fragmentation enhancement group, proline. Proline is used as it enhances cleavage of the amide bond on its N-terminal side. When used with a Q-TOF instrument the first generation tag, without the proline, only gave consistent fragmentation results when high collision energies (70V) were used. But, using high collision energies can fragment ions that provide sequence information into smaller fragments, meaning little sequence information can be obtained, without a second lower energy scan. Therefore the second generation tag which allows fragmentation at lower energies (35-40V) is beneficial, as sequence and quantification information can be obtained in one scan.

The tags are designed so that when they are analysed by collision induced dissociation the tag is released, producing an ion with a specific mass-to-charge ratio. Each tag can also have a specific functionality, which can be varied allowing different types of labelling. The tag can also contain an affinity ligand, like biotin, as in the ICAT reagents. It is also possible to create more than two tags meaning that more than two samples can be compared, or an internal standard can be incorporated into the analysis.

Pairs of TMT peptides have the same atomic composition, but are also identical in mass, unlike the ICAT tagged peptides. This means that they will exactly comigrate in separations, and can be used as extremely precise internal standards. This can give more accurate quantification, than in ICAT. Additionally the MS signal for the two sets of tagged peptides is not split into two peaks, giving a higher MS sensitivity. The fact that untagged peptides can be ignored in this technique can also help to improve data quality.

Another advantage of TMT is that the charge state of the peptide does not affect the appearance of the TMT fragment after fragmentation. Therefore scanning of the spectrum can be done without the need for adjustments to compensate for different charge states of every peptide. This can be a problem in ICAT, as mentioned previously.

1.7.3 Isobaric Tags for Relative and Absolute Quantification

ITRAQ (143) (isobaric tags for relative and absolute quantification) is another quantitative technique that is similar to TMT, and was developed in 2004 (144). ITRAQ involves a set of amine reactive isobaric tags, which derivatise peptides at the N-terminus and at lysine residues. This means they will label all peptides from a tryptic digest. In MS mode the peptides labelled with the tags cannot be differentiated, but upon fragmentation in MSMS mode signature ions are produced, in the same way as in TMT. These provide quantitative information when the peak areas are integrated. There are four different ITRAQ reagents (145), that give four unique reporter ions when fragmented in MS-MS mode.

1.7.4 Stable Isotope Labelling with Amino Acids in Cell Culture

Another widely used approach to quantitative proteomics is the stable isotope labelling with amino acids in cell culture (SILAC) (134;146-149), which was first described in 2002 (147;150). In this simple light or heavy amino acids are incorporated into proteins in vivo. The different cell cultures are grown in separate light or heavy cell media. The heavy media can contain ^{2}H in place of ^{1}H, ^{13}C in place of ^{12}C or ^{15}N in place of ^{14}N. The isolated

proteins are the combined and digested. The samples can then be fractionated and analysed by mass spectrometry. As in ICAT the labelled peptides can be identified by the mass difference of a known amount, and the peak intensities, or areas, determine their relative quantities. The sample complexity is not reduced as in ICAT, but as all peptides are labelled the sequence coverage is greatly increased. Again the accuracy of quantitation depends on the abundance and signal-to-noise ratio of the peptide pair. SILAC allows mixing of labelled and unlabelled cells, meaning that as much care does not have to be taken to ensure the two sample sets are fractionated and purified in the same way, as in ICAT. However one of the disadvantages of SILAC is that the cell culture system must be able to tolerate the isotope substituted media, which can cause problems. A further disadvantage is that due to complete labelling being required for reliable quantification the method can prove rather expensive.

A further version of SILAC that uses light, medium and heavy media can be used to compare three different samples (147;151). This involved mixing cell lysates that were derived from cells labelled with three different isotopic forms of arginine. $^{12}C_6{}^{14}N_4$-arginine was used as the light version, $^{13}C_6{}^{14}N_4$-arginine was the medium version, and $^{13}C_6{}^{15}N_4$-arginine was the heavy version. The reproducibility of this method was determined and it was found that alterations on protein abundance could be reliably measured reproducibly, even when the alterations were only 1.5 or 2 fold changes.

1.7.5 ^{18}O Labelling

A different form of stable isotope labelling that is performed subsequent to digestion is ^{18}O labelling for relative quantification (152;153). As this method is performed post-digest it circumvents the possible issues of growing cells in certain media. One part of the digested sample can be analysed using MALDI-TOF-MS for fast identification of proteins, while the remainder of the sample can be labelled with ^{18}O for relative quantification. This means that the set of samples to be compared for relative quantification does not have to be decided until after the protein identification stage. Samples can be separated on a gel and then tryptically digested, followed by trypsin catalysed ^{18}O labelling. The labelling procedure can also be performed during digestion if the whole sample is to be analysed, rather than an aliquot used for identification. The labelled and control samples can then be directly mixed on a MALDI(152) plate for analysis, or analysed by LC-ESI-MS (153).

The labelling method uses the premise that trypsin accepts the products of protein cleavage reactions as substrates and so catalyses the exchange of hydroxyl groups from ^{18}O water

into the C-terminus of peptides until equilibrium is reached. Two atoms of ^{18}O are incorporated into the C-termini of the digested peptides, resulting in a mass shift of 4 Da for the labelled peptide fragments.

^{18}O labelling is very sensitive, and can be used to quantify proteins in the low femtomole range, so could be a useful technique when samples are limited. In theory, with this method, all tryptically-digested peptides are available for quantification, increasing the reliability in the measured values. A disadvantage of this technique is that the labelled and unlabelled peptides have only a 4 Da difference, meaning a high resolution mass spectrometer may be needed to distinguish the labelled products.

1.7.6 Isotope Coded N-terminal Sulphonation

Isotope coded N-terminal sulphonation (ICenS) of peptides is a procedure recently developed (154) which has a major advantage of improved peptide identification, facilitates de novo sequencing, as well as allowing quantitative analysis. ^{13}C-labelled 4-sulphophenyl isothiocyanate (^{13}C-SPITC) and unlabelled 4-sulphophenyl isothiocyanate were synthesized, see figures 1.7 and 1.8 below:

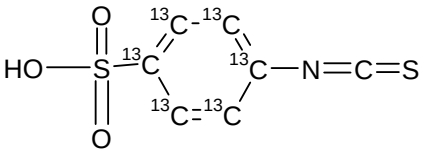


Figure 1.7 *^{13}C-SPITC*

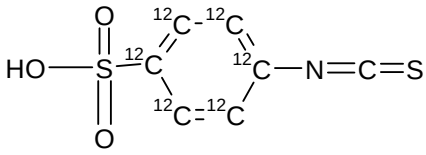


Figure 1.8 *^{12}C-SPITC*

The two tags were used to label two sets of peptides independently, these were then combined giving a mixture of peptides where each peptide isotope pair is separated by 6 Da. RPLC-MS-MS is then used to analyse the mixture. Due to the nature of ^{13}C labelling the two sets of peptides have nearly identical retention times. As the sulphonation reaction is inherently incomplete, the MS spectra can be used to distinguish between sulphonated and unmodified peptides. The MS-MS spectra can be used to differentiate between N-terminal fragments and C-terminal fragments by comparing the fragmentations of isotopic pairs. This results in improved pepdtide identification and increased confidence in de novo sequencing. Quantitation is performed in the same ways as mentioned previously.

1.8 Label Free Relative Quantitation

In recent years quantitative proteomics without labels has become an emerging area (155), as an alternative to stable isotope labelling. The label free technique is appealing as it is simple and relatively cost effective, although purchase of software can be expensive. Label free quantitation is based on comparing identical proteolytically digested peptides, which elute with the same retention time using LC-MS-MS analysis, from different samples. The relative ratios of the protein can then be calculated, providing a comprehensive quantitative overview of many protein concentrations between different samples. This relies on the observation that the peak intensity, or area, is, in most cases, proportional to the concentration of the peptide in the sample (156-159). Systematic errors are minimised by normalising peak intensities across the entire analysis (160).

Commonly the mass spectral peak intensities of the peptide ions are used for normalisation (161). The peak intensities have been shown to have a good parallel with protein abundances in complex samples. Spectral counting is an alternative method for the normalisation calculation (162). Spectral count is defined as the sum of all MS-MS signals for any peptide, including spectra that take the ion charge state into account. Spectral counting has been shown to have good reproducibility between replicate experiments, and good linearity in spectral count versus protein abundance. Spectral counting allows the opportunity to detect, identify and quantify more proteins compared to peak alignment and integration, because in spectral counting peptides in common, between datasets, are not required for the ratio calculations. However larger errors in protein ratios are seen in spectral counting than in peak area intensity measurements (160), meaning that the two methods are complementary to each other. It should also be taken into account that spectral counts will be more accurate for proteins with large numbers of spectra, and ratios from peak area intensities will be more accurate for proteins with more overlapping peptide ions. Protein ratios significant down to around 2.5 fold can be measured with high confidence using both methods. This is a lower sensitivity than that achieved by isotopic labelling, but label free methods will still be useful when labelling is difficult. It has also been shown (163) that linearity and reproducibility of label free protein quantitation methods can be improved by removing some of the most abundant proteins from complex samples.

It is very important to have a highly reproducible nano-flow HPLC separation system, as retention times must be reproducible if they are to be compared between samples. An additional important factor to consider for label free relative quantitation is the sample

processing. It is vital that the sample processing does not introduce false positive or mis-leading quantitative results. Preferably, a high resolution mass spectrometer should also be used, to prevent drifts in mass-to-charge ratio. Regular mass spectrometer calibration can also help to prevent this. Additionally, software tools must be developed to monitor changes in protein abundance within complex samples, and some commercial software is available (for example DeCyder MS). There are several data processing steps that must be performed, including peak detection, peak integration, deconvolution, chromatographic alignment of elution profiles, normalisation and statistical analysis (164). Different types of statistical analysis are discussed in more detail in following chapters.

A typical workflow for a label free quantitation experiment is as follows. Protein extracts from different samples are digested, normally using trypsin, giving complex peptide mixtures. An aliquot of this, for each sample, is analysed with the LC-MS system for the initial data acquisition. Typically, for this initial acquisition, only MS data is collected, as a high frequency of MS spectra is required to obtain the required chromatographic resolution for peak detection and integration. Following the label free analysis and identification of features of interest, the sample aliquot is re-analysed and MS-MS data is selectively acquired in a targeted manner. This can cause problems with relating the two datasets, but has the advantage that only the features of interest are targeted for identification, due to their detected quantitative differences.

One of the main advantages of label free quantitation is that there is no limit to the number of samples analysed, whereas stable isotope labelling methods will always be limited by the number of labels available to the researcher (165). There has already been work published, where label free quantitation has been used for biomarker discovery in different fields. These include breast cancer (166;167), Gaucher disease (168), proteinurea (169), schizophrenia (170) and diabetes (171).

Ru *et al* (166) developed a label free quantitation method for analysis of breast cancer samples. They investigated 120 human sera samples with 49 from invasive breast carcinoma patients, 26 from non-invasive breast carcinoma patients, 35 from benign breast disease patients and 10 from normal controls. The samples were analysed using a 2D-LC-MS method. The quantitation method used was a semi-quantitative peptide profiling method which was based on comparisons of normalised relative ion intensities. The data was analysed using two techniques, known as hierarchical clustering analysis and principal components analysis. It was concluded that the method developed has the potential to be scaled up for a larger clinical study.

Patwardhan *et al* (167) also used a label free quantitation method to study breast cancer samples, however cell lines were the samples used here. The technique employed in this case is known as an accurate mass and time strategy. For this strategy Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR-MS) is required. This is a very accurate type of mass spectrometry (accurate to <1ppm), which is required due to the accurate mass and LC elution times being used as peptide identifiers. They combined this label free method with ^{18}O labelling in an attempt to obtain a broad quantitative method which can be used for a wide range of proteins. The results of this were that multiple proteins were identified that were differentially expressed among the cancer and non-cancer cell lines. However, it was also stated that further validation was required to ensure the significance of the identified differences.

Vissers *et al* (168) investigated label free quantitation as a method to discover novel biomarkers for Gaucher disease. This is a genetic disorder, which results in lipid accumulation in certain organs such as the spleen, kidneys, lungs and brain. This is caused by a deficiency of the enzyme glucocerebrosidase. Patients with Gaucher disease show a large elevation in serum of the protein chitotriosidase. Unfortunately there is a frequent deficiency in this protein, with one in every 20 Caucasians expressing no chitotriosidase. This means that there is the need for new biomarkers for Gaucher disease. In this work serum samples from diseased and control patients were analysed using an LC-MS method. Depletion of abundant proteins was also investigated in conjunction with label-free quantitation. The results of this work were that condition-unique LC-MS protein signatures were established; confirming that label free quantitation is a useful technique for biomarker discovery.

Kemperman *et al* (169) described a label free quantitation method for the analysis of urine samples with reference to proteinuria. The urine samples were separated using LC and analysed using electrospray Ion-Trap MS. Additionally, they assessed the lower limit of detection and the reproducibility of the method. Principal component analysis was used to classify the peaks seen from the analysis. 92 peaks were selected as discriminating the sample sets, of which 6 were more intense in the majority of the proteinuric samples. Two of these peaks were identified as albumin derived peptides. Other albumin-derived peptides were found to be non-discriminatory. The group proposed that this indicated preferential proteolysis at certain cleavage sites.

Huang *et al* (170) investgated label free quantitation as a biomarker discovery method for schizophrenia patients using cerebrospinal fluid samples. Only 20 samples were used in

this work, 10 from healthy controls and 10 from diseased patients. They found a clear difference between the sample sets. 77 proteins were identified, with seven of these being newly identified. However, further development is needed for utilisation of this technique as a biomarker discovery method.

As a final example, Metz *et al* (171) used untargeted label free quantitation in a pilot proteomic analysis of human plasma and serum from control and diabetic patients, with the goal of identifying biomarkers for type 1 diabetes. FT-ICR-MS was used and five candidate biomarkers were identified. However, it was stated that further validation was needed prior to moving forward with these markers.

In the work described in following chapters a label-free relative quantitation method is developed and applied to different datasets, for biomarker discovery for hypertension and also for the evaluation of a heat treatment system for prevention of proteomic sample degradation. Statistical tools are also applied to the datasets in a novel manner, to identify and target differences between sample sets.

1.9 Protein Fractionation Technologies

As previously mentioned, complexity of biological solutions can pose problems for proteomic analysis and biomarker discovery. It is possible to address the problem of large dynamic range of biological samples using protein fractionation prior to analysis. This involves isolation of the sample into distinguishable fractions. This sample treatment can take many forms, including precipitation, centrifugation, filtration abundant protein depletion and protein of interest enrichment. The quantity and quality of protein identification and quantitation are directly related to the complexity and reproducibility of the samples used, therefore the merits and types of protein fractionation must be considered carefully. The selection of the fractionation technique depends on the type of sample being analysed, the physicochemical properties of the proteins, their subcellular location and the objective of the study in question. There is no general fractionation protocol that is universally accepted and widely used (172).

A relatively simple technique for protein fractionation is sample precipitation. This can be performed using acetone, TCA, ethanol, diethyl ether, isopropanol, chloroform, methanol, ammonium sulphate or PEG (173;174). Ammonium sulphate is used to induce protein de-stabilisation, otherwise known as ‘salting out’. Organic solvents are used to increase attraction between particles of opposite charge in the sample, which leads to protein

precipitation. The sample is then redissolved in a smaller volume. Precipitation is not a selective fractionation method, and may not be reproducible between samples. It is, however, a quick and simple fractionation technique.

Samples can also be fractionated into their subcellular fractions (nucleus, mitochondria, Golgi apparatus, lysosomes, exosomes, peroxisomes and phagosomes), and is another straightforward method by which to simplify proteomic samples. There are two main steps, first the disruption of the cellular organisation and secondly the fractionation of the homogenate to allow separation of the different organelles. This is often performed using centrifugation (175). The cells are collected using a low speed centrifugation step and then mechanically homogenised. Following this, the nuclei are removed using low-speed centrifugation and can be purified from the cell debris and unbroken cells. The remaining supernatant contains the cytosol and other organelles in suspension. These can then be separated using differential gradient centrifugation (176). Different media can be used for the separation, which will affect the degree of separation attained and sucrose is the most commonly used medium. Free flow electrophoresis and immunoisolation have also been used for subcellular fractionation of organelles (177). For analysis of the total organelle proteome, purity of the isolated organelles is of importance, although cross contamination between fractions can make complete purification impossible. This can be monitored by the use of markers. Enrichment of certain subcellular fractions can then be used to detect proteins of low abundance, and also to track their change in abundance (178). This technique has been used to identify proteins from the Golgi location in rat liver (179).

A more specific method of protein fractionation is the use of beads with differently functionalised surfaces (180). This sample process is effective and versatile, and is a useful technique for discovery and identification of biomarkers or other proteins of interest. A commercial system using magnetic beads with functionalised surfaces known as ClinProt has been introduced by Bruker Daltonics (see figure 1.9 below). The body fluid sample is incubated with the magnetic beads (available with different functionalised surfaces), and washed by manipulating the beads with a magnetic strip. The proteins or peptides of interest can then be eluted and analysed using MALDI-TOF-MS. It has been shown that using the ClinProt system with Anchor Chip Targets from Bruker and human plasma samples can be used to generate reproducible protein/peptide profiles (181), which can be used for scanning for potential biomarkers for disease. Differences between diseased and normal samples can be seen, when statistical analysis software package (ClinProTools, provided by Bruker) is used to analyse the profiles.

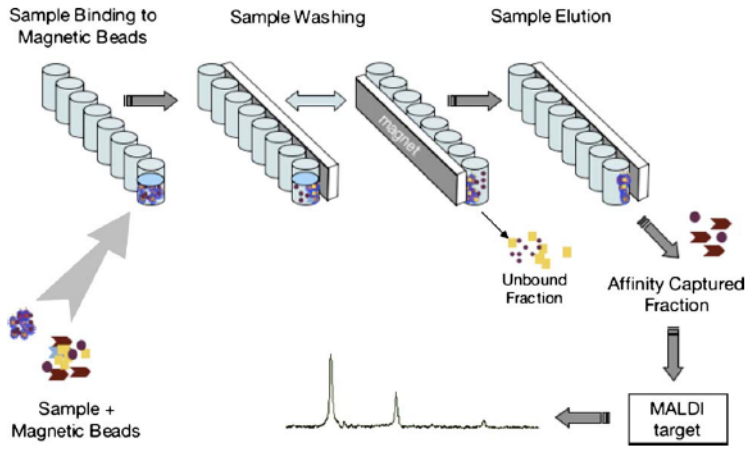


Figure 1.9 ClinProt Workflow(180). Protein samples are incubated with magnetic beads with functionalised surfaces. They are then washed to remove unbound material. (Magnetic strips are used to move the beads around). Bound proteins or peptides are then eluted and can be spotted onto a MALDI target plate for MS analysis

SELDI-TOF-MS is similar technique to the bead based capture technologies that has been used as a biomarker discovery platform. In SELDI-TOF-MS a planar array of different surface chemistries are used for simultaneous processing and sample fractionation (180) prior to analysis by TOF-MS. Proteins are isolated on the surfaces and can be used to reduce sample complexity in the same way as the ClinProt bead system. Unfortunately there is limited binding to the SELDI surfaces and extra noise can be obtained from the surfaces. A commercial SELDI system is available from Ciphergen, the ProteinChip Array System. This system allows for simple sample preparation, and lessens the need for expertise in mass spectrometry (182).

Different surface chemistries can be used including traditional chromatographic separation media, such as reverse phase or ion exchange, or biochemical surfaces, such as antibodies or enzymes (see figure 1.10).

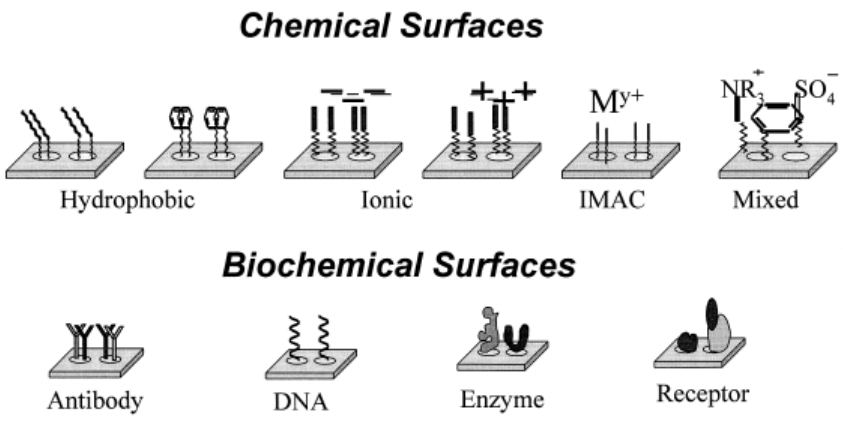


Figure 1.10 Possible SELDI surface chemistries (182)

The chemical surfaces have broad binding properties and are usually used in protein profiling experiments and for biomarker discovery, in a similar way to the ClinProt beads. The biochemical binding surfaces are much more specific, so provide high enrichment of particular captured analytes.

There are various methods and technologies used to deplete high abundance proteins and enrich for low abundance proteins. Centrifugation forces can be used to force a liquid against a semi-permeable membrane, meaning that low molecular weight solutes and the liquid will mass through the membrane and high molecular weight solutes and solids will be retained, dependent on the pore size of the membrane. These are available with different molecular weight cut-offs, and can be used with different centrifugation speeds (183;184). This is a simple method that can be used to remove high abundance proteins such as albumin from plasma samples, as their molecular weight is known. It must be considered, however, that low abundance features with a high molecular weight will also be removed from the sample using this method. Additionally, protein-protein interactions may need to be disrupted to disassociate low molecular weight proteins from albumin, and allow them to pass through the membrane (185).

Abundant proteins can also be removed from plasma using antibody affinity systems. The multiple affinity removal system (MARS) is used to remove the most abundant proteins from plasma samples. It has been shown (186) that using the MARS system, prior to mass spectrometry analysis gives a good, reproducible proteome coverage. The MARS column removes the six most abundant proteins from the plasma samples (albumin, transferrin, IgG, IgA, anti-trypsin and haptoglobin) helping to reduce interference from these, and allowing proteins of lower abundance to be detected and analysed. The advantages of this technique are that it is quick and easy, however it is an expensive technique and the columns have a limited capacity. There are also concerns regarding techniques such as this when performed under non-denaturing techniques, as low abundance proteins bound to the target proteins may be simultaneously removed from the sample, via the ‘albumin sponge effect’. It has been shown that about 210 proteins are associated with the six most abundant plasma proteins, and also the removal of albumin causes a significant loss of cytokines from the sample (187). This and similar techniques for removal of abundant proteins are reviewed in more detail in the introduction to the chapter ‘Discovery and Identification of Biomarkers for Hypertension’.

Another protein enrichment technique is known as N-linked glycopeptide capture (188), and can be used for body fluid samples such as plasma and serum. Protein glycosylation is a relatively common post-translational modification. Carbohydrates can be linked to serine or threonine residues (O-linked glycosylation) or to asparagine residues (N-linked glycosylation) (189). Many biomarkers and drug targets are glycoproteins, as N-linked glycosylation is often found in extracellular proteins (190), located in body fluids.

Peptides that are N-linked glycosylated in the native protein can be isolated (after an enzymatic digestion), reducing the complexity of the sample, as every serum protein contains (on average) only a few N-linked glycosylation sites. The capture of the glycosylated proteins is based on oxidation of hydroxyl groups on adjacent carbons of the carbohydrate to aldehydes, by sodium periodate. The aldehydes then covalently couple to amine or hydrazine containing molecules, immobilised on beads. The formerly N-linked glycosylated peptides can be released via a catalysed reaction with PNGase F, for mass spectrometry analysis. It was shown by Zhang *et al* (189) that this method is reproducible and gives more information about low abundance proteins in serum samples, compared to mass spectrometry analysis of untreated serum samples. It was also shown (188) that the peptide pattern of a mouse with cancer could be differentiated from a genetically identical mouse without cancer.

In conclusion, proteomics is an ever-growing field with many exciting and interesting applications. In the chapters that follow several of these are examined and developed, mainly concentrating on the theme of biomarker discovery for disease. The literature reviewed here concentrates on research relating to that in following chapters, and gives an interesting context and background in which to place this work.

The biological changes in stroke induced mouse brain tissue are studied as an introductory experimental chapter. This work was a useful preliminary to the following research and helped develop experimental knowledge, planning skills and highlighted the importance of careful sample preparation.

The prevention of degradation to tissue samples by a novel heat treatment method is studied. The issue of possible sample degradation was a logical point to address following the analysis of stroke induced mouse brain tissue.

Finally, the changes to plasma samples from hypertensive, wild type and a congenic strain of rat are studied. The focus of this chapter is the development of a label free relative

quantitation method for analysis of a rich data set. The application of novel statistical methods to the data was the main interest, and this is considered to be an under discussed and studied area in this field of research.

Within the chapters that follow the following aims are addressed:

- Optimisation of mouse brain tissue sample preparation for proteomic analysis and determination of protein concentration in samples

- Identification of features by LC-ESI-MS-MS analysis, previously discovered by MALDI-MSI analysis, followed by confirmation of identifications using traditional 1D gel and western blotting analysis

- Evaluation of the overall effect of heat treatment on the mouse brain tissue sub10 kDa proteome

- Development of a label free relative quantitation method in order to quantify the differences seen between heat treated samples and snap frozen samples in both digested and intact sub 10 kDa proteome

- Optimisation of sample preparation of plasma samples for LC-MS analysis and determination of protein concentration in the rat plasma samples used

- Evaluation of plasma partitioning spin columns using 2D mini gels and LC-MS analysis

- Evaluation of two types of label free quantitation software, followed by label free relative quantitation of features within plasma sample as a possible biomarker discovery method

- Development of a novel statistical analysis for the label free relative quantitation dataset

- Targeted MS-MS analysis for identification of any features of interest, followed by western blots as validation of any biomarkers discovered

In my research presented here, special attention is paid to sample preparation, the importance of experimental design and the application of relevant and useful statistical methods to enable the mining of useful information from rich datasets.

2 Biomarker Identification in Stroke Brain Guided by MALDI-imaging

2.1 Introduction

Ischaemic stroke is the second most common cause of death and also the most common cause of acquired disability in adults (191;192). Additionally in western countries stroke causes 10-12 % of all deaths, with 12% of those deaths in the under 65 age range (193). In 2002, stroke related disability was judged to be the sixth most common cause of increased disability adjusted life-years (DALYs). DALYs are the sum of life-years lost as a result of premature death and years lived with disability, adjusted for severity (191). Due to the increasing elderly population in western societies, it is estimated that by 2030 stroke related disability, in western societies, will be rated as the fourth most important cause of DALYs (194). Stroke is also very expensive to society, consuming 2-4 % of total healthcare costs, worldwide, and in industrialised countries accounting for over 4 % of direct healthcare costs. Costs to UK society were estimated at £7.6 billion, based on 1995 prices, and the costs to US society were estimated at $40.9 billion, based on 1997 prices. This relates to a large cost of $100 per head of US population per year (195). Regardless of these high costs to society, the proportion of research funds targeted towards stroke remains disproportionately low (196).

Risk factors for stroke can be classified as modifiable (such as hypertension, diabetes and smoking) or fixed (atrial fibrillation and transient ischaemic attack). Risk factors that have been identified account for only around 60 % of the attributable risk. In contrast about 90 % of ischaemic heart disease is explained by identifiable risk factors (197;198).

Strokes can be classified as being either ischaemic or haemorrhagic, and the management of these subtypes is quite different. Therefore, the clinical distinction between these types is one of the most important steps in stroke management. This distinction has been made much easier by the introduction of CT and MRI scanning (199). Haemorrhagic stroke is also known as intracerebral haemorrhage, and the most common cause is hypertensive small-vessel disease causing rupturing of small lipohyalinotic aneurysms (200). Haemorrhage into a previous infarction can be another contributing factor (201). Around two thirds of patients with primary cerebral haemorrhage have either pre-existing or recently diagnosed hypertension (202). Ischaemic strokes are more common, accounting for around 80 % of all strokes (203). The mechanism that causes vessel occlusion, and

therefore ischaemic stroke can be cardioembolic, artery to artery ambolism or in-situ small vessel disease. The identification of which mechanism is present is important as it can influence acute treatments and secondary prevention strategies (192), meaning that a biomarker based assay could have life changing implications. Following vessel occlusion, a volume of structurally intact but functionally impaired tissue surrounds the ischemic core (204). This is commonly known as the ischaemic penumbra. It is often the target for therapeutic treatments, as its recovery is associated with neurological improvement. In the ischaemic penumbra a cascade of neurochemical events starts with energy depletion. This is followed by disruption of ion homoeostasis, release of glutamate, calcium channel dysfunction, release of free radicals, membrane disruption, inflammatory changes and necrotic and apoptotic cell death triggering. The infarct core contains unsalvageable tissue, however in animal models the cascade can be arrested at various points, which can form the basis for neuroprotective therapies (205).

Stroke prognosis is not good, with around a quarter of stroke patients being dead within a month, a third within six months and a half within a year (206;207). Prognosis is worse for those with intracerebral and subarachnoid haemorrhage, with the one month mortality approaching 50 %. The main cause of early mortality is neurological deterioration; in addition contributions are made from infections. Later deaths are more usually caused by cardiac disease or further complications (207). The main predictors of stroke recovery at three months are initial neurological deficit and age, with other factors being high blood glucose concentrations, body temperature and history of previous stroke (208). Following the occurrence of minor strokes, the risk of further strokes is high, even as high as 30% in some subgroups (209-211). Patients at particular high risk of reoccurrence can be identified by their age, blood pressure and characteristics of their symptoms, including unilateral weakness and speech impairment (212). Imaging strategies, such as magnetic resonance imaging are also used to predict possible high risk of recurrence. The presence of diffusion weighted image lesions, or of occluded vessels can help to identify these patients (213).

Mortality rates for stroke are fortunately in decline, and the main reason for this is due to improved control of risk factors, including hypertension (214) and blood pressure reducing agents have been ever increasing in effectiveness since the 1950s (215). Other risk factors may also have had an effect, including reduction in smoking rates, cholesterol, diabetes and atrial fibrillation (216). It has been suggested that patients with hypertension, without previous stroke should be treated with warfarin if they have atrial fibrillation, lipids

reduction can be targeted with statins in patients with pre-existing ischaemic heart disease, and women over 45 can be treated with aspirin (217-221).

In the past 30 years there have been advances in the prevention of recurrent stroke. Aspirin was introduced as a prevention strategy in 1978 (222). Evidence has also been presented showing that oral administration of aspirin within 48 hours of onset of ischaemic stroke reduces 14 day morbidity and mortality (223;224). However, this benefit is relatively small, as only around 9 patients per 1000 treated are saved from death or disability, and after exclusion factors are taken into account, only 4 patients per 1000 treated are saved from death or disability (225). Aspirin is a straightforward treatment, as it low cost, easily administered and has low toxic effects (225). Aspirin is not only useful for secondary prevention, but penumbral salvage may also be possible (226). Further secondary prevention strategies include aspirin with dipyridamole (introduced in 1987) (227), warfarin for patients with atrial fibrillation (introduced in 1993) (228), carotid andarterectomy for symptomatic carotid-artery stenosis of greater than 70 % (introduced in 1991) (229;230), clopidogrel (introduced in 1996) (231), blood pressure reduction using perindopril and indapamide or ramipril (introduced in 2001) (232;233) and cholesterol reduction with atorvastatin (introduced in 2006) (234). Most patients will qualify for at least one of these secondary prevention strategies.

A treatment under recent investigation is carotid angioplasty with stenting. This is now generally combined with distal protection devices. This is a minimally invasive procedure, which may replace carotid endarterectomy as the treatment of choice in many patients. Initial trials have suggested that the surgical risks associated with this procedure are similar to those seen in carotid endarterectomy (235;236), but it has been suggested that these risks may be increased when the procedure is undertaken by less skilled hands (237).

Surprisingly, prevention of secondary stroke by modification of risk factors such as blood sugar reduction in diabetics (238), reduced alcohol consumption, increased exercise and smoking cessation is not substantiated with evidence from clinical trials. However, observational studies have shown many potential benefits, small costs and small risk of adverse effects, meaning that patients should be advised to stop smoking, drink in moderation, eat a balanced diet and to exercise (239-242).

With regards to treatment of stroke there have also been advances in recent years. One of the most effective advances has been the routine management of patients in Stroke Care Units (SCUs). This is appropriate for all stroke subtypes, reduces mortality by around 20

%, and also improves functional output by around 20 % (243). In addition, a space identified as an SCU sees improved outcomes when compared with a dedicated stroke team visiting patients on general wards (244). This is thought to be due to strict blood pressure control, early mobilisation and good adherence to best practice (245;246).

Thrombolytic treatment with recombinant tissue plasminogen activator (tPA) is a biologically effective treatment for acute ischaemic stroke, if used within a three hour window of stroke onset. However, it does not improve mortality rates, only disability rates (225;247), and is therefore not widely used as a treatment method. Decompressive surgery is another possible intervention method, but is suitable for only a very small proportion of patients. These are young patients with malignant middle-cerebral-artery-territory infarction and space-occupying brain oedema. This occurs in 1-10 % of patients with supratentorial hemispheric infarcts, and usually arises between 2-5 days after stroke (248).

There are also various intervention treatments currently under evaluation. Haematoma growth in the first few hours of onset of stroke has been shown to be a key factor for poor clinical outcome (249). Recombinant factor VII is normally given to patients with haemophilia, or to reduce haemorrhagic complications of surgical procedures. It has now been shown to attenuate haematoma growth, with additional secondary clinical benefits in a phase II trial (250). Another treatment that is under evaluation is thrombectomy devices. These are mechanical devices which have been developed to enable removal of a blood clot from major vessels, including middle-cerebral and basilar arteries. One of these is the mechanical embolus removal in cerebral ischemia (MERCI) retrieval catheter (Concentric Medical, Mountain View, CA, USA). This device has a corkscrew shape. It has been shown, in a phase II trial, that this device could be used to remove the blood clot, with a complication rate similar to that seen when the treatment used is intravenous tPA (251). This device has now been FDA approved for clot removal, but not as a stroke therapy, which has proved to be a fairly controversial decision (252;253).

In summary, stroke management and treatment has improved over the past 10 years, but still poses interesting challenges for clinicians and researchers. In my research stroke brain tissue is investigated, using a mouse model, with the aim to identify markers that are associated with stroke. It is proposed that the discovery of biomarkers for stroke could aid diagnosis and influence early treatment methods, as well as providing further information with regards to tissue recovery.

An imaging technique has previously been used to identify features that differ between stroke brain and wild type brain (unpublished work by Dr Richard Goodwin, IBLS, Glasgow University). Features are categorised by mass only, and it can be a challenge to identify these features. In this work, an effort has been made to use downstream proteomic techniques to help improve identification of these features. The imaging technique used is known as MALDI-MSI (matrix assisted laser desorption ionisation mass spectrometry imaging). This is a technique that is used to investigate the distribution of molecular features across thin (μm) slices of tissue, using a MALDI mass spectrometer. It was pioneered by Caprioli, Chaurand and Stoeckli (254-257), and has been used to detect different types of biomolecules including peptides, proteins, lipids and small molecules (258-260). Recently, MALDI-MSI has shown potential for detection of disease progression and biomarkers (261-268).

MALDI-MSI uses conventional MALDI-MS technology to collect a mass spectrum over a series of points across a thin section of tissue. These spectra can then be used to plot the relative intensity of each m/z peak in the mass spectrum across the tissue area. Therefore, the distribution of each feature can be visualised across the tissue, using a pixilated intensity map for each feature identified. Assuming the sample preparation and mass spectra acquisition is reproducible between tissue samples, these intensity maps can then be compared between related samples, in this case between a stroke induced mouse brain and a normal, wild type mouse brain. In this way markers for disease can be identified.

One of the major challenges in MALDI-MSI is the MS based identification of the features of interest. Currently, performing MALDI MS-MS analysis directly from the tissue has been a difficult process (260). In this work, it is proposed that the mass of interest discovered using MALDI-imaging can be used to identify the feature using a different method of identification, in this case LC-MS-MS analysis. The mass of interest is targeted using the conventional proteomic technology. It is suggested that these techniques will prove to be complimentary. There are certain disadvantages to this method, which must be taken into account. Certain assumptions must be made regarding the mass shown in the MALDI-MSI data, for example it must be assumed that the mass imaged directly from the tissue relates exactly to the mass targeted using the LC-MS-MS analysis. One of the most important steps to take to assure this assumption is true, is to ensure both mass spectrometers are calibrated accurately and often. Additionally experiment validation must be undertaken. It is proposed here that validation can be performed using quantitative western blotting and the traditional biochemical technique of immunohistochemistry.

In this work, a stroke-induced mouse model is used, but in the future these techniques could have the potential to be applicable to different tissue types, the diagnosis and monitoring of many different disease states, and also different types of complex biological samples. Excision of regions of interest can also be used in conjunction with these techniques to give useful information from specific areas and also to help avoid low abundance changes being missed.

2.2 Aims

There are several important aims in this chapter, which will be discussed in more detail in the following sections.

- Optimisation of mouse brain tissue sample preparation for proteomic analysis
- Determination of the protein concentration in tissue samples
- Identification of features by LC-ESI-MS-MS analysis, previously discovered by MALDI-MSI analysis
- Confirmation of identifications using traditional 1D gel and western blotting analysis

2.3 Methods

2.3.1 Sample Preparation Protocol

Male ICR mice (6-8 weeks) brain tissue samples were used in this work, and these were obtained from Dr Hilary Carswell from the Institute of Pharmacy and Biomedical Sciences, University of Strathclyde. Global ischaemia was induced in the mouse brain. Both common carotid arteries were isolated and then occluded using microaneurysm clips applied bilaterally for 15 minutes. The clips were removed, and blood flow through the arteries was confirmed prior to suture of the wound. The body temperature was controlled at 37 ºC during the surgical procedure, and then maintained by an incubator until the animals had fully recovered from aneathesia. Wild type, healthy mouse brains were also used as a control. 6 stroke induced brains and 10 wild types brains were used in this work.

Mice were euthanatized by cervical dislocation in accordance with the U.K. Animals (Scientific Procedures) Act, 1986 and local ethical guidelines. Brains were removed and sliced using a cryostat microtome (Leica Microsystems CM 1900UV, UK) maintained at minus 15 ºC.

Two of the stroke induced brains were sliced to a thickness of 10 µm, two were sliced to a thickness of 15 µm and two were sliced to a thickness of 20 µm. The wild type brains were sliced in the same way, with the remaining four wild type brains all sliced to a thickness of 15 µm. A protein extraction from brain tissue protocol was optimised for LC-MS-MS analysis. The samples were cooled with ice during handling. The protocol consisted of covering the brain slice with 50µl of a solution of 1% formic acid in distilled H_2O, in a sample tube (Eppendorf) and then centrifuging for 10 seconds at 20,000 g. The sample was then vortexed vigorously prior to probe sonication for 5 seconds, on ice. The sample was then vortexed vigorously and then centrifuged for 30 minutes at 20,000 g. The supernatant was removed, and size filtered using Microcon Nominal Molecular Weight Filters, 10,000 Da, (Millipore, YM10) according to manufacturers protocol. The sub 10 kDa fraction filtrate was then removed and stored at minus 80 ºC prior to analysis.

2.3.2 Protein Concentration Determination

Concentration of tissue sample extractions was determined using a standard protein concentration assay. This was performed on extracted tissue samples prior to size

fractionation, and also following size fractionation, for 10 µm slices of wild type brain tissue and also for 20 µm slices of wild type brain tissue.

10 µl of tissue extraction solution was used for protein concentration determination. BSA standards were used with protein concentrations of 0.0625, 0.125, 0.25, 0.375, 0.5 and 0.75 mg/ml. The absorbtivity measurement of each standard and sample was repeated three times, and the mean signal was calculated. A standard curve was created from the absorbtivity and concentration of the known standards. The concentration of tissue extraction solution was then determined using this curve.

2.3.3 LC-MS-MS for Protein Identification

Extracted protein samples were analysed using a Dionex (4 Albany Court, Camberley, Surrey) Ultimate nano-flow LC system and an Applied Biosystems (120 Birchwood Boulevard, Warrington) Q-Star Pulsar i. A 90 minute gradient (5% solvent A to 80% solvent B) was used. The LC separation was performed on a 15cm C18 PepMap reverse phase column with a 75 µm internal diameter, (Dionex). Solvent A was a 2 % acetonitrile, 0.1 % formic acid solution and solvent B was an 80 % acetonitrile, 0.1 % formic acid. The loading solution used was a 2 % acetonitrile, 0.5 % tri-fluoroacetic acid solution. The loading flow, onto the trap was maintained at 30 µl/min. The column flow was maintained at 0.3 µl/min. This LC system does not have a column oven; therefore the column was maintained at room temperature throughout the separation.

MS and MS-MS data was collected in positive ion mode, using the following Q-Star settings:

Accumulation time: 3 seconds, TOF masses: m/z 200 – 1500, Product ion masses: m/z 50 – 2000, dynamic exclusion was used, and rolling collision energy. Electrospray was performed using ‘picotip’ needles (New Objective) and the standard Q-Star electrospray source.

Q-star ‘wif’ files were extracted using Analyst QS MASCOT plug in (v1.1) and searched using MASCOT version 2.1 (Matric Science, London, UK) against database NCBInr 20070413 (4848751 sequences; 1676925347 residues), using MASCOT search settings as follows:

Species: Rodentia, Enzyme: No enzyme, Report top hits: Auto, Variable modifications: Oxidation of Methionine, Charge: 1+, 2+ and 3+, Peptide tolerance: 1.8 Da, MS/MS tolerance: 0.8 Da, Instrument: ESI-TOF.

From the MASCOT results, a list of detected, and identified, features was constructed. The equivalent 1+ m/z was calculated and this was cross referenced with the MALDI-MSI data. The MALDI-MSI data used, in this instance, was a list of features that demonstrated differing intensity distributions between stroke induced and wild type brain. All MALDI-MSI work was performed by Dr Richard Goodwin (IBLS, Glasgow University) using a Bruker Ultraflex MALDI-TOF mass spectrometer. Any features that occurred in the LC-MS-MS data set were noted.

A two dimensional MudPIT analysis was also performed on the brain tissue, size filtered samples. This process is described in more detail in the introduction section 1.3. For the MudPIT analysis the size filtered protein extraction from tissue samples, prepared as described in section 2.3.1, were dried down for 30 minutes to remove solvent and subsequently resuspended in 50 µl of 25 mM ammonium bicarbonate and digested using sequencing grade modified porcine trypsin.

The mass spectrometer settings used were as previous. A 90 minute gradient was used for each salt step in the MudPIT analysis. Sodium chloride solutions were used at the following, increasing concentrations:

0mM, 25mM, 50mM, 100mM, 150mM, 200mM, 250mM, 300mM, 350mM, 400mM, 1M

2.3.4 1D gels and Western Blotting for Confirmation of Identification

Following identification of features that possibly related back to MALDI-MSI data, 1D gels and western blots were performed as a validation method. Samples used were a 10 µm brain slice, extracted as previously described in section 2.3.1, and size filtered <10 kDa, a 20 µm brain slice, extracted as previously and size filtered <10 kDa, a 10 µm brain slice, extracted as previously without size filtration, and two 20 µm brain slices, extracted as previously described without size filtration. From each sample 4 µl was used, and diluted with 16 µl of protein loading buffer before being denatured at 100 ºC for 5 minutes. Protein loading buffer consisted of 1.6ml Glycerol, 1.6ml 10% SDS, 0.5ml 1M Tris (pH 6.8), 3.9ml water, 0.4ml β-mercaptoethanol, and 0.1g of bromophenol blue. This was

diluted 1:5 prior to use. Each sample was then split into two 10 µl aliquots. Each sample was then loaded onto a 12 % pre-cast polyacrylamide gel (Invitrogen) and separated for 1 hour at 200 V, using a pre-made MOPS running buffer (Invitrogen). A protein molecular weight marker was loaded in the first lane. This was repeated, in an identical manner for the 2 aliqouts of each sample, on two separate gels. One gel was used for silver staining for protein visualisation and the second gel was used for western blotting.

Western blots were performed for several proteins:

Primary antibody: Anti-myelin basic protein (mouse – developed in rat), Secondary antibody: Anti-rat IgG horse radish peroxidase conjugate

Primary antibody: Anti-MHC Class II antigen (mouse – developed in rat), Secondary antibody: Anti-rat IgG horse radish peroxidase conjugate

Primary antibody: Anti-ATP synthase (mouse – developed in chicken), Secondary antibody: Anti-chicken IgY horse radish peroxidase conjugate

Western blots were performed as follows. The gels were immersed in transfer buffer for 15 minutes at room temperature. The gel size was measured, and filter papers and western blot membrane cut to the same size (1 piece of membrane per gel, and 6 pieces of filter paper per gel). The filter paper was soaked in transfer buffer for 5 minutes. The membrane was wet in methanol for 15 seconds, soaked in water for 2 minutes, and then in transfer buffer for 5 minutes. The transfer stack was assembled in the SemiPhor western blot apparatus. This consisted of 3 pieces of filter paper, the western blot membrane, gel and 3 further pieces of filter paper. 27 mA of current was applied, per gel, for 1 hour. The membranes were then removed and left to air dry overnight. The gels were discarded at this stage. The membranes were wet in methanol and then immersed in blocking solution for 1 hour. Blocking solution consisted of 100 ml PBS with 5 g of dried milk powder. The membranes were then removed and washed in PBS with 0.05 % Tween. The membranes were then each immersed in the primary antibody wash for 1 hour, consisting of 20 ml PBS with 0.05 % Tween, 5 % milk and 10 µl primary antibody. The membranes were then washed 6 times, for 5 minutes each time, in PBS with 0.05 % Tween, and immersed in secondary antibody solution for 1 hour. Secondary antibody solution consisted of of 20 ml PBS with 0.05 % Tween, 5 % milk and 10 µl secondary antibody. The six 5 minute washes were repeated. Each membrane was then covered in Pierce ECL reagent for 60 seconds. Membranes were dabbed dry and exposed onto UV film (Kodak) and developed using an

Xomat developer. Five 30 second exposures were used initially, followed by five 2 minute exposures.

The 1D gel and western blot for antibody anti-ATP synthase was repeated using laser microdissected tissue samples. These samples consisted of 2 wild type samples and three stroke samples. All LMD work was performed by Alastair Michael Lang (IBLS, Glasgow University). Laser microdissection is a method used to isolate small regions of tissure sample and is discussed in more detail in the introduction section 1.1.

The tissue slices were placed under a microscope and the stroke region of the brain identified visually. A laser was then used to outline and exise the region of interest. This was then captured in a sample tube. Hence, LMD selects the tiny tissue sample by area alone and protein concentration is not considered.

In this work the stroke regions were microdissected using LMD apparatus, areas of 900 x 560 µm. Identical sized regions were dissected from the wild type brains, from the corresponding brain region. Protein extraction was performed as described previously without size filtration. The 1D gel was also performed as described previously. The 1D gel was silver stained for visualisation of the protein bands. The western blot was repeated in an identical way to the previously described blot.

An additional western blot was performed, with a protein loading standard, to normalise between protein content of samples. The loading standard used was an antibody for anti-beta tubulin III (developed in rabbit), with a secondary antibody of anti-rabbit IgG horse radish peroxidase conjugate. These antibodies were simply mixed with the original antibodies (as described previously) and the blot performed in the same way.

2.4 Results

2.4.1 Protein Extraction

Several differing protein extraction methods were considered in this research, prior to finalising the method as described in section 2.3.1.

Probe sonication time was considered. Sonication times of 2, 5, 10, 20 and 30 seconds were trialled. It was found that for sonication times of over 5 seconds that there was a large amount of heating of the sample. Even when performed on ice, the sample tube became hot to the touch. It was concluded that this would be detrimental to the samples and could cause protein degradation. When a sonication time of 2 seconds was used, it could be seen that the tissue sample remained in a pellet at the bottom of the sample tube. Therefore, it was concluded that an optimal probe sonication time of 5 seconds would be used in the final method. This was performed on ice, in order to preserve the sample integrity and to keep sample degradation to an absolute minimum.

Additionally, the time of centrifugation following sonication was considered. This was undertaken by initially centrifuging samples for 10, 20 and 30 minutes (all at 20,000 g). The supernatant was examined by eye following each centrifugation time. It was observed after the 10 and 20 minute centrifugation that the pellet was not fully formed and that there were tissue particles present in the supernatant. After 30 minutes of centrifugation the pellet was fully formed and the supernatant was clear. Therefore, it was concluded that a 30 minute centrifugation was necessary for this sample protocol.

2.4.2 Protein Concentration Determination

The protein concentration was determined (as described in section 2.3.2) in the samples extracted from the mouse brain tissue using a standard Bradford assay. It was found that a 10 μm brain slice extraction contained 37 (±2) μg of protein and a 20 μm brain slice extraction contained 69 (±4) μg of protein. All brain slices were from mid-brain regions. See appendix 1 for the standard curve produced for determination of protein concentration.

Additionally, it was found that the protein content in the size filtered <10kDa extraction samples was too low to measure using this method. This is as expected, due to the majority of protein in tissue being larger than 10 kDa in size.

2.4.3 LC-MS-MS for Protein Identification

Two brain slices from each stroke induced mouse brain were extracted for LC-MS analysis as described in section 2.3.1. In addition two brain slices from six of the wild type mouse brains were extracted for LC-MS analysis. This resulted in four 10 μm slices, four 15 μm slices and four 20 μm slices from stroke induced and wild type mouse brains.

All LC-MS-MS data was searched using MASCOT version 2.1 against database NCBInr 20070413 (4848751 sequences; 1676925347 residues), as described in section 2.3.3. Masses of proteins identified were compared with masses of interest identified using MALDI-MSI, from the stroke brain regions.

A total of 67 proteins were identified. These identifications are shown in appendix 1. Three features were identified which matched masses from the MALDI-MSI analysis. These features were considered to be possible markers for stroke; that is features that show up or down regulation in stroke brain when compared with wild type brain.

These features were identified by the LC-MS-MS analysis as being the following proteins: ATP synthase, MHC Class II Antigen and Myelin Basic Protein. The MASCOT results and are shown in figures 2.1 – 2.3. It can be seen that the MASCOT scores for these identifications range from 68 to 185. The mass of each identified protein can be seen in the top line of each figure. It should be noted that the scores for the peptides seen in ATP synthase and Myelin Basic Protein are higher than those seen in MHC Class II Antigen, meaning the identifications are more statistically likely for the peptides with higher scores. In addition, the differences between the expected and identified peptides for MHC Class II Antigen are rather high.

gi|258788 **Mass:** 4964 **Score:** 79 **Queries matched:** 2

H(+)-ATP synthase subunit e (N-terminal) [rats, liver, Peptide Mitochondrial Partial, 44 aa]

☐ Check to include this hit in error tolerant search or archive report

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
938	596.3511	1786.0314	1786.0141	0.0173	0	44	7.2	1	-.VPPVQVSPLIKFGRYS.A
957	620.0221	1857.0446	1857.0512	-0.0066	0	69	0.019	1	-.VPPVQVSPLIKFGRYSA.L

Proteins matching the same set of peptides:

gi|17978459 **Mass:** 8249 **Score:** 79 **Queries matched:** 2

ATP synthase, H+ transporting, mitochondrial F0 complex, subunit E [Rattus norvegicus]

gi|83715998 **Mass:** 8230 **Score:** 79 **Queries matched:** 2

ATP synthase, H+ transporting, mitochondrial F1F0 complex, subunit e [Mus musculus]

gi|59808056 **Mass:** 8231 **Score:** 79 **Queries matched:** 2

ATP synthase, H+ transporting, mitochondrial F1F0 complex, subunit e [Mus musculus]

gi|26373885 **Mass:** 4225 **Score:** 79 **Queries matched:** 2

unnamed protein product [Mus musculus]

Figure 2.1 MASCOT results from mouse stroke brain, extracted and analysed using Dionex Ultimate nano-flow LC system and Applied Biosystems Q-Star in MS-MS mode. ATP synthase subunit E is identified.

gi|18958188 **Mass:** 6937 **Score:** 68 **Queries matched:** 8

MHC class II antigen [Rattus rattus]

☐ Check to include this hit in error tolerant search or archive report

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑ 348	444.1340	886.2535	887.4613	-1.2078	1	9	2.3e+02	1	R.HRAAVATY.-
408	457.1263	912.2381	913.4770	-1.2389	1	(8)	2.5e+02	10	R.HRAAVPTY.-
☑ 413	458.1445	914.2744	915.4926	-1.2182	1	20	19	1	R.HRAAVVTY.-
414	458.1445	914.2744	913.4770	0.7974	1	18	31	3	R.HRAAVPTY.-
☑ 417	459.1352	916.2558	915.4926	0.7631	1	(12)	1.2e+02	1	R.HRAAVVTY.-
☑ 420	460.1269	918.2393	919.4334	-1.1941	1	17	35	1	R.HRAAVCTY.-
462	474.1433	946.2720	945.4668	0.8052	1	9	2.5e+02	2	R.HRAAVETY.-
540	502.1587	1002.3029	1002.5035	-0.2006	1	7	3.4e+02	10	R.HRAAVWTY.-

Figure 2.2 MASCOT results from mouse stroke brain, extracted and analysed using Dionex Ultimate nano-flow LC system and Applied Biosystems Q-Star in MS-MS mode. MHC Class II antigen is identified.

gi|69885073 **Mass:** 14202 **Score:** 185 **Queries matched:** 8

myelin basic protein isoform 6 [Mus musculus]

☐ Check to include this hit in error tolerant search or archive report

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
157	478.6675	955.3204	955.4399	-0.1195	0	33	66	2	Q.DENPVVHF.F
☑ 235	539.7158	1077.4171	1077.5818	-0.1647	0	58	0.23	1	D.TGILDSIGRF.F
☑ 318	411.1549	1230.4429	1230.6033	-0.1604	0	34	61	1	Q.DENPVVHFFK.N
338	628.9924	1255.9703	1257.6982	-1.7278	0	21	1.3e+03	5	V.HFFKNIVTPR.T
☑ 412	449.1633	1344.4680	1344.6462	-0.1782	0	22	8.9e+02	1	Q.DENPVVHFFKN.I
482	497.5246	1489.5520	1488.6779	0.8740	0	13	8.7e+03	9	K.YLATASTMDHARH.G + Oxidation (M)
☑ 677	478.6967	1910.7575	1911.0002	-0.2427	0	31	1.3e+02	1	Q.DENPVVHFFKNIVTPR.T
709	500.7145	1998.8288	1999.0751	-0.2463	0	20	1.5e+03	3	G.FLPRHRDTGILDSIGRF.F

Figure 2.3 MASCOT results from mouse stroke brain, extracted and analysed using Dionex Ultimate nano-flow LC system and Applied Biosystems Q-Star in MS-MS mode. Myelin Basic Protein is identified.

The MALDI-MSI image of mouse stroke brain is shown in figure 2.4. As mentioned previously, this analysis was undertaken by Dr Richard Goodwin. This image shows a 'heat map' of feature with mass 4964 Da.

The images are successive sections taken from a stroke induced mouse brain. The stroke region occurs in the bottom part of the brain image. The heat map indicates decreased concentration of the 4964 Da feature in this region. The 4964 Da feature is identified in the LC-MS-MS analysis as ATP-Synthase subunit E. However, it must be considered that all spatial information is lost in the LC-MS-MS analysis procedure. This means that further validation steps are needed to confirm this identification, simply matching the two masses together is not sufficient to conclude this investigation.

In addition, in the LC-MS-MS data there was no signal seen for a 4964 Da protein. The peptides identified by LC-MS-MS are fragments of the ATP-Synthase subunit E, making it difficult to come to the conclusion that ATP-Synthase subunit E is also identified in the MALDI-MSI data.

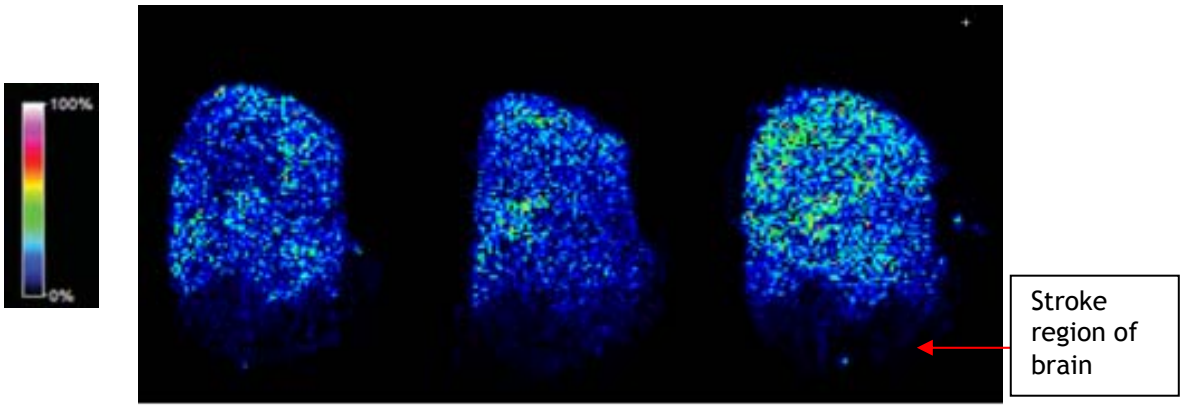


Figure 2.4 MALDI-MSI image of feature 4964 Da, in mouse stroke brain, possibly relating to ATP-Synthase subunit E. Three successive sections of the brain are shown, with the stroke region occurring in the bottom part of the image. This figure is a heat map based demonstration of the concentration of this mass, with the colour scale shown to the left. It can be seen that there is a decreased concentration of the feature in the stroke regions of all three images of the brain.

The LC-MS-MS identifications reported above were all made using the simpler one dimensional LC separation, as described at the start of section 2.3.3.

It should be noted here that hese samples have not been digested in the traditional proteomic way using trypsin, therefore the proteins and protein fragments within the sample are not 'tryptic'.

In the digestion process the enzyme trypsin cleaves the proteins at the carboxyl side of the amino acids arginine and lysine, resulting in tryptic peptides which have terminal groups of either arginine or lysine, the structures of these are shown in figures 2.5 and 2.6.

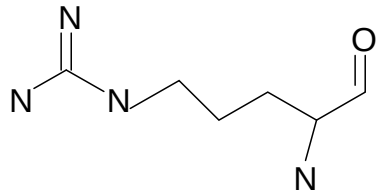


Figure 2.5 Structure of amino acid arginine

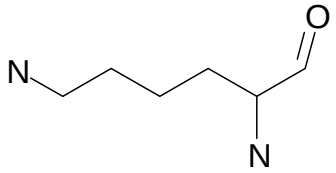


Figure 2.6 Structure of amino acid lysine

The fragments produced in MS-MS fragmentation are dependent on many factors. This includes the primary sequence of the peptide, the amount of internal energy, the charge state of the ion and the way the energy is introduced into the ion. Additionally, the fragments will only be detected if they have at least one charge. The fragmentation method used here is collision induced dissociation (CID), which is discussed in the introduction section 1.3. This involves collision of the peptide with an inert gas to induce fragmentation. This usually occurs preferentially at the peptide bond to produce y and b ion fragments as shown in figure 2.7. This nomenclature was introduced by Roepstorff and Fohlman in 1984 (269). The sequence of these ions then allows identification of the peptide.

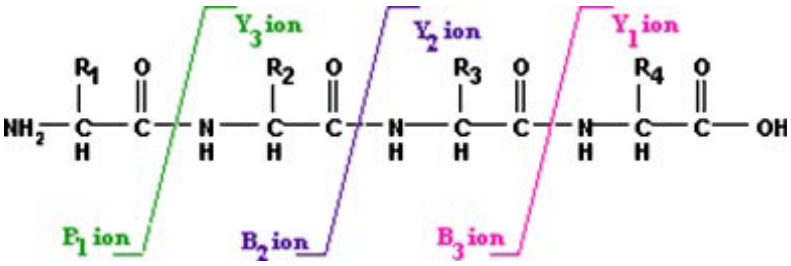


Figure 2.7 Cleavage of peptide bond to produce b and y ions, from Barton and Whittaker 2008 (270)

Tryptic peptides are particularly well suited to CID MS-MS fragmentation, as used here, and are usually identified fairly easily. It is proposed here that the non-tryptic peptides produced from the intact samples do not fragment well in this way, and are hence difficult to identify in this manner.

Additionally, a two-dimensional separation was performed, described in the latter part of section 2.3.3. This two-dimensional separation had a tryptic digestion step included in the

protocol. The problem with using a digestion step, in this type of work, is that the features identified by the following analysis will be likely to relate to fragments of those seen in the MALDI-MSI, rather than intact features. This means that it is very challenging to relate these two types of datasets with one another. Additionally, the two-dimensional analysis is very time consuming, taking 24 hours for one separation, in comparison with 90 minutes for a standard one dimensional separation. However, the advantages of using the two-dimensional analysis is that low abundance peptides may be identified, which cannot be seen in a conventional 1-dimensional LC-MS-MS analysis.

The two-dimensional analysis did result in further protein identifications, however, it was decided that the results were not suitable for this research due to the digestion step in the protocol and the challenges this brings with relating the LC-MS-MS data with the MALDI-MSI data. For this reason these results will not be considered any further in this research.

2.4.4 Gels and Western Blots as Confirmation of Identification

A silver stained 1D gel of laser microdissected tissue (900 x 560 μm in area) from mouse stroke brain and wild type brain is shown in figure 2.8 below. All samples were taken from 10 μm thick tissue slices. As previously mentioned all LMD was performed by Alastair Michael Lang.

Lane 1 shows a protein size molecular marker. Lanes 2 and 3 are from wild type mouse brain, and lanes 4, 5 and 6 are taken from the stroke region from stroke induced mouse brain. The tissue from the wild type brain was taken from the corresponding brain regions to the stroke regions from the stroke brains.

It can be seen that there are many different bands, as would be expected of a complex tissue sample. It is also observed that there are certain bands which are stronger in the wild type brain lanes.

Unfortunately, this gel shows some streaking and overloading. In particular the protein marker lane is particularly overloaded. The image of the gel is also not ideal. Towards the validation stages of this work the tissue sample amounts available for experimentation became extremely limited. Ideally, this gel would have been repeated to obtain a higher quality image; however, this was not possible due to sample limitations.

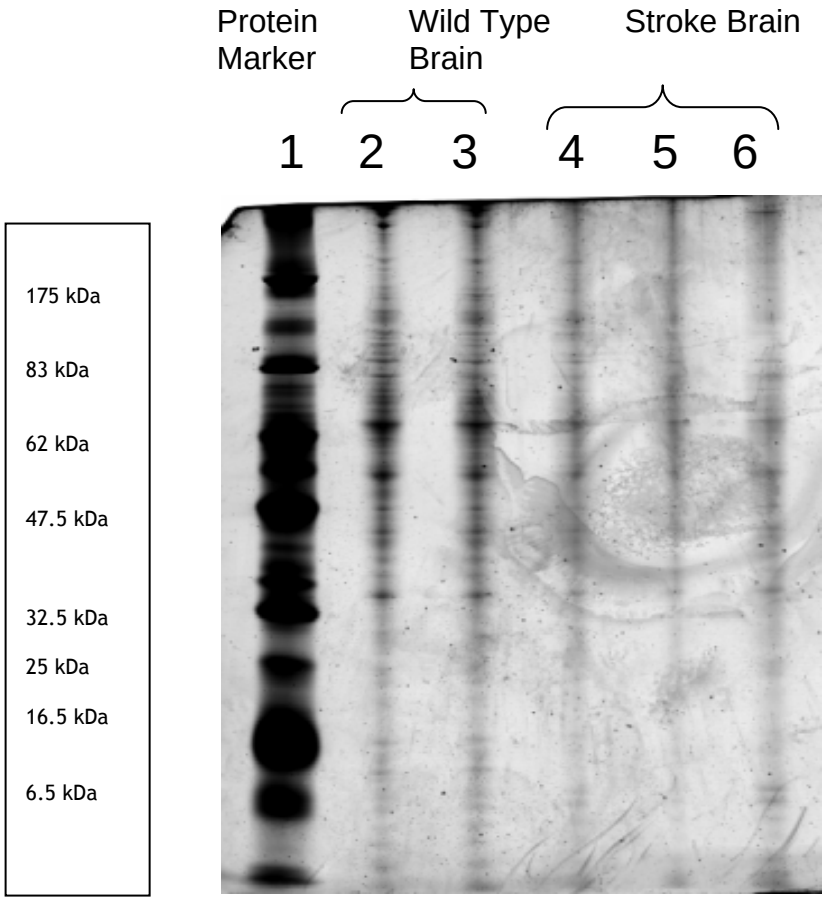


Figure 2.8 1D gel of mouse brain samples, laser microdissected. Lanes 2 and 3 are from wild type healthy brain. Lanes 4, 5 and 6 are stroke region from stroke induced mouse brain. Lane 1 is a protein size molecular marker. Laser microdissected area is 900 x 560 µm. Tissue slices are 10 µm thick. The gel was visualised using silver staining as discussed in section 2.3.4.

Western blots were performed using three antibodies; ATP synthase, MHC Class II antigen and Myelin basic protein, as these were the proteins identified in the LC-MS-MS analysis that corresponded to features from MALDI-MSI data. All antibodies used were polyclonal and hence will recognise multiple epitopes on any one antigen. The advantages of using polyclonal antibodies are that they are more likely to be able to recognise denatured proteins and generally provide more robust detection. In contrast monoclonal antibodies recognise only one epitope and will therefore give more specific detection than polyclonal antibodies. Polyclonal antibodies were used here to attempt to obtain a robust, first detection of the proteins in question.

No clear bands were seen for blots of MHC Class II antigen and Myelin basic protein, therefore this data is not presented here. This was thought to be due to poor performance of antibodies, which is a well known problem with this type of validation technique. It was

proposed that these western blots be repeated with alternative antibodies; however this was not possible due to lack of sample availability. The initial western blot for ATP synthase, using entire brain slices, was very over-exposed, even with a short 30 second exposure time and this is not presented here. The repeated western blot for ATP synthase, with laser microdissected brain regions is shown in figure 2.9 below. Brain tissue was laser microdissected from stroke region of three brain slices, of an area 900 x 560 x 10 µm. This was repeated for two wild type brain slices, of an identical size from the same brain region. It can be seen that the bands for the wild type brain are larger and more intense than the bands for the stroke brain region. This is as predicted by the data obtained from the MALDI-MSI.

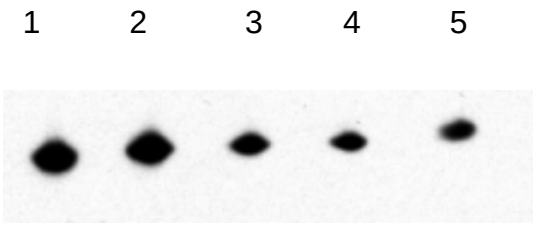


Figure 2.9 Western blot of laser microdissected stroke induced and wild type mouse brains, for ATP synthase. Lanes 1 and 2 are from wild type healthy brain. Lanes 3, 4 and 5 are stroke region from stroke induced mouse brain. Laser microdissected area was 900 x 560 µm. Tissue slices were 10 µm thick. The bands from the wild type mouse brain are more intense than those from the stroke induced brains, as predicted by MALDI-MSI.

However, if the corresponding 1D gel image is examined, it can be seen that the wild type brain lanes are more intense than those observed in the stroke brain lanes. It is proposed that this may be due to oedema in the stroke brain, causing swelling of the tissue. This would result in a smaller amount of tissue being sampled in the laser microdissection of an equivalent area of stroke brain, as the tissue is sampled by area only and not protein concentration. To investigate this, the western blot was repeated using a tubulin loading standard. The loading standard will allow confirmation of whether equal protein concentrations are being sampled. The results of this are shown in figure 2.10.

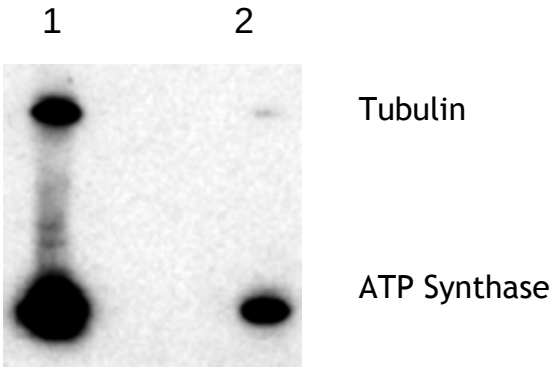


Figure 2.10 Western blot of laser microdissected stroke induced and wild type mouse brains, for ATP synthase, with a loading standard of Tubulin. Lane 1 is from wild type healthy brain. Lane 2 is stroke region from stroke induced mouse brain. Laser microdissected area is 900 x 560 µm. Tissue slices are 10 µm thick. The bands from the wild type mouse brain are more intense than those from the stroke induced brains, for both ATP synthase and for the loading standard, confirming the theory that oedema induced swelling has caused a larger amount of tissue to be sampled from the wild type brain with comparison to the stroke brain.

The western blot with loading standard concurs with the aforementioned theory of oedema induced swelling causing the same area of tissue, but differing overall protein concentration to be sampled by laser microdissection.

It was proposed that western blots be repeated and normalisation performed against loading standards. It was also proposed that immunohistochemistry could be used as an additional validation technique. Immunohistochemistry would give a direct comparison with MALDI-MSI results and would be an ideal validation technique. However, unfortunately due to a lack of sample availability this work could not be completed.

2.5 Discussion

The work in this chapter aimed to develop a biomarker identification method, in conjunction with, and guided by MALDI-MSI, a previously developed technique. MALDI-MSI can be used to investigate spatial differences of bio-molecules between diseased and healthy tissue samples and identify regions that change and markers that define these regions; however it often proves to be a challenge to identify the peptide and protein defining differences using MALDI-MSI alone. In this work, a conventional proteomic technique LC-MS (described in section 2.3.3) is used in an attempt to identify these previously defined features, in this example from mouse stroke induced brain. Previous work conducted by Dr Richard Goodwin, found features of interest, but these could not be identified.

LC-MS techniques, including 1D and 2D liquid chromatography were used in an attempt to identify the features in question from brain tissue. It was concluded that the 2D analysis was not suitable for this type of research, due to the digestion step introducing global changes to the sample. This means that the data from the 2D LC-MS cannot be easily correlated with the MALDI-MSI data.

The 1D LC-MS data was searched using MASCOT against NCBInr database, and the masses of identified proteins were compared with features of interest previously identified in the MALDI-MSI work. 1D gels and western blotting were used to validate the identifications made. However, due to lack of sample availability it was not possible to complete the validation. It was, however, discovered that samples of different protein concentration were unwittingly sampled from the stroke and wild type brain tissue, as discussed in section 2.4.4. It was theorized that the stroke induced brain tissue may have been swollen due to oedema, causing the difference in sampling by laser microdissection, leading to potentially mis-leading results in the western blots. Any future western blots from stroke brain tissue will be normalised using a loading standard in order to help to avoid this. It was also proposed that immunohistochemistry could be used as a useful validation technique, giving a direct comparison to the MALDI-MSI data.

As this work progressed it became apparent that certain assumptions must be made when attempting to relate two datasets of different types. One of the main advantages of MALDI-MSI is that spatial information is not lost, allowing features to be investigated in their original location. This can allow investigation into different areas of tissue, and is very applicable to stroke induced tissue, as the stroke region can be easily visualised. The

ideal way to identify these features would be direct MS-MS fragmentation of the feature in question from the tissue. In this way it can be certain that the correct feature is being targeted for identification. Unfortunately this is a difficult technique to perfect, and often direct fragmentation from tissue simply is not possible. When proteins are extracted into solution, all spatial information is lost, meaning that the only way to target specific features is by mass alone. This has clear disadvantages. Tissue is a complex sample and there are likely to be more than one feature with the same mass present in a tissue sample. Enzymatic degradation can also cause changes in the features present in samples, and it is therefore entirely possible that a feature identified with the same mass by MALDI-MSI and LC-MS-MS analysis may in fact be an entirely different feature. For this reason validation techniques are vital when using a method such as the one developed here, and in this case it is unfortunate that the validation could not be completed.

During the course of this work direct MS-MS fragmentation from tissue has been improved and optimised, and it is thought to be likely that this will be the main target for improvement in the future for identification of features found via MALDI-MSI, due to the advantages of spatial information, and the reduced need for time-consuming validation with traditional biochemical techniques. However, this technique is still only applicable to masses above 2.5 – 3 kDa. Additionally, the work here has been a useful exercise in developing a method for analysis of brain tissue using LC-MS, and has been applied to the following chapter ‘Assessment of Denator Heat Treatment for Prevention of Proteomic Sample Degradation using Label Free Relative Quantitation’. In this chapter the issue of enzymatic degradation changing the composition of tissue samples is also addressed. This is an important issue as uncontrolled sample degradation has been known to give mis-leading and difficult to interpret results.

This work in this chapter has clearly demonstrated the importance of sampling issues, as changes to tissue due to disease have been seen to cause global changes in protein and peptide concentration. Rigorous controls and standards can help to prevent this; however physical treatment of biological samples, such as tissue, must also be tightly controlled. One of the main conclusions made from this research has been the importance of experimental planning and thought prior to experimentation. Sample integrity is one of, if not the most important problems in biological proteomic research, and is of vital importance when sample availability is limited, as it was here. Prevention of sample degradation and maintenance of sample integrity is investigated further in the following chapters.

3 Assessment of Heat Treatment for Prevention of Proteomic Sample Degradation using Label Free Relative Quantitation

3.1 Introduction

The use of proteomic technologies for analysis of tissue continues to be a growing field with many important applications, including biomarker discovery and identification. These techniques can be applied to a variety of different disease states and tissue types, such as pancreatic cancer, lung cancer, and ovarian cancer (271-274). However, sample degradation of protein in tissue samples continues to be an ongoing challenge for proteomic analysis, as demonstrated in the previous chapter. Endogenous proteolytic enzymatic action can cause large changes in samples following collection and storage to point of analysis. Degradation can be rapid and sometimes semi-random, especially if cells are ruptured by freeze-thawing or tissue sectioning (275;276). Such degradation can distort any qualitative and especially quantitative proteomic analysis, particularly if degradation is variable. For instance, if there is the same degree of degradation in each sample this can be dealt with, however if degradation is variable between samples, this is when disortion of results can be particulary crucial. Traditionally the solution for this problem is to store samples at a low temperature (typically below minus 80 ºC) and to use protease inhibitors as early in the sample preparation as possible. However this can cause further problems. Protease inhibitors can interfere with mass spectrometric analysis, and snap freezing of tissue can cause damage and sample loss. Understanding and, if possible, controlling these degradation processes will be of significant importance within many areas of proteomic research. A sample preservation step that is applied as early as possible after sampling would seem like the ideal solution to this problem.

It has previously been reported that microwave irradiation can prevent changes of labile molecules, including cyclic AMP and phosphoproteins (277;278). The principle behind this is thermal inactivation of enzymes in the tissue. Further work has demonstrated that animals sacrificed with focused microwave irradiation have higher levels of detectable neuropeptides, than traditionally sacrificed animals (279-281). Specifically, Svensson *et al* (282) showed that neuropeptides were detected in mice and rats sacrificed using a focused microwave irradiation method. A major disadvantage of this method is that the animals must be restrained in a stressful manner prior to sacrifice, and additionally the cost of the

focused irradiation device (over $40,000). It was hypothesized by Che *et al* (283) that protein degradation occurred in the time between decapitation and dissection of the brain. They demonstrated that a conventional microwave oven can be used post-mortem to prevent protein breakdown in mouse brain tissue, simply by placing the entire animal head in the microwave. The temperature of the brain was monitored using a probe thermometer, and varied from 60 – 90 ºC. This work was a good proof of the principle behind this theory, however there were certain issues with the research. Of particular worry is the large variation in temperature which cannot be controlled within a microwave oven.

In my work presented here a denaturing device (Denator AB, Gothenburg, Sweden) has been evaluated for sample preservation using brain tissue as a sample tissue. This instrument works using a combination of pressure and heat to rapidly and homogenously heat the tissue sample preventing proteolytic activity by denaturing enzymes and proteins. It is proposed that this fast, homogenous heat distribution will give more effective, reproducible sample preservation than alternative denaturing methods, or by sample processing at low temperature alone. The use of heat-treatment on tissues immediately postmortem, to halt enzymatic degradation by denaturing both enzymes and proteins within tissue, is described in this work. The precise and rapid heating of tissue can be demonstrated to have a range of beneficial affects such as reducing the number of peptide fragments derived from abundant proteins present in samples, and also allowing a significant increase in the period that can be spent processing samples at room temperature without degradation of protein and peptide molecules.

High performance liquid chromatography coupled to electrospray mass spectrometry (HPLC-ESI-MS) is a well established technique for protein identification, as discussed in the previous chapter. It is especially applicable to identification of proteins in complex biological mixtures, such as body fluids or tissue extracts, meaning it is an ideal technique to quantify differences caused by sample degradation. In this work HPLC-ESI-MS is used for identification of proteins, in conjunction with database interrogation. A label free relative quantitation method is applied to quantify the differences seen between conventionally treated snap frozen samples and heat treated samples, both initially and following sample manipulation. Label free relative quantitation is based on the principle of comparing identical proteolytically digested peptides, which elute with the same retention time using LC-MS analysis, from different samples. The relative ratios of the protein can then be calculated. The benefits of label free quantitation are discussed in more detail in the introduction in section 1.9.

Mouse brain tissue is used from WKY (wild type) mice. In addition, the time the mouse brain can be left at room temperature, following heat treatment or snap freezing, without detrimental effect on the proteomic data is also evaluated.

3.2 Aims

There are several important aims in this chapter, which will be discussed in more detail in the following sections.

- Evaluate the overall effect of heat treatment on the mouse brain tissue sub10 kDa proteome

- Evaluate the overall effect of heat treatment on the mouse brain tissue digested sub10kDa proteome

- Development of a label free relative quantitation method in order to quantify the differences seen between heat treated samples and snap frozen samples in both digested and intact sub 10 kDa proteome

3.3 Methods

3.3.1 Sample Preparation

As described in section 2.3.1, Male ICR mice (6-8 weeks) were euthanatized by cervical dislocation in accordance with the U.K. Animals (Scientific Procedures) Act, 1986 and local ethical guidelines. The brain was rapidly removed and processed. Brains were dissected into the two hemispheres. One hemisphere was heat-treated (using the Denator system) and then snap-frozen in liquid nitrogen, while the other hemisphere was only snap-frozen in liquid nitrogen. The removal and treatment of the brain was accomplished in approximately 60 seconds. Samples were then stored at minus 80 ^{o}C until required. All subsequent movement of tissue was done on dry ice, with all containers pre-chilled to minimize the possibility of localized tissue warming. Tissue sections were cut on a cryostat microtome to 10 µm thick sections (Leica Microsystems CM 1900UV, UK), with the microtome chamber pre-chilled to minus 20 ^{o}C. Samples were then placed directly into chilled sample tubes (Eppendorf).

As a time point study of degradation was to be performed, samples were allowed to thaw at room temperature for differing lengths of time:

- Heat treated 0 minutes at room temperature
- Heat treated 5 minutes at room temperature
- Heat treated 10 minutes at room temperature.

- Snap frozen 0 minutes at room temperature
- Snap frozen 5 minutes at room temperature
- Snap frozen 10 minutes at room temperature.

Each group contains 3 biological replicates. Each sample was split into two aliquots to create two separate sample sets.

This allows a comparison of protein degradation over time, between frozen and heat treated samples. Samples were then split into two further sets. One set was analysed as intact protein samples, and the second set of samples were prepared for tryptic digestion. Therefore there will be 3 replicates for each sub sample set.

For generation of intact protein samples 50 µl of a 1 % formic acid solution was added to the tissue sample, in the sample tube. This was then centrifuged for 10 seconds at 14,000 rpm, before being vortexed vigorously and centrifuged again. A 10 second probe sonication was then used for cell lysis. The sample was then vortexed vigorously and centrifuged at 14,000 rpm for 30 minutes at 4 ºC. The supernatant was removed and applied to a Microcon 10,000 Da molecular weight filter (as described in section 2.3.1). This was centrifuged for 30 minutes at 14,000 rpm at 4 ºC, prior to the addition of 50 µl of a 1 % formic acid solution. It was then centrifuged again for 30 minutes at 14,000 rpm at 4 ºC. The below 10,000 Da molecular weight fraction (the filtrate) was collected and then stored at minus 20 ºC prior to analysis. The samples for tryptic digestion were processed in the same way except that the 1 % formic acid solution was replaced with a 25 mM ammonium bicarbonate solution. Following the size filtration step, 2 µl of a 0.5 µg/µl sequencing grade porcin trypsin (Promega, catalogue number V5111) was added to each sample, and left to incubate overnight at 37 ºC.

3.3.2 LC-MS

Following size filtration and digestion (as described in section 3.3.1) the samples were then analysed by HPLC-ESI-MS using a Dionex 3000 LC system and a Bruker (Banner Lane, Coventry) HCT Ultra. A 60 minute gradient separation was used (5 % Solvent B to 80 % solvent B). The LC separation was performed on a 15 cm C18 PepMap reverse phase column with a 75 µm internal diameter (Dionex). This was stored in a 37 ºC oven for the duration of the separation. Solvent A was a 2 % acetonitrile, 0.1 % formic acid solution and solvent B was an 80 % acetonitrile, 0.1 % formic acid. The loading solution used was a 2 % acetonitrile, 0.5 % tri-fluoroacetic acid solution. The loading flow, onto the trap was maintained at 30 µl/min. The column flow was maintained at 0.3 µl/min. MS data was collected from m/z 150-2000, in positive ion mode, with an accumulation time of 3 seconds. Decyder MS (GE Healthcare, Amersham Place, Little Chalfont, Buckinghamshire) was used for label free relative quantitation, using the following settings:

Typical peak width of 0.5 minutes, Trap resolution of 0.4 u, charge states of 1-4, uniform background model, signal-to-background detection threshold of 4.0, background subtracted quantitation, charge assignment enabled where possible, charge assignment from two peaks limited, LC peak shape tolerance of 20 %, m/z shift tolerance of 0.1 u, m/z shape tolerance of 5 %, remove peptides below S/N of 4.0, remove peptides of unspecified

charge below S/N of 4.0, remove peptides with low quality LC peaks, remove overlapping peptides, spectrum vs model tolerance of 15 %.

Settings for spectrum alignment in Decyder MS were as follows:

Max/stretch compress: 2, Max leader and trailer: 10 %, Stretch compress penalty: 0.1

Settings for peptide matching in Decyder MS were as follows:

Time tolerance: 1 minute, Mass tolerance: 0.5 Da, Use aligned retention times, Cross detection: Add peptides to intensity maps if present in 1 map

Normalisation was performed using a measured intensity distribution. This means that the entire peptide population is used for normalisation. It is assumed that most samples contain peptides that do not vary in intensity and the normalisation is performed without any further specifications. This is the recommended method of normalisation, by the maufacturers of the software, unless very few peptides are present in the samples.

The settings used for the Decyder MS analysis were optimised using instructions and advice from the manufacturer. The optimisation process included ensuring the correct peak width and resolution are used for the instrument in question. Additionally, it ensures noise is filtered efficiently and charge assignment is performed correctly.

As previously mentioned there are 6 sample groups, and two separate sample sub-sets. This is summarised as follows:

Heat Treated	**Snap Frozen**
0 minutes at room temperature 3 samples – Intact 3 samples – Digested	0 minutes at room temperature 3 samples – Intact 3 samples – Digested
5 minutes at room temperature 3 samples – Intact 3 samples – Digested	5 minutes at room temperature 3 samples – Intact 3 samples – Digested
10 minutes at room temperature 3 samples – Intact 3 samples – Digested	10 minutes at room temperature 3 samples – Intact 3 samples – Digested

Table 3.1 Table summarising samples sets

3.3.3 Statistical Analysis of LC-MS data for Label Free Quantitation

The LC-MS data was analysed using one-way ANOVA (Analysis of Variance). This is discussed in more detail in the results section 3.4.2.

3.3.4 LC-MS-MS for Targeted and Untargeted Protein Identification, using Ion Trap MS

Subsequent to masses of interest being identified (as described in section 3.3.2), the samples were re-analysed, using the Dionex 3000 nano-flow LC system and a Bruker HCT Ultra. As in section 3.3.2, a 60 minute gradient was used. The LC separation was performed on a 15cm C18 PepMap reverse phase column with a 75 µm internal diameter, from Dionex. This was stored in a 37 ºC oven for the duration of the separation. Solvent A was a 2 % acetonitrile, 0.1 % formic acid solution and solvent B was an 80 % acetonitrile, 0.1 % formic acid. The loading solution used was a 2 % acetonitrile, 0.5 % tri-fluoroacetic acid solution. The loading flow, onto the trap was maintained at 30 µl/min. The column

flow was maintained at 0.3 µl/min. MS data was collected from m/z 150-2000, in positive ion mode, with an accumulation time of 3 seconds. Targeted MS-MS (as described in the introduction, in section 1.3) was used in conjunction with MASCOT database searching for peptide and protein identification. The targeted MS-MS was performed using a simple 'Include List' strategy, in which the mass spectrometer preferentially performs MS-MS fragmentation of the specified masses. Again, this was repeated for the lower than 10,000 Da tryptically digested fraction, and the lower than 10,000 Da intact fraction.

Electrospray was performed using a Proxeon 30um x 50mm stainless steel nano-electrospray emitter. Settings used were as follows:

Ion Mode: Positive, Enhanced, Mass Range: m/z 500-3000, Max Fill Time: 250 ms, MS/MS Scans: 5 per MS (Ultrascan).

Samples were re-analysed in MS-MS mode, without an include list to identify any peptides that were missed in the targeted MS-MS experiment. In this case MS-MS was performed on the most intense masses detected, as is the usual setting for MS-MS.

All files were converted into MASCOT compatible 'mgf' files using Data Analysis (Bruker). They were then searched with MASCOT version 2.1 against database NCBInr 20080422 (6468149 sequences; 2207779427 residues), using MASCOT search settings as follows:

Species: Rodentia, Enzyme: Trypsin/No enzyme (depending on sample), Report top hits: Auto, Variable modifications: Oxidation of Methionine, Charge: 1+, 2+ and 3+, Peptide tolerance: 1.2 Da, MS/MS tolerance: 0.8 Da, Instrument: ESI-Trap.

3.3.5 LC-MS-MS for Targeted and Untargeted Protein Identification using Q-TOF-MS

The samples were also analysed using a Dionex Ultimate nano-flow LC system and an Applied Biosystems Q-Star, again both with and without an MS-MS include list. In this case an identical 60 minute gradient was used. The LC separation was performed on a 15cm C18 PepMap reverse phase column with a 75 µm internal diameter, from Dionex. Solvent A was a 2 % acetonitrile, 0.1 % formic acid solution and solvent B was an 80 % acetonitrile, 0.1 % formic acid. The loading solution used was a 2 % acetonitrile, 0.5 % tri-fluoroacetic acid solution. The loading flow, onto the trap was maintained at 0.2 ml/min.

The column flow was maintained at 0.3 µl/min. This LC system does not have a column oven, therefore the column was maintained at room temperature throughout the separation.

Q-Star settings used were as follows:

Accumulation time: 3 seconds, TOF masses: m/z 200 – 1500, Polarity: Positive, Cycle time: 15 seconds, Product ion masses: m/z 50 – 2000, dynamic exclusion was used, and rolling collision energy. Electrospray was performed using 'picotip' needles (New Objective) and the standard Q-Star electrospray source.

Q-star 'wif' files were searched against database NCBInr 20080422 (6468149 sequences; 2207779427 residues), using MASCOT search settings as follows:

Species: Rodentia, Enzyme: No enzyme or trypsin (as appropriate), Report top hits: Auto, Variable modifications: Oxidation of Methionine, Charge: 1+, 2+ and 3+, Peptide tolerance: 1.2 Da, MS/MS tolerance: 0.8 Da, Instrument: ESI-TOF

3.4 Results

3.4.1 Visual Evaluation of the Heat Treatment effect on Tissue

The effect of the heat treatment on the structure of the brain tissue was observed. The rigidity of the tissue was seen to increase. Little change was observed in the overall brain anatomy. The tissue appeared darker in colour following the heat treatment. Following heat treatment the tissue was snap frozen and thawed without cracking or loss of structure that is often seen when snap freezing without heat treatment. There was also an observed reduction in odor from the thawed tissue when compared with untreated tissue.

3.4.2 LC-MS and Label Free Relative Quantitation

The LC-MS data obtained from the tissue samples, as detailed in section 3.3.2, was analysed using one-way ANOVA (Analysis of Variance). The one-way ANOVA method performs comparisons on an arbitrary number of populations. In Decyder MS, each non-empty experimental group is considered as an ANOVA population. The ANOVA is used for each feature to test the null hypothesis that there is no difference in the abundance of the feature between the groups. The test does not indicate which groups are different from which other groups; it only indicates that there is an overall difference. The ANOVA null hypothesis is that the means of all populations are equal. The test is used to compare the variance between populations to the variance within populations. The ratio of these variances is known as the F-ratio, and this is used to calculate the p-value. A large F-ratio corresponds to a small p-value, which therefore indicates that the variation between populations is large in comparison to the variation within populations. This would mean that the populations may show a difference in abundance of the particular feature being tested. If the ANOVA p-value between two populations is 0.01, then under the null hypothesis the probability of obtaining the observed difference in abundance by chance is 1 in 100. The ANOVA test assumes that the data points are normally (or approximately normally) distributed, and also that the variance of the populations are equal. Small deviations from a normal distribution are unlikely to affect the calculation greatly. However, if the variance among the groups differs greatly the ANOVA p-value may not be valid. This must be taken into account when using ANOVA p-values to determine differences in feature abundance between groups. In this analysis one of the most important decisions to be made is at what level to set a significant p-value. A common method is to simply choose an arbitrary value of 0.01, 0.005 or 0.001, and this is the

method of choice used in this chapter. This would mean that the probability of obtaining the observed difference by chance would be 1 in 100, 0.5 in 100, or 0.1 in 100, respectively.

There are many other ways of statistically analysing data-sets such as these. This is explored and discussed more fully in the following chapter, “Discovery and Identification of Biomarkers for Hypertension”. However, in this chapter the fairly universally accepted method, of using arbitrarily defined ANOVA p-values is the method of choice.

The samples analysed are summarised as follows

- Sub 10kDa intact samples (heat treated and snap frozen) see section 3.4.2.1
- Sub 10kDa digested samples (heat treated and snap frozen) see section 3.4.2.2

3.4.2.1 Sub 10kDa Intact Sample Results

The LC-MS datasets, as generated using the sub 10kDa samples (described in section 3.3.2), were analysed using Decyder MS label free relative quantitation software. Each mass detected throughout the dataset will be referred to as a ‘Feature’, as the only information available at this stage in the analysis is mass, retention time, and relative intensity between groups.

The total number of features detected across all 18 samples was 6034. An ANOVA score was calculated for each feature, across all 6 groups. This indicates whether that particular feature has a significant change between the groups. ANOVA calculates the within group variance, and compares it to the between groups variance. Therefore, the value does not tell us anything about the change in intensity, other than whether it is significant, or not. This means that this ANOVA score gives a starting point for the analysis. The expression profiles of these significant features must be analysed further for quantitation purposes.

Various significant p-values were defined, and peptides with a p-value equal to or less than this were deemed to have a significant change between sample sets. This is summarised in the table 3.2:

Significant p value	Number of features with significant changes	Percentage of features showing significant changes
≤ 0.001	50	0.83 %
≤ 0.005	80	1.33 %
≤ 0.05	200	3.31 %

Table 3.2 Summary of label free relative quantitation of heat treated and snap frozen samples (<10kDa intact samples). Shows three differing arbitrarily defined ANOVA p-values, and the affect this has on the percentage of features demonstrating significant changes between the six sample groups. (Total number of features identified was 6034)

It is shown that even with a fairly high significant p-value, only 3.31 % of the total features identified have significant changes. This indicates that the heat treatment does not substantially alter the detected proteome, when compared with the snap frozen samples, and that the majority of proteins and peptides in the brain tissue do not degrade, or degrade to the same extent, regardless of the treatment applied to the samples. The small percentage of features that do have significantly different abundances in the differentially treated samples are of interest in this work These features may be proteins and peptides that are either preserved, or stabilised, by the heat treatment, or equally could be destroyed by the treatment. The expression profiles across the sample sets are investigated for these particular features, and they will also be targeted for identification using MS-MS analysis. This will allow the heat treatment system to be assessed, in comparison with the conventional snap freezing treatment.

All expression profiles for the 50 features with p-values ≤ 0.001 are shown in appendix 3. From visual examination of the profiles it is demonstrated that 7 show expression profile as predicted by figure 3.2 C and 25 show expression profile as predicted by figure 3.2 D. Predicted profiles are shown below.

The average number of detected features across the sample groups is summarised in figure 3.1. The average number of features detected across all six groups was 1186 with a standard deviation of 188.4.

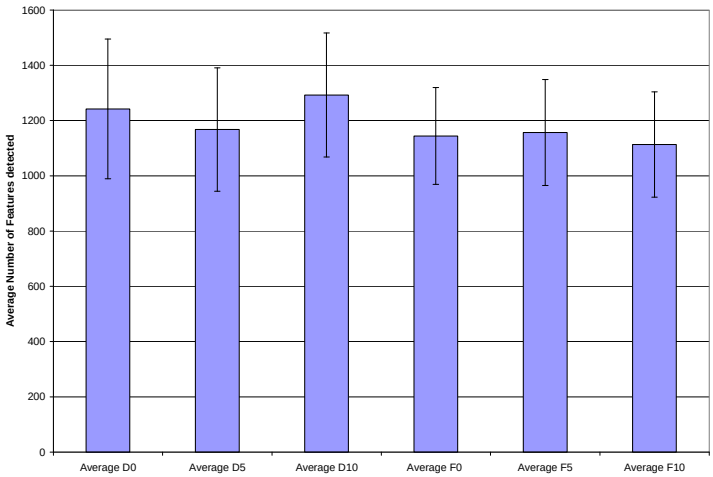


Figure 3.1 Bar chart of average number of features detected by label free relative quantitation, for samples, sub 10kDa

As the label free relative quantitation gives information for each feature regarding, mass, retention time and relative intensity, each feature can be visualised using what is known as an 'Expression Profile'. This allows a simple visualisation of the relative intensity of the feature in question between the different groups.

'Relative intensity' is a measure of how abundant the feature is, within the sample. In mass spectrometry the abundance of a feature is measured by the current produced by the number of ions arriving at the detector, therefore the relative, not the absolute, intensity is measured. In addition, the LC-MS data expressed here has been normalised against the total intensity of all peptides in the sample, as discussed in section 3.3.2, therefore the 'relative intensity' of each feature presented in this work refers to the abundance of each feature relative to the other features present in the sample set.

As mentioned previously the features which change significantly across the sample groups are of particular interest here. The predicted profiles that would demonstrate different effects caused by the heat treatment process on features within the samples are shown in figure 3.2 (A-D):

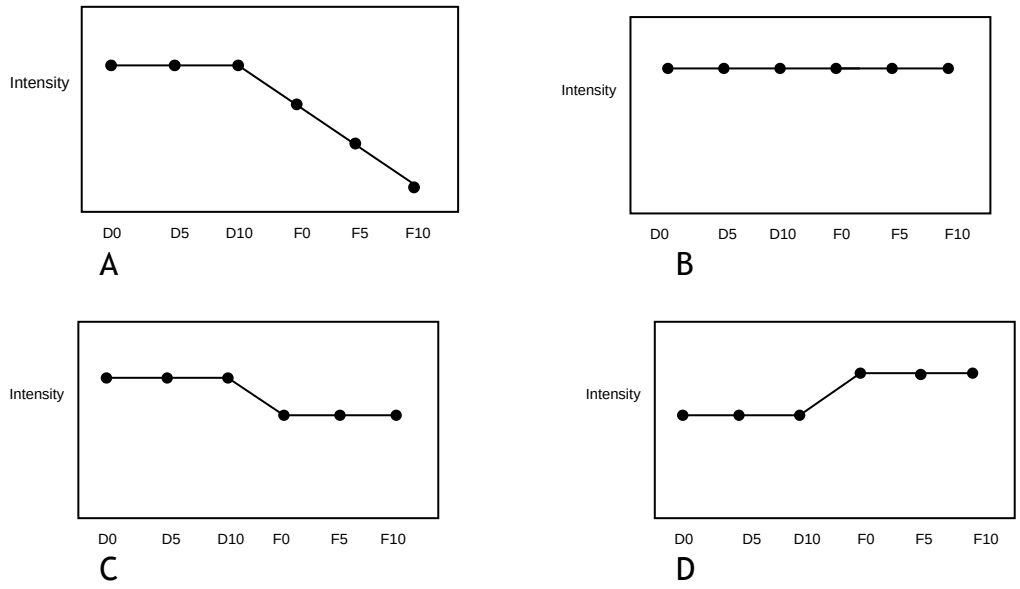


Figure 3.2 D0, D5 and D10 refer to the three timepoints, denator heat treated. F0, F5 and F10 refer to the three time points, snap frozen. These potential expression profiles have several possible explanations, which are expanded below.

Figure 3.2 (A) shows the predicted expression profile indicating that the heat treatment has stabilised this feature prior to, during and following slicing of the tissue. Figure 3.2 (B) shows the predicted expression profile indicating that the heat treatment has had no effect on the tissue. Figure 3.2 (C) shows the predicted expression profile indicating that the heat treatment has stabilised this feature prior to and during slicing of the tissue. Figure 3.2 (D) shows the predicted expression profile indicating that the heat treatment has stabilised this feature prior to and during slicing of the tissue, suggesting a larger mass has degraded to produce a higher intensity of a smaller mass. Conversely, this could show that the heat treatment has degraded this feature prior to and during slicing

In an attempt to focus purely on the features with the largest and most significant changes, only those with an ANOVA p-value of < 0.001 were focused upon. Example expression profiles for a selected number of these features are shown in figures 3.3 to 3.6.

These figures relate to the mean relative intensity within each group, with error bars representing +/- one standard deviation. The y-axis shows the Average Log_2 Intensity, and the x-axis represents the six different sample groups. The expression profiles seen here are visually compared with the predicted expression profiles shown in figure 3.2.

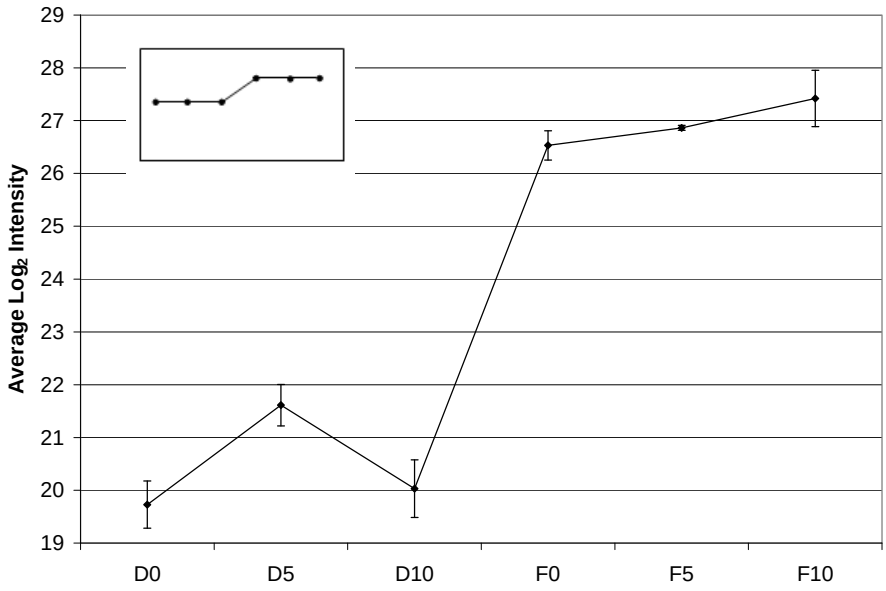


Figure 3.3 Expression profile showing relative mean intensity across all six groups of the feature with mass 1832.980 Da and p value of 1.66E-11. Insert shows predicted expression profile for comparison.

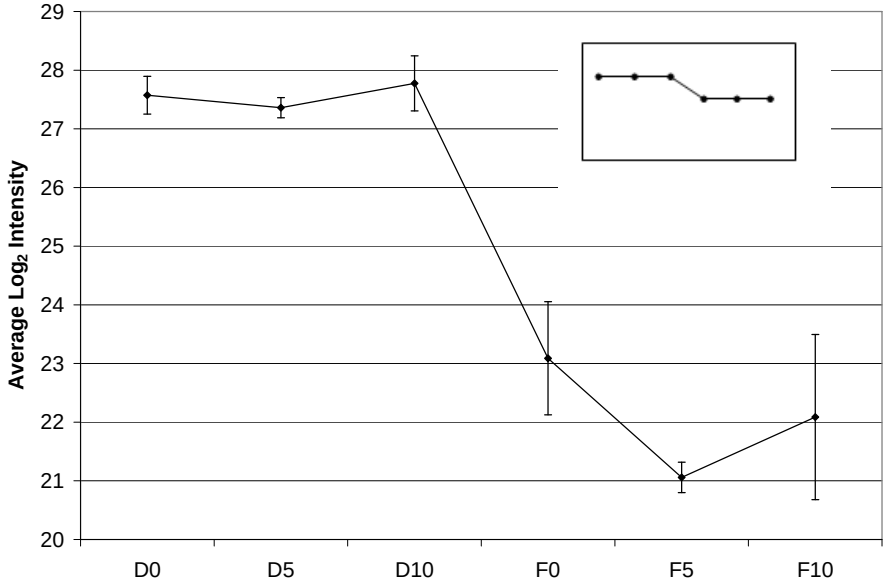


Figure 3.4 Expression profile showing relative mean intensity across all six groups of the feature with mass 1228.512 Da and p value of 1.11E-07. Insert shows predicted expression profile for comparison.

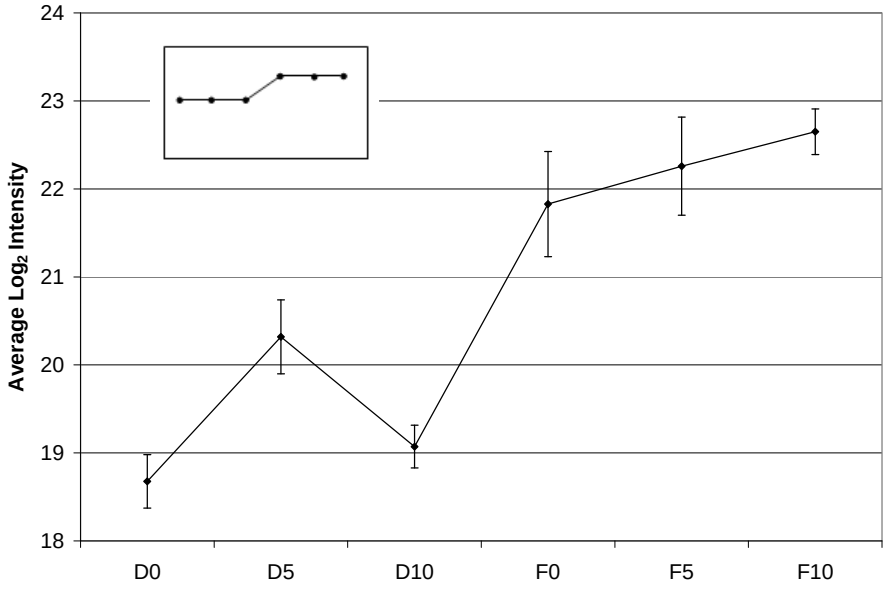


Figure 3.5 Expression profile showing relative mean intensity across all six groups of the feature with mass 2021.021 Da and p value of 1.52E-07. Insert shows predicted expression profile for comparison.

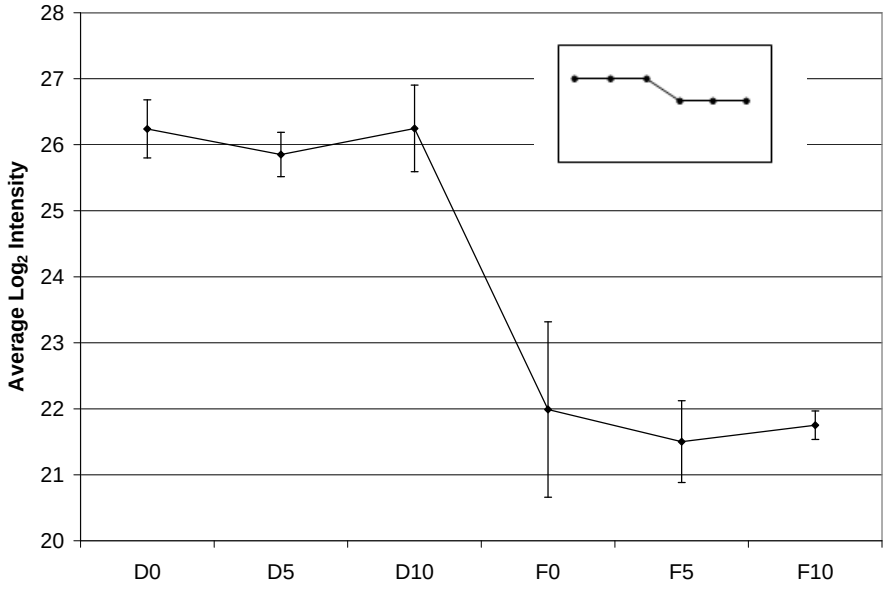


Figure 3.6 Expression profile showing relative mean intensity across all six groups of the feature with mass 543.124 Da and p value of 8.87E-07. Insert shows predicted expression profile for comparison.

These selected expression profiles, as seen in figures 3.3 to 3.6, demonstrate that the heat treatment has a range of effects on different features. It may degrade some proteins and peptides, and conversely it may denature enzymes, stabilising and preserving other proteins and peptides. Therefore, the heat treatment may enable detection of features that are otherwise degraded prior to mass spectrometry treatment. This may mean detection of interesting biomarkers and disease targets. The features that have differing abundances

between the groups are targeted for identification, to enable further insight into the effects of the heat treatment, as opposed to conventional snap freezing.

3.4.2.2 Sub 10kDa Digested Sample Results

The sample data from the digested samples, prepared as described in section 3.3.1, was analysed as described in section 3.3.2, using Decyder MS label free quantitation software. The total number of features detected across all 18 samples was 4282. Again (as in section 3.4.2.1), various significant p-values were defined, and peptides with a p-value equal to or less than this were deemed to have a significant change. This is summarised in the table 3.3:

Significant p value (from ANOVA)	Number of Features with significant changes	Percentage of Features showing significant changes
≤ 0.001	39	0.91 %
≤ 0.005	104	2.43 %
≤ 0.05	376	8.78 %

Table 3.3 Summary of heat treatment analysis of <10kDa digested samples. Shows three differing arbitrarily defined ANOVA p-values, and the effect this has on the percentage of features demonstrating significant changes between the six sample groups. (Total number of features identified was 4282).

The distribution of total features detected is summarised in figure 3.7 below. There are less total features detected in the digested samples in comparison to the intact samples; however there are more features detected in each individual sample. This indicates that there are a larger number of the same features detected in each digested sample in comparison to the intact samples. This is possibly due to the tryptic digestion creating reproducible samples, which are easier to analyse using LC-MS. Peptide samples may have more reproducible LC retention times, and are less likely to block columns and be lost due to sample processing, in comparison with the undigested samples.

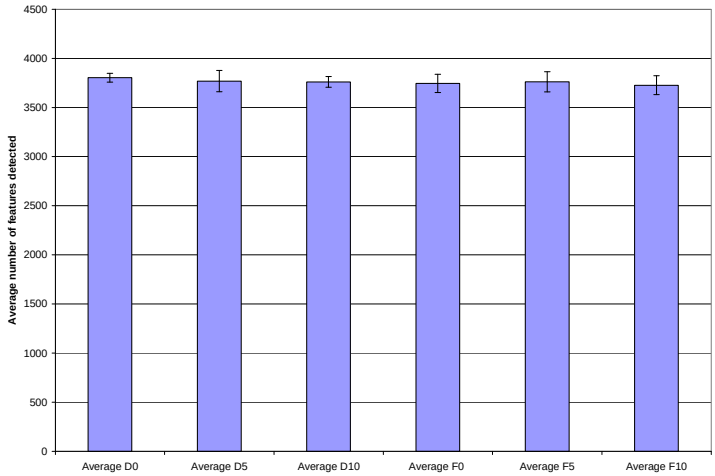


Figure 3.7 Bar chart of average number of features detected by label free relative quantitation, for samples, sub 10kDa tryptically digested

Again as in section 3.4.2.1, in an attempt to focus purely on the features with the largest changes, only those with an ANOVA p-value of < 0.001 were focused upon. Example expression profiles for a selected number of these features are shown in figures 3.8 to 3.11 below. These are also visually compared with the predicted expression profiles, shown in figure 3.2.

All expression profiles for the 39 features with p-values ≤ 0.001 are shown in appendix 4. From visual examination of the profiles it is demonstrated that 6 show expression profiles which match figure 3.2 C and 5 show expression profile which match figure 3.2 D. Predicted profiles are shown below.

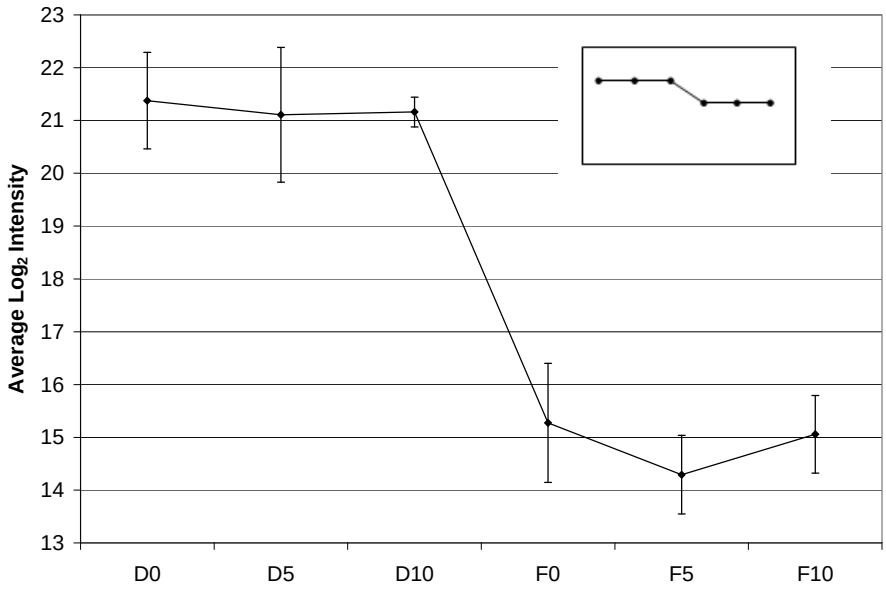


Figure 3.8 Expression profile showing relative mean intensity across all six groups of the feature with mass 654.489 Da and p value of 4.53E-06. Insert shows predicted expression profile for comparison.

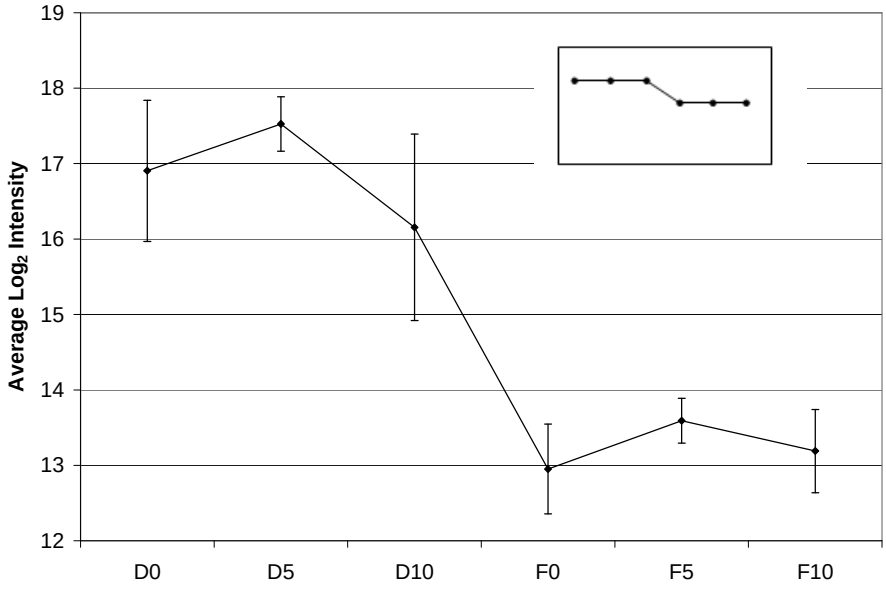


Figure 3.9 Expression profile showing relative mean intensity across all six groups of the feature with mass 289.713 Da and p value8.95E-05. Insert shows predicted expression profile for comparison.

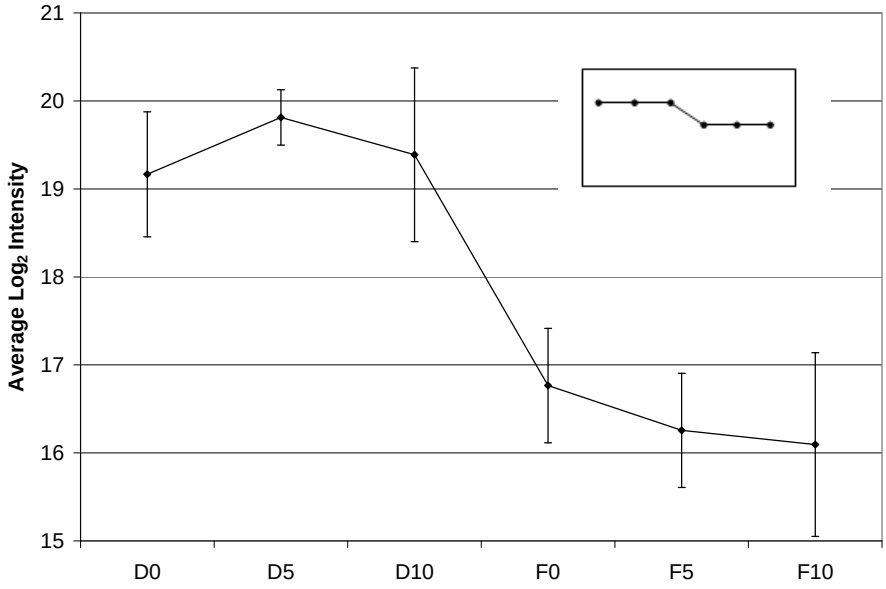


Figure 3.10 Expression profile showing relative mean intensity across all six groups of the feature with mass 1114.749 Da and p value of 0.000127. Insert shows predicted expression profile for comparison.

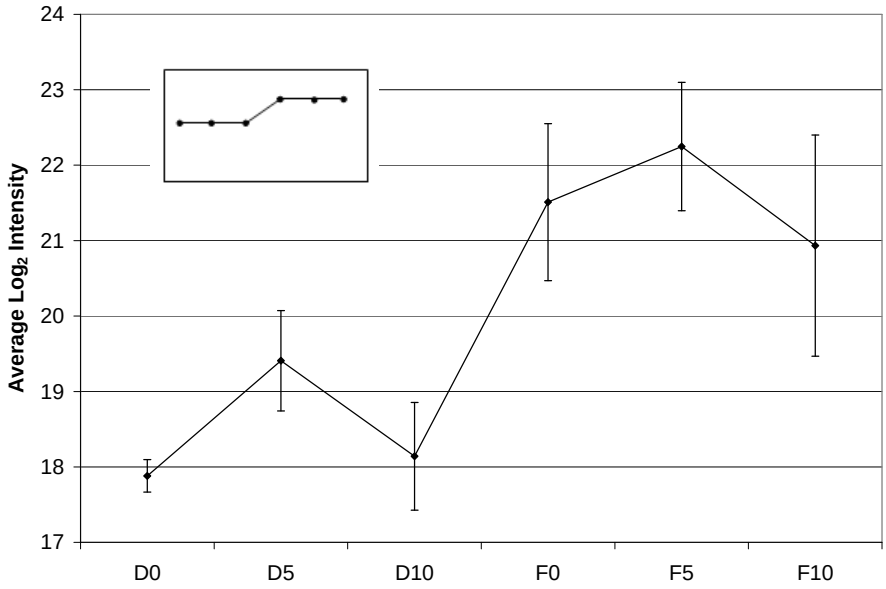


Figure 3.11 Expression profile showing relative mean intensity across all six groups of the feature with mass 537.157 Da and p value of 0.000646. Insert shows predicted expression profile for comparison.

As with the sub10kDa undigested samples, discussed in section 3.4.2.1, these selected expression profiles demonstrate that heat treatment will have a differing effect on different types of features. It may degrade some proteins and peptides, and conversely it may denature enzymes, stabilising and preserving other proteins and peptides. This means that the heat treatment may enable detection of features that are otherwise degraded prior to mass spectrometry treatment. This may mean detection of interesting biomarkers and disease targets. Again, the features that have differing abundances between the groups are

targeted for identification, to enable further insight into the effects of the heat treatment, as opposed to the traditional snap freezing.

3.4.3 LC-MS-MS for Targeted Protein ID using Ion Trap MS

3.4.3.1 <10kDa Intact Sample Results

It should be noted that the features of interest are identified by mass, retention time and relative quantity across the six groups only. There is still no information available regarding the actual identity of the feature. To remedy this situation, two separate mass spectrometers were used to target the masses of interest for fragmentation and subsequent identification. The features within the <10kDa intact samples proved extremely challenging to identify. Targeted MS-MS was performed for the features of interest, on these samples. Very poor quality MS-MS spectra were obtained, of low intensity. Unfortunately no identifications were able to be made from these, and this data is not presented here.

As in chapter 2, the samples here have not been digested in the traditional way using trypsin, meaning it can be challenging to identify the peptides using collision induced dissociation fragmentation. The reasons for this are discussed in more detail in section 2.4.3.

Potentially a different manner of fragmentation could be used to gain identifications, such as electron transfer dissociation (284) or electron capture dissociation (285). These methods were not available in this research, but could be potential avenues for investigation in the future.

3.4.3.2 Sub 10kDa Digested Sample Results

Targeted MS-MS was also performed for the features of interest, on the sub10kDa digested samples. These features in the digested samples are likely to be tryptic peptides, as discussed in section 3.4.2.1, and hence proved rather more straightforward to identify using the CID fragmentation method. Therefore, this analysis resulted in the identification of a number of the significantly changing features, and these are summarised in table, 3.3.

Table 3.3 shows that the percentage changes in masses between the original LC-MS data, and the following, targeted LC-MS data is very small (less than 0.04 %). This is as

expected, as the mass spectrometer was calibrated weekly, to minimise mass inaccuracy. The LC retention times show a larger percentage difference between the original LC-MS data, and the following, targeted LC-MS data. This is possibly due to slightly differing conditions between the two experiments, although in theory this should be minimised with strict temperature and flow control on in the LC system. Other possible explanations could be slight differences in solvents used, slight voids in the system or slightly differing column properties. It can be seen, from examination of table 3.3, that there are several identifications which have two peptide identifications. This can be used as a useful validation of results, as peptides from the same protein should, show similar expression profiles. Identifications were matched back to original expression profiles, and these are shown in figures 3.15 to 3.25.

Mass (Da)	RT (min)	Identification	gi number	RT in MS-MS run (min)	Change in RT (min)	% Change in RT	m/z in MS-MS run	Mass in MS-MS run	Change in Mass (Da)	% Change in Mass	Higher abundance seen in:	Abundance difference (average fold change)
1114.749	13.9	Prosomatostatin	gi|207019	15.6	1.7	12.23	558.384	1114.6814	0.0676	0.0061	Heat treated	8.5
1089.384	27.6	Purkinje cell protein 4	gi|6679227	24.3	-3.3	-11.96	545.87	1089.7254	-0.3414	-0.0313	Heat treated	7.1
1459.848	33.5	Myelin basic protein	gi|199051	31	-2.5	-7.46	487.616	1459.8262	0.0218	0.0015	Snap frozen	20.1
1130.705	15.2	Myelin basic protein	gi|199051	16.6	1.4	9.21	566.34	1130.6654	0.0396	0.0035	Snap frozen	7.0
1153.619	37.7	Peptidylprolyl isomerase A	gi|6679439	32.1	-5.6	-14.85	557.886	1153.5655	0.0535	0.0046	Heat treated	19.9
2004.98	52.6	Peptidylprolyl isomerase A	gi|6679439	44.5	-8.1	-15.40	669.333	2004.9772	0.0028	0.0001	Heat treated	16.0
1478.882	43.8	Alpha-fetoprotein	gi|191765	36.8	-7	-15.98	740.449	1478.8834	-0.0014	-0.0001	Heat treated	15.0
1698.979	38.6	Proprotein convertase	gi|7305455	31.8	-6.8	-17.62	850.488	1698.9614	0.0176	0.0010	Heat treated	2.1
1108.863	21.2	Keratin 10	gi|57012436	29.2	8	37.74	555.334	1108.6534	0.2096	0.0189	Snap frozen	3.6
1675.534	30.7	Secretogranin II	gi|6677865	25.7	-5	-16.29	838.953	1675.8914	-0.3574	-0.0213	Heat treated	3.1
2468.978	43.8	Secretogranin II	gi|6677865	36.2	-7.6	-17.35	824.094	2469.2602	-0.2822	-0.0114	Heat treated	3.4

Table 3.4 Summary of identifications from targeted MS-MS experiment, for <10kDa digested samples. A proportion of the target masses were identified. It can be seen that there is only a very small percentage change in mass from the target mass. However, some of the retention time changes are rather significant. It is proposed that this is due to poorly reproducible chromatography.

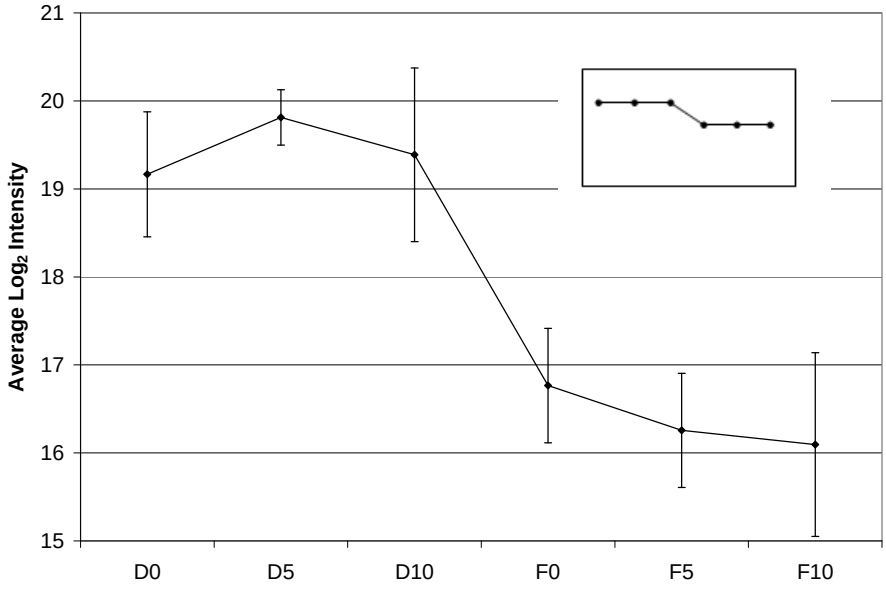


Figure 3.12 Expression profile showing relative mean intensity across all six groups of the feature identified as a peptide from Prosomatostatin with mass 1114.749 Da. Insert shows predicted expression profile for comparison.

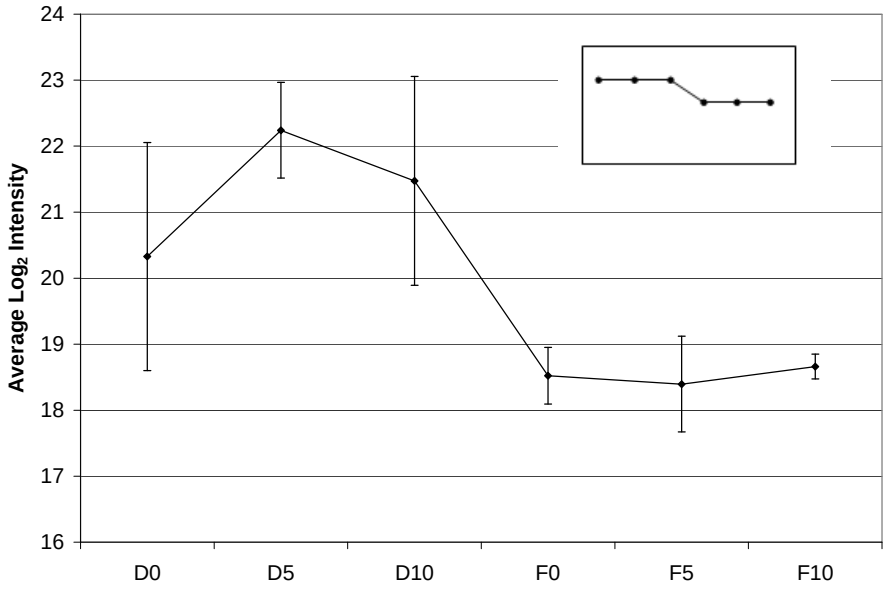


Figure 3.13 Expression profile showing relative mean intensity across all six groups of the feature identified as a peptide from Purkinje cell protein 4 with mass 1089.384 Da. Insert shows predicted expression profile for comparison.

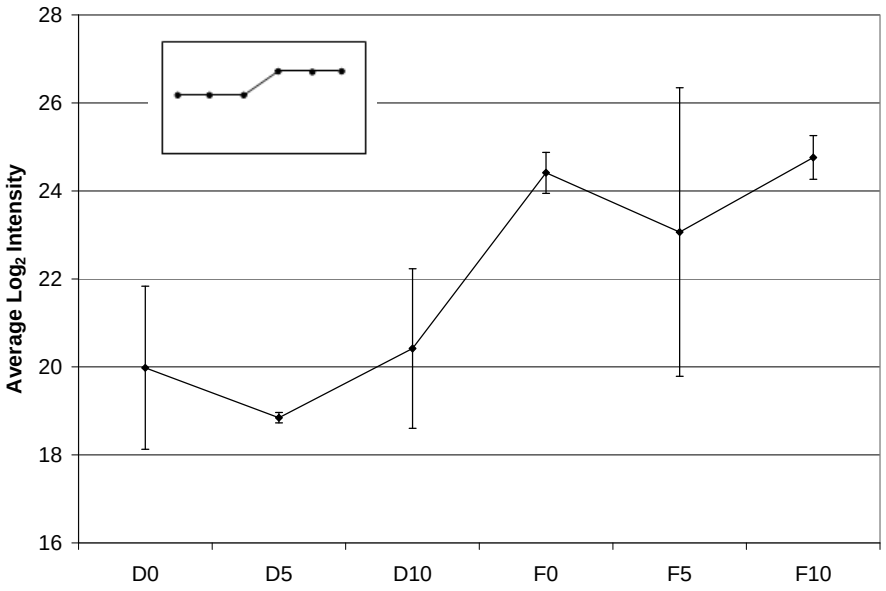


Figure 3.14 Expression profile showing relative mean intensity across all six groups of the feature identified as a peptide from Myelin Basic Protein with mass 1459.848 Da. Insert shows predicted expression profile for comparison.

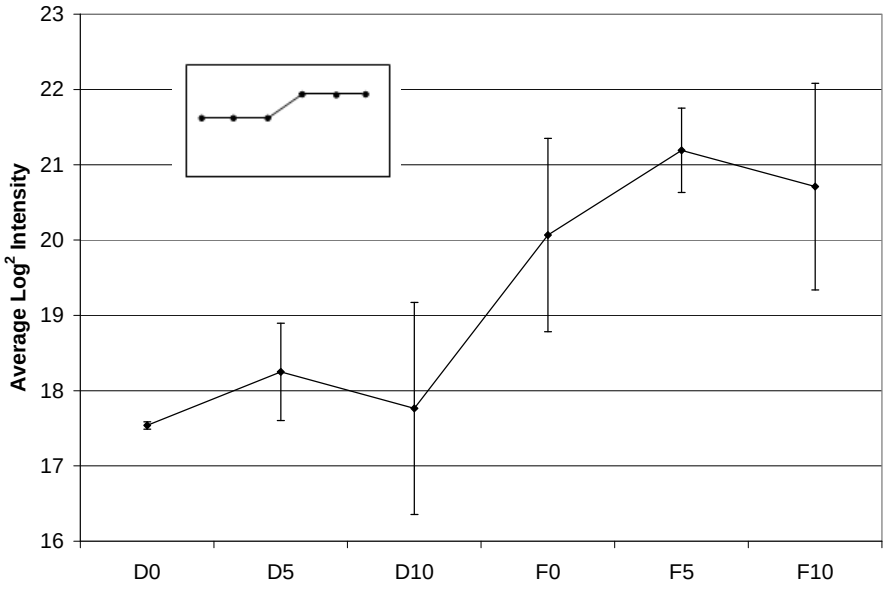


Figure 3.15 Expression profile showing relative mean intensity across all six groups of the feature identified as a peptide from Myelin Basic Protein with mass 1130.705 Da. Insert shows predicted expression profile for comparison.

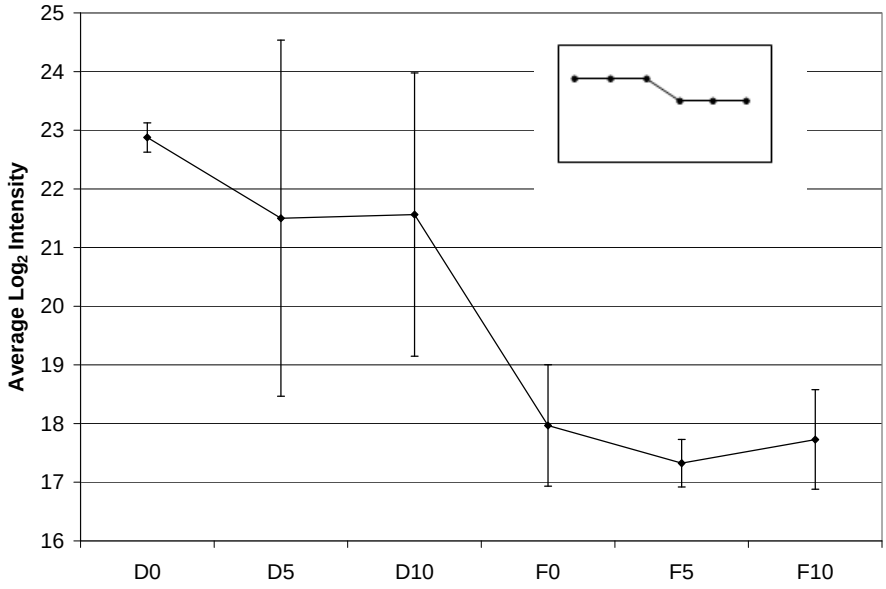


Figure 3.16 Expression profile showing relative mean intensity across all six groups of the feature identified as a peptide from Peptidylprolyl isomerase A with mass 1153.619 Da. Insert shows predicted expression profile for comparison.

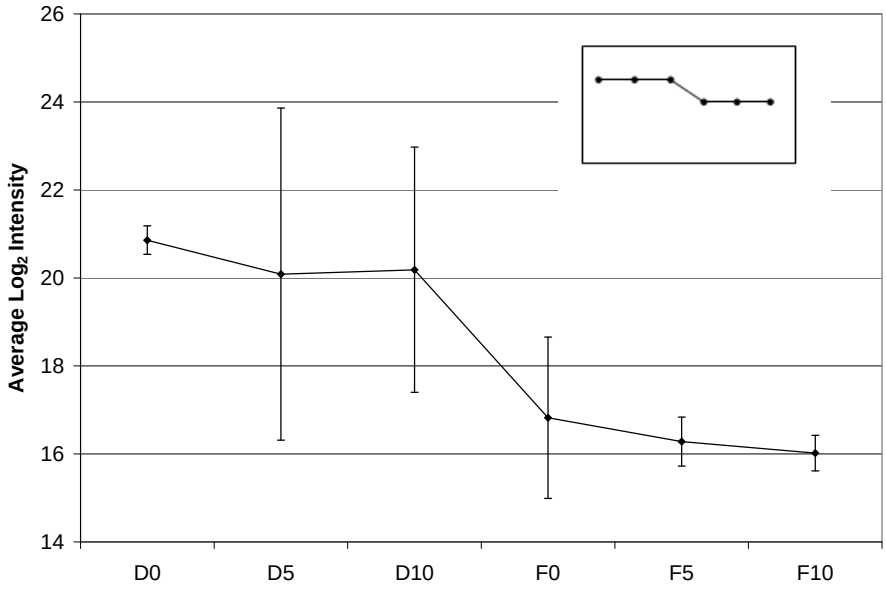


Figure 3.17 Expression profile showing relative mean intensity across all six groups of the feature identified as a peptide from Peptidylprolyl isomerase A with mass 2004.980 Da. Insert shows predicted expression profile for comparison.

It should be noted that this particular expression profile is very similar to that seen in figure 3.19. This is as predicted, as both these peptides are identified as being from the same protein.

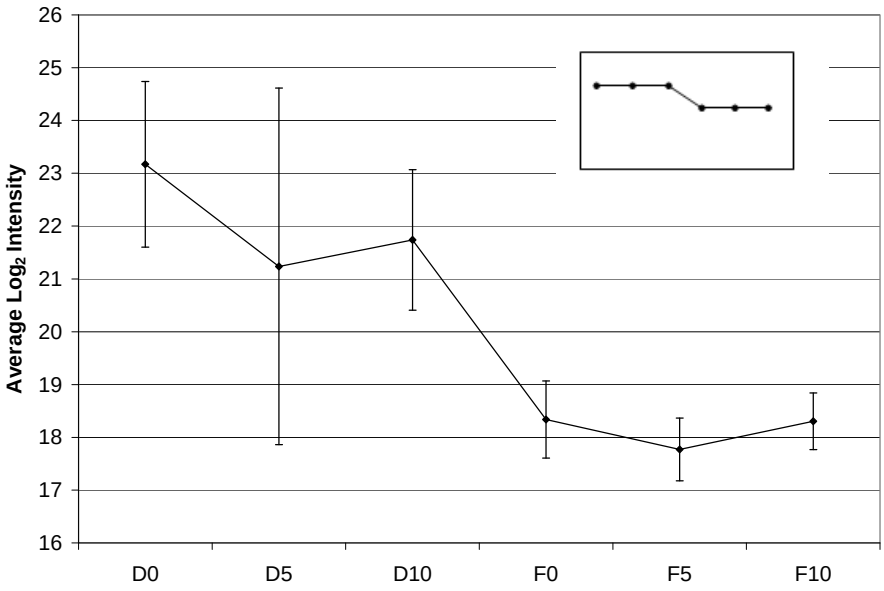


Figure 3.18 Expression profile showing relative mean intensity across all six groups of the feature identified as a peptide from Alpha-fetoprotein with mass 1478.882 Da. Insert shows predicted expression profile for comparison.

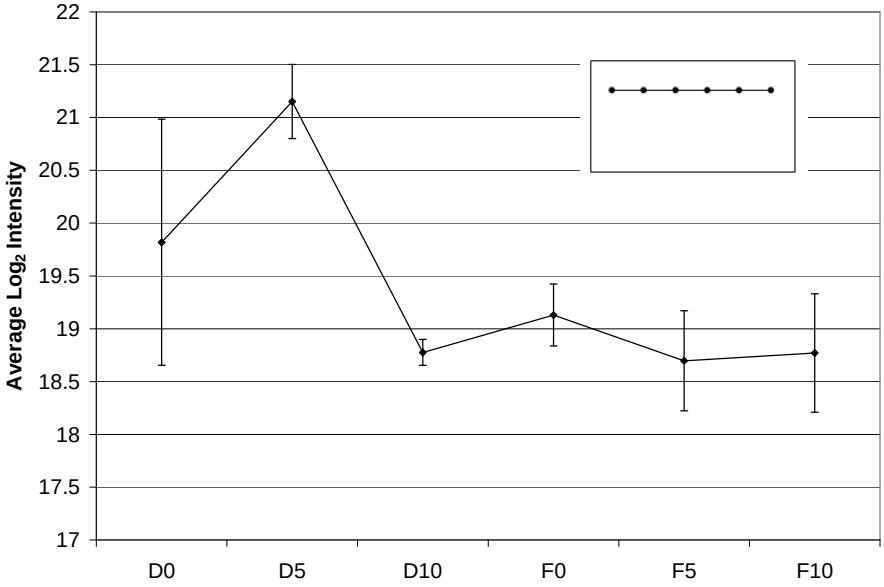


Figure 3.19 Expression profile showing relative mean intensity across all six groups of the feature identified as a peptide from Proprotein convertase with mass 1698.979 Da. Insert shows predicted expression profile for comparison.

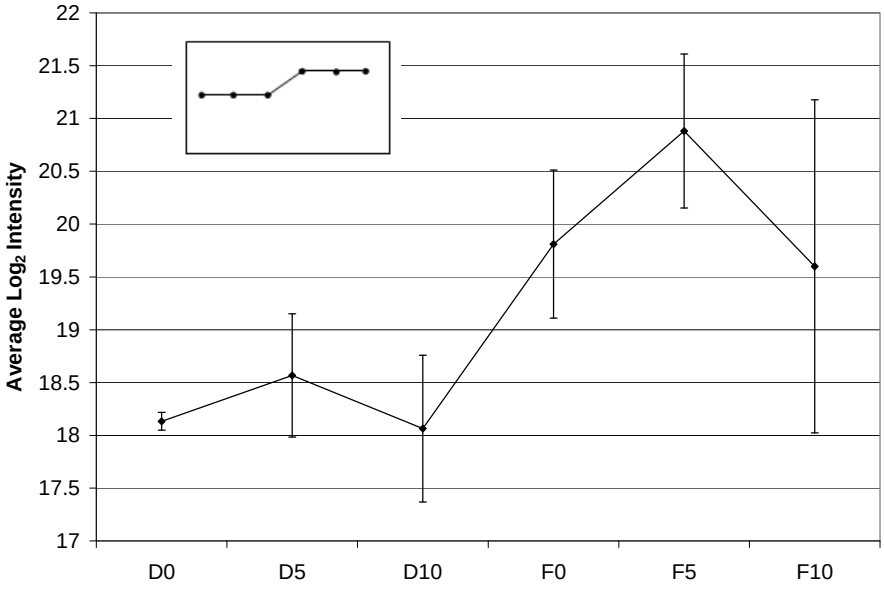


Figure 3.20 Expression profile showing relative mean intensity across all six groups of the feature identified as a peptide from Keratin 10 with mass 1108.863 Da. Insert shows predicted expression profile for comparison.

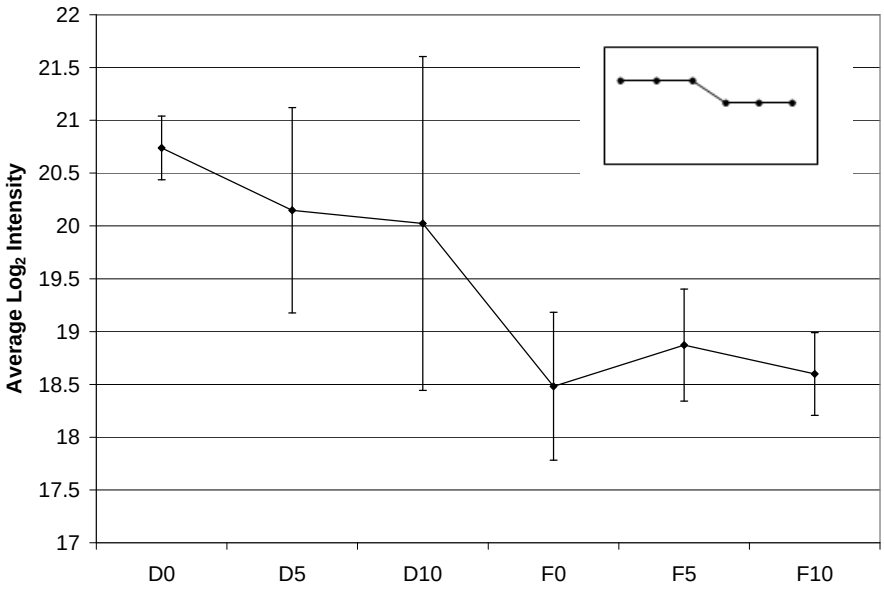


Figure 3.21 Expression profile showing relative mean intensity across all six groups of the feature identified as a peptide from Secretogranin II with mass 1675.534 Da. Insert shows predicted expression profile for comparison.

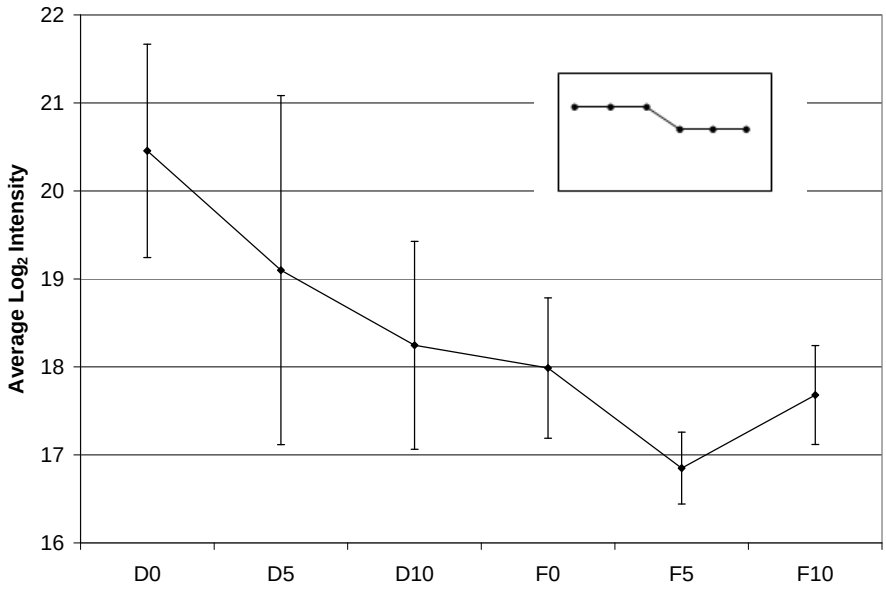


Figure 3.22 Expression profile showing relative mean intensity across all six groups of the feature identified as a peptide from Secretogranin II with mass 2468.978 Da. Insert shows predicted expression profile for comparison.

It should be noted here that the only expression profiles, from the predicted profiles, identified are as shown in figures 3.2 C, D and B.

From the top 39 features, with p-values less than 0.001, 6 show expression profile as predicted by figure 3.2 C and 5 show expression profile as predicted by figure 3.2 D. The remaining profiles show expression profile as predicted by B or are not able to be categorised.

It is therefore concluded that the heat treatment either stabilises the features identified prior to and during tissue slicing, degrades the features identified prior to and during tissue slicing or has no effect on features identified.

However, examination of the expression profiles does demonstrate that there are some overlapping error bars between sample sets. This is a result of sample variation within the sample sets analysed. This analysis would benefit from larger sample numbers to attempt to minimise the sample variation within sets.

It should, again, be noted at this point that the majority of features, not targeted for identification, show no significant effect from the heat treatment. Predicted expression profiles in figures 3.2 A was not identified in the features targeted for identification.

From this, the conclusion can be drawn that if the heat treatment is going to have a significant effect on any feature, it will occur prior to, and during tissue slicing. This is as predicted, as any enzymes which are inactivated due to the treatment, will be activated immediately following dissection and prior to slicing. A further advantage of this effect is a more efficient tryptic digestion will take place in the heat treated samples, as opposed to the snap frozen samples, as unwanted enzyme action should be minimised.

Peptides identified from the same protein show similar expression profiles as demonstrated in figures 3.16 and 3.17 giving an extra validation to the profiles.

3.4.4 LC-MS-MS for Targeted and Untargeted Protein Identification using Q-TOF-MS

3.4.4.1 Sub 10kDa Intact Sample Results

Targeted MS-MS and untargeted MS-MS was performed on an alternative mass spectrometer (Applied Biosystems Q-Star) in an attempt to identify features from the <10 kDa intact fraction. As previously discussed (in section 3.4.2.1), identification of the features from these samples proved to be very challenging. The Q-Star mass spectrometer also uses CID fragmentation, in the same way as the Bruker HCT.

Unfortunately, again, no peptides or proteins could be identified using this method. Again the spectra obtained were of very low intensity and the data is not presented here. This reinforces the conclusion that the poor quality MS-MS data is due to the fact that the non-tryptic peptides in these samples do not fragment efficiently, using CID fragmentation.

The conclusion drawn from this is that a different method of MS-MS fragmentation may be the only way to obtain identification of the features in the intact <10kDa samples used in this work. As mentioned previously in section 3.4.2.1 possible choices for this could be electron transfer dissociation or electron capture dissociation fragmentation. It was unfortunate that these methods were not available in this research; however they could provide a future investigation.

3.5 Discussion

In my research presented here a heat treatment system (Denator AB) has been evaluated in comparison with conventional snap freezing for preservation of proteomic data in tissue samples, specifically mouse brain tissue. This evaluation was undertaken using LC-MS and Label Free Relative Quantitation, using Decyder MS software from GE Healthcare. From this analysis it is concluded that the treatment has little affect on the sub 10 kDa proteomic data. This conclusion was reached, due to the fact that for over 6000 detected features in the intact sample and over 4000 detected features in the digested samples, less than 1% of features show significant changes between the groups, when an ANOVA p-value of 0.001 is set. This is discussed in sections 3.4.2.1 and 3.4.2.2.

It should be noted at this stage that the changes in intensity discussed here, relates to the relative signal intensity between groups and not the absolute intensity of the features in question. The features that do show significant changes between the sample groups showed a number of different types of expression profiles, varying from showing feature preservation when the heat treatment is used, to sample degradation when the treatment is used. An additional conclusion, as discussed in section 3.4.3.2 was that if any significant effect was observed on a a feature, it occurred prior to and during tissue slicing. This was as hypothesized, as any enzymes that are denatured by the heat treatment will be active in the untreated tissue immediately following dissection and prior to tissue slicing. An advantage of this effect is thought to be that a more efficient tryptic digestion will take place in the heat treated samples, as any unspecific enzyme action should be minimised.

It is concluded from this work that the heat treatment system may be a useful technique when particular targets cannot be detected using conventional snap freezing treatment. However, for analysis of the sub 10 kDa LC-MS proteomic data from brain tissue sample studied here, it is not expected that the Denator heat treatment will become a widely used sample preparation technique, due to the small change it demonstrates on the majority of the data.

This result can be compared with work undertaken by Goodwin *et al* (286), who studied the effects of heat treatment on proteomic samples using MALDI-MS imaging and 2D-DIGE. It was concluded in this work that rapid and controlled heat treatment of brain tissue reduces the non-specific proteomic degradation caused by endogenous enzyme action, making the Denator heat treatment system a useful process for quantitative proteomic analysis and may also increase the opportunity for tissue handling at room temperature,

such as sample dissection. However, it was also found that the heat treatment system did reduce the quality of tissue sections that could be obtained for MALDI-MS imaging. This result confirms the conclusion reached in the work presented here, that the Denator heat treatment system may prove useful in certain types of analysis for particular targets, however it may not be necessary or may even prove detrimental to other studies.

It has also been found here that MS-MS identification of non-tryptically digested protein and peptide fragments in sub 10 kDa brain tissue samples is extremely challenging. The use of trypsin for protein analysis generates peptides with a basic C-terminal amino acid (287). When these peptides are protonated and fragmentated a predominant y-ion series is formed, simplifying the interpretation of the product ion spectra. It has been reported previously (288) that fragmentation of non-tryptic peptides can result in a mixture of b and y fragment ions, which can prevent unambiguous assignment of fragment ion types. This phenomenon may impede the interpretation of the product ion spectra.

Two different types of mass spectrometer were used in this work, in an attempt to identify features of interest from these samples. However, it proved to not be possible with either of these methods. As discussed in section 3.4.2.1, this is thought to be due to poor fragmentation from non-tryptic peptides and protein fragments by collision induced dissociation. Different techniques have been developed that may help to alleviate this problem such as electron transfer dissociation fragmentation (289), electron capture dissociation fragmentation (290), charge-directed fragmentation (291) and sodiation induced tryptic peptide-like fragmentation (288). Unfortunately, these techniques were not available in this work however they could be possible future avenues of investigation.

It is also possible that the features of interest in these samples may not be peptides or protein fragments, and they may in fact be other types of molecules, such as lipids. This would explain why these features could not be identified using the proteomic MS-MS techniques employed in this research. This could potentially indicate that the heat treatment may have a more significant effect on molecules which are not peptides and proteins. This concept has not been investigated in this work, but could prove to be an interesting future project.

As this work progressed it became apparent that the widely used ANOVA statistical analysis may not be the most appropriate method to discover significant differences in datasets of this type, even though this is a widely accepted and utilised method. Further ways of discovering significant changes in datasets such as this one, especially in datasets

with multiple groups, has been further and fully investigated. This is discussed and implemented in chapter 4, entitled ‘Discovery and Identification of Biomarkers for Hypertension’.

4 Discovery and Identification of Biomarkers for Hypertension

4.1 Introduction

Systemic hypertension is one of the most common diseases in the developed world, and is responsible for many symptoms including heart failure and stroke. Hypertension is defined as persistent raised blood pressure of ≥ 140/90 mm Hg (292). Worldwide, the number of adults with hypertension by 2025 is expected to increase by up to 60 %, to a total number of 1.56 billion (293). By this time cardiovascular disease is predicted to be the most common cause of death worldwide, with hypertension being the most reversible factor (293). The risks of raised blood pressure are well documented and studied (294;295). Meta-analysis of pooled data sets has confirmed the relationship between continuous raised blood pressure and cerebrovascular disease and coronary heart disease (296). Hypertension has also been shown to be directly related to left ventricular hypertrophy, heart failure, peripheral vascular disease, renal disease and atherosclerosis (294). Hypertensive patients are also more likely to have type 2 diabetes and lowered levels of high density lipoprotein cholesterol (297), the common factor may be insulin resistance, due to high levels of co-existing hypertension and obesity (298).

In recent years there has been a greater emphasis on prevention of hypertension (299). In order to accomplish this goal it is important to be able to understand the mechanisms of the disease, as well as to establish new screening techniques which can predict the occurrence of hypertension. Essential hypertension has many causes (300), that can differ greatly from patient to patient, and not all are well understood or documented. The relationship between high levels of dietary salt and increased blood pressure has been well documented for many years (301-305). The current opinion being that a high level of dietary salt is associated with increased blood pressure in later life, and other cardiovascular diseases (301). In order to understand the reasons for this relationship, it is first important to discuss the mechanisms that actually lead to abnormal arterial constriction and elevated blood pressure. The pressure required to circulate the blood is provided by the pumping of the heart (known as cardiac output) and the tone of the arteries (known as peripheral resistance). The cardiac contraction propels the blood through the arteries, but it is the dynamic regulation of artery diameter, especially in the smaller arteries that regulates flow in the periphery and hence blood pressure. Due to the episodic ejection of blood from the

heart, pressure in the vessels has a periodic variation. These peaks and troughs of pressure refer to the systolic and diastolic pressures that can be measured intra-arterially or by a blood pressure cuff. Blood pressure is closely regulated in the body to ensure continuous perfusion of the vital organs (306).

In the short term, flow and blood pressure can be adjusted via the control of neural and humoral factors that can rapidly constrict or dilate arteries in order to meet short term circulatory demands (307). Baroreceptor reflexes then play an important role in resetting blood pressure, following acute changes, and could also have a role in long term blood pressure regulation via sympathetic nerve activity and renal Na^+ excretion in hypertension (308). Over the long term blood pressure is primarily controlled by water and salt balance, due to the infinite gain property of the kidneys to rapidly eliminate excess water and salt. If renal function is reduced this results in an increase in the extra-cellular fluid volume (ECV), inevitably causing a rise in blood pressure (309). It has long been known that total body sodium content determines extra-cellular fluid volume (ECV), via osmotic activity (302). Hypertension that results from salt retention is preceded by a rise in total plasma volume (310). In this state the elevated blood pressure promotes a pressure natriuresis, meaning that normal ECV is restored, at the expense of elevated pressure. The basis of the ECV is an isotonic salt solution, mainly consisting of NaCl. As the kidneys are the primary regulator of salt and water balance, it follows that renal function and salt are widely recognised as critical factors in the pathogenesis of hypertension. It has also been known for sometime that diuretic/natriuretic agents, such as hydrochlorothiazide, which counteract the tendency for salt and water retention, can be effective treatments for hypertension (299;311). In addition renal transplantation studies in rats and humans show that the hypertension travels with the hypertensive donor kidney (312-314).

There is substantial evidence implicating genetic causes of hypertension. Monogenic defects in renal salt transport have a clear effect on blood pressure, those that promote salt retention being associated with hypertension and conversely those that promote salt wasting are associated with hypotension (306;315;316). Additionally, mutation or knock out of genes which affect blood pressure induce either salt dependent hyper or hypotension, or unusual forms of salt independent alterations in blood pressure. These unusual forms are, in general, associated with genes that affect the synthesis and secretion of humoral vasoconstrictors or vasodilators (317). Further evidence of genetic influence on hypertension comes in the form of twin studies, showing a greater concordance of blood pressures of monozygotic than dizygotic twins (318), and population studies, showing a greater similarities of blood pressures within families, than between families (319).

In western societies the frequency of hypertension increases with age. It has been suggested that the reason for this may be that the primary cause of the disease, in many patients, is from a subtle renal injury rather than a genetic cause (320). To summarise, hypertension is a complex disease with genetic and environmental influences in which the role of the kidneys and the salt/water balance is central.

Genetic studies in hypertension, especially in humans, can be very complex, due to multifactorial determination. A variety of factors, in addition to genetic, may contribute to hypertension, including environmental and demographic. Therefore genetic study design in the general population is, at best, challenging. One possible approach is to investigate rare forms of blood pressure variation, in which a single gene mutation causes large changes. Success in this work has been able to define pathways fundamental to hypertension (306), in particular 8 genes have been identified in this way that cause hypertension in humans. Interestingly, all of these 8 genes act in the same pathway in the kidneys, which alters net renal salt reabsorption. As mentioned previously the kidneys play a vital role in maintaining the salt/water balance in the kidneys. In a normal day they filter around 170 litres of water, containing 23 moles of salt. To maintain salt homeostasis (on an average diet) the kidneys must re-absorb 99.5 % of the filtered salt. The kidneys accomplish this via a system of ion channels, exchangers and transporters, with 60 % of the filtered sodium reabsorbed in the proximal tubule of the nephron, mainly by Na^+/H^+ exchange and 30 % in the thick ascending limb of Henle by Na-K-2Cl co-transport. A further 7 % is reabsorbed by Na-Cl co-transport in the distal convoluted tubule. The final 2 % is reabsorbed through the epithelial Na^+ channel in the cortical collecting tubule. Although this final step accounts for the smallest fraction of salt reabsorption, it is also the main site where net salt balance is determined. Mutations that affect the pathway involved in the reabsorption at this site have been identified by investigation of families with extreme variations in blood pressure (either high or low), firmly establishing this pathway as central in hypertension. Increased net renal salt reabsorption causes increased water reabsorption to maintain plasma sodium concentration. This results in increased intravascular volume, increasing blood return to the heart and raising cardiac output. This then leads directly to elevated blood pressure.

Further genetic mutations involved with hypertension have been identified, including mutations affecting circulating mineralocorticoid hormones (which is the main regulator of the epithelial Na^+ channel) (321-324) and mutations affecting the mineralocorticoid receptor (325). This mutation affecting the mineralocorticoid receptor is involved with hypertension which is exacerbated in pregnancy (pre-eclampsia), which occurs in 6 % of

all pregnancies and is a major cause of maternal and foetal morbidity. Although these genetic mutations have been very insightful to certain pathways surrounding hypertension, and also tend to cause large blood pressure changes in the individuals involved, they are actually very rare. Hence, they likely account for a very small percentage of patients with hypertension in the general population.

As discussed more extensively in the main introduction, in recent years, focus has started to turn to proteomics as a tool to understand disease (1), and also to identify possible biomarkers to diagnosis and monitor disease (6). Problematically the increased importance of proteomics has not been accompanied with an improvement in analytical tools and methods, in particular if compared to tools available for the study of genomics, meaning method development in this field is as, if not more, important than the biological research itself. The theory behind biomarker discovery is in essence very simple. Body fluids pass through diseased tissue and collect biomolecules, such as peptides, that can indicate status of health. If these biomolecules can indicate the presence of a disease they are called biomarkers. A biomarker is used as a way to diagnose, classify disease and also to measure progress and therapeutic response of disease. The ideal biomarker is a substance or symptom that is present only in diseased patients and absent in healthy controls, or vice versa. In this work biomarkers are considered as biomolecules and in this respect a biomarker will often be a change in biomolecule abundance, meaning that it is important to characterise as many biomolecules as possible, with high accuracy and resolution (7).

There have been past studies which have identified possible biomarkers for hypertension, such as C-reactive protein, which is a marker for inflammation, serum aldosterone, which has a neurohormonal activity, and brain natriuretic peptide, which is an indicator of ventricular wall stress. These have been seen at elevated levels, prior to the onset of hypertension in humans (326-329). These types of studies are often limited by their focus on only one pathway, or marker. It is thought that a multi-marker process would be able to give a more accurate prediction of the occurrence of hypertension. A study conducted by Wang *et al* (330) did just that. The levels of nine previously detected biomarkers were measured in a large group of non-hypertensive individuals. From this, a set of three markers were detected to be strongly associated with the risk of development of hypertension. The nine initial possible markers investigated were C-reactive protein (inflammation), fibrinogen (inflammation and thrombosis), plasminogen activator inhibitor-1 (fibrinolytic potential), aldosterone, rennin, B-type natriuretic peptide, and N-terminal proatrial natriuretic peptide (neurohormonal activity), homocysteine (renal function and oxidant stress) and urinary albumin/creatinine ratio (glomerular endotheliar

function). These possible markers were all taken from pathways previously implicated in the pathogenesis of hypertension. The levels of these markers were measured in 1456 patients, of which 232 developed incident hypertension over three years. Of the nine potential markers, three were found to be significantly related to incident hypertension; C-reactive protein, plasminogen activator inhibitor-1 and urinary albumin/creatinine ratio. Higher levels of these three biomarkers were found to be strongly associated with the development of incident hypertension.

One of the major limitations of this work is that the biomarker discovery step is guided by previous knowledge and availability of arrays alone. As the occurrence of, and possible predisposition to, hypertension is not fully understood this may limit possible biomarker discovery significantly.

An unguided biomarker discovery process has been utilized and reported by Kiga *et al* (40) to identify haptoglobin as a biomarker for stroke. In this work changes in the expression of plasma proteins in spontaneously hypertensive stroke-prone rats were investigated using surface-enhanced laser desorption/ionization time-of-flight mass spectrometry. The study found that decreased levels of haptoglobin were seen in the diseased rats compared with the wild type controls. This finding was validated using the traditional biochemistry technique of western blot analysis. This result shows that unguided biomarker discovery can be used in a meaningful way to identify biomarkers and learn more about disease.

Unguided methods have also been used and reported for other applications, such as detection of preeclampsia (331). In this work by Carty *et al* urine samples taken from women throughout pregnancy. The samples were analysed using capillary electrophoresis coupled to time-of-flight mass spectrometry and disease specific peptide patterns were then generated using support vector machine-based software. Potential biomarkers were then identified using liquid chromatography and tandem mass spectrometry. A model of 50 biomarkers was developed which could be shown to predict preeclampsia from gestational week 28, however it was noted that further validations were required.

Silbeger *et al* (41) have reported the use of surface enhanced laser desorption/deionisation time of flight mass spectrometry in an unguided manner to identify differentially expressed proteins in human myocardial infarction plasma samples. In this case statistical analysis was used to identify potential biomarkers for early stage acute myocardial infarction, however it was noted that the sample size used was rather small and further validation would be required.

Dayarathna *et al* (332) have reported the use of an unguided biomarker discovery technique to quantify differences in plasma samples between several disease groups; obesity, diabetes and hypertension. In this work human plasma samples from the three disease states and healthy controls were analysed using liquid chromatography and tandem mass spectrometry. In addition, the use of depletion of abundant proteins was used (this method is dicussed in more detail in the following pages). Dayarathna *et al* were able to identify changes in the level of several proteins. Angiotensinogen was found to be present at high levels in patients with obesity, diabetes and hypertension. Apolipoprotein CI was shown to be elevated in all disease groups. These findings were supported by a review of the literature.

It is proposed here that biomarkers for hypertension are currently still not particularly well characterised. In the work presented here biomarkers for hypertension are investigated in a completely unguided manner. A label free relative quantitation method is used to examine differences between diseased and healthy samples, mainly concentrating on plasma samples. In this instance, no previous knowledge of the disease is used, meaning that it is possible to discover entirely new markers, which may come from unknown functions of hypertension. This not only enables biomarker discovery for early detection of disease, but also a new method for understanding hypertension and identifying new disease pathologies. This unguided biomarker discovery method could be applied to any disease state in the future, without the requirement of prior knowledge of the disease in question.

To perform these studies a rat model is used. The model systems are a Stroke Prone Spontaneously Hypertensive Rat (SHRSP), Wistar-Kyoto wild type (WKY) and a congenic strain showing intermediate symptoms. The effect of salt treatment on the peptides in this study is also investigated, as salt/water balance and renal function is so pivotal to the mechanisms surrounding hypertension. In this model, a major technique used for identifying genetic determinants which underlie disease is the identification of quantitative trait loci (QTL). This is then followed by the development of the congenic and sub-strains in order to evaluate candidate genes (333).

Genetic models used were as follows, and as pictured in figure 4.1:

- SHRSP – Spontaneously Hypertensive Rat Stroke Prone
- Wistar Kyoto Rat (wild type)
- Congenic Strain – SP.WKYGla2k

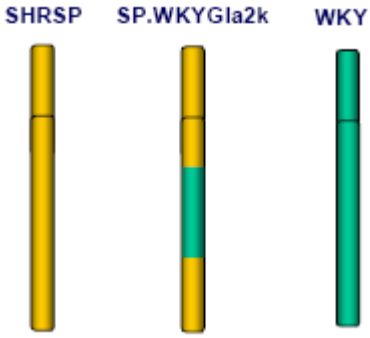


Figure 4.1 Schematic demonstrating genetic models used in this study. The genetic models used are SHRSP, WKY and a congenic strain where WKY segments are introgressed into the SHRSP strain genetic background. The congenic strain is discussed in more detail in section 4.3.1.

The SHRSP strain is a well characterised model for essential hypertension. It is known to show left ventricular hypertrophy, endothelial dysfunction and salt sensitivity (334-336). The WKY rat model has a contrasting phenotype, and demonstrates less of a blood pressure response to salt treatment (334). In a study by Clark *et al* (334) quantitative trait loci were identified for baseline and salt-sensitive blood pressure on rat chromosome 2. Following this, these quantitative trait loci were confirmed, using reciprocal chromosome 2 congenic strains (336;337). Studies combining congenic mapping and microarray analysis identified the functional candidate gene *Gstm1* (337;338). This gene is involved in the endogenous defence against oxidative stress. A combined strategy of congenic substrain production and microarray expression profiling has identified two positional candidate genes (Edg1 and Vcam1) for salt sensitive hypertension in the SHRSP model (333).

Congenic strains are models where the chromosomal region containing the quantitative trait loci of interest in one strain (recipient) has been replaced by the homologous region from a second strain (donor). If the blood pressure of the congenic strain is significantly different from the recipient strain, it can be inferred that this chromosomal fragment does possess a quantitative trait loci which contributes to a difference in blood pressure between the donor and recipient strains. The traditional way to produce congenic strains is to

serially backcross the donor and recipient strain, at every generation selecting for progeny heterozygous for the desired chromosomal region. Mendelian principles dictate that 8 to 12 backcrosses are necessary, in order to ensure that > 99 % of the donor strain's genetic background is replaced by that of the recipient strain. After these backcrosses brother/sister mating is used to make the desired chromosomal region homozygous for the donor strain's alleles (336). The samples used in this work are created using a 'speed' congenic breeding strategy. The theory of this strategy is based on calculations by Lander and Schork (339) which showed that the congenic strain production can be speeded up, using repeated screening of polymorphic marker loci distributed throughout the genetic background, allowing specific selection of a male from each backcross with the fewest amount of donor alleles remaining in the genetic background. Breeding these males would then allow the rate of background donor elimination to be accelerated. This, in turn, reduces the number of generations necessary for creation of a congenic strain.

In my research presented here, an animal model is used for biomarker discovery which can be problematic. The differences between a study performed on a set of human samples and animal samples are multiple. One of the reasons for choosing to use an animal model is to minimise the sample heterogeneity. Human samples may differ in many ways that cannot be controlled and that can significantly affect the outcome of a study. For instance lifestyle choices such as diet and exercise and underlying genetic differences can cause large differences between individuals. In contrast, conditions surrounding animals can be tightly controlled and maintained. This should help to minimise the sample variance, meaning that the disease being studied should cause any significant differences between sample groups (340). Clearly, experimental design will be simpler when using an animal model, as the sample sets are clearly defined from the start of the experiment, as opposed to in an 'observational' human study, in which a large group is monitored over a significant time period. Additionally, an animal model study can be conducted over a much shorter period of time. It must, however, be considered that using an animal model for biomarker discovery does not give a true view of the normal sample variation between individuals. It is therefore often used as a precursor to a wider, human study.

However, experimental design must still be carefully considered even when using an animal model for a biomarker discovery study, to avoid bias in the results. Certain threats to a valid study have been reported (340) as being 'chance', 'generalizability' and 'bias'. Chance refers to a type I error, otherwise known as a false positive result, or a type II error, which is known as a false negative result. These 'chance' errors cannot be completely avoided, but can be minimised by increasing sample size. A further chance based error,

more specific to the discovery based '-omics' type research, is known as over-fitting (341). Over-fitting can occur when multivariate models are made to fit a data set, in order to discriminate one set from another. Sometimes a pattern will discriminate between the two sample sets purely by chance, and will not be reproducible in any other sample sets. To avoid this error, a validation sample set should be used in order to assess reproducibility of the results. The second type of possible error is 'generalizability', and is also known as 'external validity' (340). This refers to who the results of the study would apply to. For instance, if an animal model is used initially, the results of the study may not be applicable to humans with the same disease. This is perfectly acceptable as a proof of principle study, and it would be expected that positive results could then be moved forward to a more expensive, larger study in humans, giving the results a wider generalizability. It has been proposed that this step-wise method for biomarker discovery is the most efficient and cost-effective way of conducting such a study (342). Bias is a more difficult error to monitor, as it is usually unintentional and unconscious. It can be defined as 'the systematic erroneous association of some characteristic with a group in a way that distorts a comparison with another group' (340). The most important step to avoid bias in a study is to keep every possible factor throughout the study as equal as possible, during the design, conduct and also the interpretation of results. In addition, randomisation of the order of sample analysis is important to help avoid the introduction of bias into the study.

In my research presented here, plasma is the sample of choice. This further complicates experimental design as plasma can be a difficult sample to analyse. Plasma is an interesting biofluid, in many senses. It is complex, containing a wide dynamic range of proteins, spanning over ten orders of magnitude. The complexity of plasma samples is even more problematic when subjected to LC-MS analysis, as the samples will often be protease digested, adding an additional level of complexity to the samples. This makes the high dynamic range of proteins even more challenging to cope with experimentally. Plasma also contains 'leakage proteins' from tissue, meaning as well as the classical plasma proteome, it also contains a sub-set of tissue proteomes. It is also notoriously difficult to analyse due to the high percentage presence of a single protein, albumin (343). One of the greatest challenges in biomarker discovery from plasma is identifying these low abundance proteins, that make up less than 1 % of the entire plasma proteome, whereas the 20 most abundant plasma proteins make up the remaining 99 % of the proteome (344;345). Biomarker discovery can be assisted by removal of these highly abundant, well characterised proteins from biological solutions (346). Potential markers and interesting proteins may be masked by the presence of this small number of highly abundant proteins. Protein depletion has been used in the past to remove albumin from plasma for biomarker

discovery (347;348), but the removal of more than simply albumin may be desirable. There are a number of different strategies available to the researcher for plasma depletion, also known as protein pre-fractionation. Removal of the abundant proteins is probably the most common technique used. There are different depletion systems available, and they are mostly centred around antibody affinity. Depletion systems that remove one (349), two (350-352), three (353), six (350;351;354-356), seven (357), 12 (356;358;359), 14 (360), 20 (361) and 58 (362) abundant proteins have been described. Alternatively, ProteoMiner beads is a new technology available for plasma depletion (363). They use a library of hexa-peptides coupled to beads, with each bead being present in equal numbers. As proteins bind to the peptide ligands, proteins are retained on the beads. As there is a limited number of each bead, only a limited number of each protein will be retained, reducing the high dynamic range of proteins present without removing any specific protein. The disadvantage of this system is that protein concentration differences between samples may be altered, making downstream quantitation difficult. A further strategy is as an alternative to removing abundant proteins, is to attempt to enrich for interesting proteins. This is done, for example, by enriching for carbonylated proteins. Reactive oxygen species (ROS) are well known for their role in diseases such as cancer, heart disease and Alzheimer's (364-366). Exposure to ROS can cause irreversible oxidation of amino acid side chains, which introduces aldehyde or ketone groups (367). Biotinylation of oxidised proteins with biotin hydrazide and affinity selection with monomeric avidin affinity chromatography can be used to enrich carbonylated proteins from plasma. They may then be analysed using LC-MS-MS. It was shown (368) that when this enrichment was performed 92 % of the 146 proteins identified were tissue derived, and not classical plasma proteins, showing the potential of this method for biomarker discovery. However, as the enrichment is for a specific modification, the method is somewhat limited. Other types of pre-fractionation used are peptide based. In general this is used to reduce complexity caused by digestion, or can also be used to target and enrich sub-proteomes within the plasma. This can take the form of cysteine containing peptide enrichment (369). This has been shown to have no positive effect on the number of proteins identified (351), suggested to be due to the fact that 95 % of cysteinyl peptides in human plasma are from two abundant proteins, albumin and transferrin.

Combined protein/peptide pre-fractionation can also be used. Around 50 % of all proteins are thought to be glycosylated (370), and glycoproteins secreted into plasma are a main part of the plasma proteome. As protein glycosylation is known to play an important role in disease, such as cancer (371) these can be an interesting set of proteins to analyse. Enrichment for glycoproteins is mainly based around lectin affinity (372), hydrazide

affinity (373) or other carbohydrate affinity (374;375). These techniques both reduce the complexity of plasma by looking at a sub-proteome and also reduce the additional complexity caused by digestion, meaning the end enriched product is only the glycosylated peptides from the glycoproteins.

Metal affinity based fractionation is used quite frequently in proteomics (376), however to date immobilized metal affinity chromatography has only been used in a limited number of studies with plasma or serum samples, and these are primarily with MALDI-TOF-MS (377-379).

It must, however, be considered that plasma depletion or pre-fractionation for biomarker discovery may not always be advantageous. Additional sample processing steps can cause sample loss and poor sample reproducibility. In addition, removal of abundant proteins may cause further loss of less known and low abundance associated proteins and peptides (380) . Another disadvantage of plasma depletion is the high cost. This can be around £12 for each 15 µl sample, as for the spin columns evaluated in my research, making it an expensive step in sample preparation, especially for the high number of samples that must be analysed for biomarker discovery. In this work Beckman Coulter plasma partitioning spin columns are evaluated, this is discussed further in section 4.3.2. These are based on antibody affinity capture, and remove the seven most abundant proteins from rat plasma.

4.2 Aims

The main aims in this chapter are as follows:

- Optimisation of sample preparation of plasma samples for LC-MS analysis
- Determination of protein concentration in the rat plasma samples used
- Evaluation of plasma partitioning spin columns using 2D mini gels and LC-MS analysis
- LC-MS analysis of plasma samples
- Evaluation of two types of label free quantitation software
- Label free relative quantitation of features within plasma sample as a possible biomarker discovery method
- Development of a novel statistical analysis for the label free relative quantitation dataset
- Targeted MS-MS analysis for identification of any features of interest
- Western blots as validation of any biomarkers discovered

4.3 Methods

4.3.1 Sample Preparation

Samples for this study were provided by Dr Martin McBride of the Institute of Cardiovascular and Medical Sciences, Glasgow University. Inbred colonies of SHRSP and WKY rats have been kept at the Cardiovascular Research Group, in the University of Glasgow since 1991. Strain SP.WKYGla2a and SPWKYGLA2c* are generated using the marker-assisted 'speed' congenic strategy, where WKY segments are introgressed into the SHRSP strain genetic background (339). The strains are named with the first abbreviation belonging to the recipient strain, and the second abbreviation belonging to the donor strain. Gla indicates that the strain originates from Glasgow colonies, and the number 2 refers to rat chromosome 2. Congenic substrains are generated by backcrossing male SP.WKYGla2a congenic rats with SHRSP females. Progeny generated from this are heterozygous throughout the congenic interval. Brother/sister mating is carried out to generate and fix substrains, which contain smaller congenic intervals. These include SP.WKYGla2b, SP.WKYGla2e, SP.WKYGla2f, SP.WKYGla2g, SP.WKYGla2i and SP.WKYGla2k. In this research the substrain used is the SP.WKYGla2k. A genetic map of the congenic strains is shown in figure 4.2 (taken from Graham *et al* (333)).

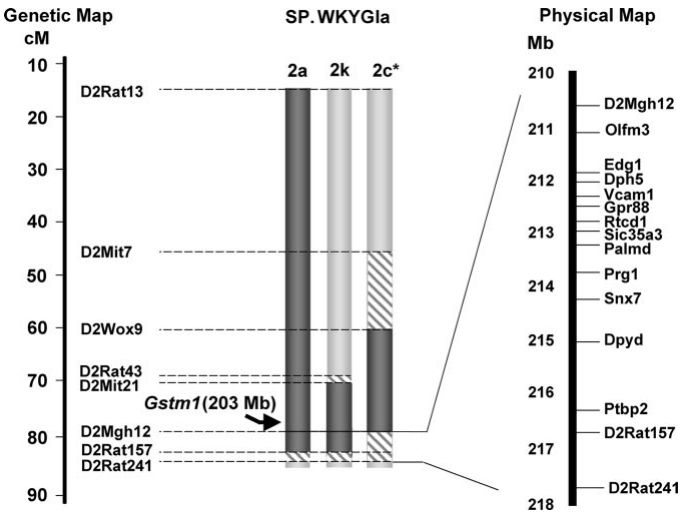


Figure 4.2 Comparison of chromosome 2 genetic map among SP.WKYGla2a, SP.WKYGla2k, and SP.WKYGla2c* strains. Dark gray bars indicate regions of WKY homozygosity. Lighter gray bars represent regions of SHRSP homozygosity. Hatched bars represent regions of recombination and heterozygosity. To the right, a physical map of an _6-Mb region, which delineates the different lower boundary markers between the SP.WKYGla2c* strain and the SP.WKYGla2a or SP.WKYGla2k strains, is expanded to illustrate a selection of genes contained within this congenic interval. The location of Gstm1 is indicated outside of the implicated region. From Graham *et al* (333).

From weaning the rats were fed on normal rat chow. All rats were euthanized at 21 weeks old. A salt treatment was carried out on a proportion of the animals. This involved a 1 % NaCl solution as drinking water from 18 weeks onwards. A total of 38 rat plasma samples were used in this work, shown in table 4.1.

Prior to euthanasia blood was obtained by resection at the end of the rat tail under anesthesia and collected into EDTA-tubes. Plasma was separated by centrifugation (3000g, 10min) and then stored at minus 80ºC until further analysis. The rearing of rats and collection of plasma was undertaken by Dr Martin McBride (Institute of Cardiovascular and Medical Sciences, University of Glasgow).

From this point forwards the congenic SP.WKYGla2k strain is referred to simply as Congenic Strain. Blood pressure was measured using the Dataquest IV telemetry system, implanted at 12 weeks of age. This directly measures systolic blood pressure. An example of blood pressure differences recorded by the Cardiovascular Research Group at Glasgow University is shown in figure 4.3 (taken from Graham *et al* (333)).

WKY no salt	WKY salt	SHRSP no salt	SHRSP salt	Congenic (SP.WKYGla2k) no salt	Congenic (SP.WKYGla2k) salt
A4659	A4631	C5952	C5847	N6690	N6686
A4660	A4637	C5953	C5848	N6692	N6687
A4695	A4803	C5989	C5849	N7036	N6688
A4696		C6157	C5850	N7037	N6718
A4792		C6169	C5851		N6839
A4798			C5990		N6840
A4800			C5991		N6841
A4802			C5992		N6842
			C6155		
			C6156		

Table 4.1 Hypertension study sample numbers

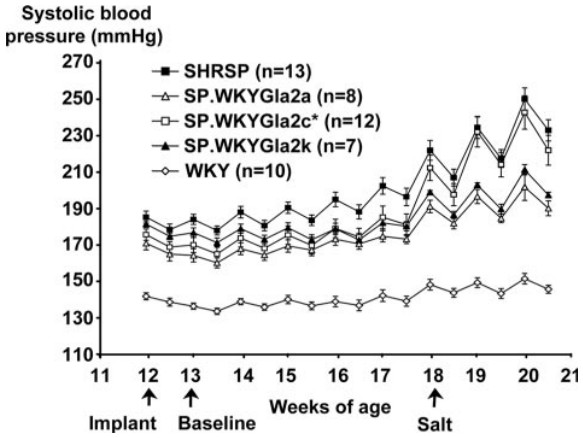


Figure 4.3 This figure shows an example of blood pressure levels in the strains of rat used in this work and was produced by the Cardiovascular Research Group at Glasgow University. Systolic blood pressure measured by radiotelemetry in male SHRSP (n_13), SP.WKYGla2a (n_8), SP.WKYGla2c* (n_12), and SP.WKYGla2k (n_7) strains. Baseline systolic blood pressure was significantly reduced in SP.WKYGla2a (F_12.98; P_0.002), SP.WKYGla2c* (F_6.35; P_0.021), SP.WKYGla2k (F_4.47; P_0.049), and WKY (F_130.8; P_0.00001) strains vs SHRSP (repeated-measures ANOVA). Systolic blood pressure was significantly lower in WKY rats vs SHRSPs during the salt-loading period (F_110.3; P_0.00001). Salt-loaded systolic blood pressure was significantly and equivalently reduced in SP.WKYGla2a (F_18.82; P_0.0001) and SP.WKYGla2k (F_11.06; P_0.004) strains but was not significantly reduced in SP.WKYGla2c* (F_1.94; P_0.179; repeated-measures ANOVA) vs SHRSP strains. From Graham *et al* (333).

All plasma samples were aliqouted into 50 µl portions and stored at minus 80ºC until required for analysis. For initial work plasma samples were treated with protein de-salting spin columns (Pierce, product number 89849) as described by the manufacturer's protocol. All plasma samples were tryptically digested for LC-MS analysis, using a trypsin to protein ratio of 1:100 by weight. Sequencing grade porcine trypsin (Promega, as used in chapter 3) was dissolved in 25 mM ammonium bicarbonate, added to plasma samples and left to digest overnight at 37 ºC.

4.3.2 Determination of Plasma Protein Concentration

Concentration of plasma samples was determined using a standard protein concentration assay, as used in chapter 2 (Bradford Assay). 1 µl of each plasma sample was diluted to 100 µl with distilled water. 10 µl of this was then used for protein concentration determination. BSA standards were used with protein concentrations of 0.0625, 0.125, 0.25, 0.375, 0.5 and 0.75 mg/ml. The absorbtivity measurement of each standard and sample was repeated three times, and the mean signal was calculated. A standard curve

was created from the absorbtivity and concentration of the known standards. The concentration of plasma samples was then determined using this curve.

4.3.3 Plasma Partitioning Spin Columns – 2D mini gels and LC-MS analysis

Plasma partitioning spin columns were purchased from Beckman Coulter (ProteomeLab IgY Spin Columns). The effectiveness of this process was evaluated using 2D mini-gel electrophoresis along with LC-MS analysis. The columns are designed to deplete the seven most abundant proteins from rat plasma samples via an affinity capture method. These proteins are albumin, IgG, Transferrin, Fibrinogen, IgM, a1-antitrypsin and haptoglobin. Plasma samples were depleted in 15 µl portions, following the protocol given by the manufacturer. This effectively partitions the plasma sample into the abundant protein fraction and the remaining plasma fraction, with additional wash fractions. As the 'depleted' proteins are recovered and available for analysis, the protocol is referred to as partitioning rather than depletion.

This affinity capture plasma partitioning was compared with a simpler, size fractionation of plasma samples. The size separation was performed using Microcon Nominal Molecular Weight Filters, 30,000 Da Microcon (Millipore, YM30). These filters separate the plasma sample into 2 fractions, based on the molecular weight of the component molecules, in this case less than, and above 30kDa. Again, these were used as directed by the manufacturer.

Each of the fractions from the plasma partitioning spin column and the Microcon Nominal Molecular Weight Filter fractionation, as well as an untreated plasma sample was separated using a 2D mini gel, 4-12 % gradient (Invitrogen) with a pH 4-7 isoelectric focussing strip. One plasma sample was used only, from a WKY rat (A4696). For the untreated plasma sample 3 µl of plasma was diluted to 20 µl with water. For the partitioned sample fractions 15 µl of each fraction (from an original plasma sample of 15 µl) was diluted to 20 µl with water. For the size filtered samples, 15 µl of untreated plasma was size fractionated and the resulting fractions diluted to 20 µl with water.

Each of these 20 µl samples was then diluted to 125 µl by addition of rehydration buffer. This consisted of Urea (7 M), Thiourea (2 M), CHAPS (4 %), bromophenol blue (0.002 %), DTT (100 mM) and IPG buffer (0.5 %). The samples were vortexed and each was layered onto a separate ceramic focussing strip holder. The plastic backing was removed from the IPG strips (pH 4-7) (GE Healthcare), and these were then placed, gel side down,

on top of the samples. Each sample and strip was then overlaid with 1 ml of oil. The samples were then left to focus overnight on an IPG machine, until 20,000 Vhrs were accumulated with program 30 V step and hold for 12 hours, 300 V step and hold for 2 hours, 1000 V gradient for 2 hours, 8000 V gradient for 5 hours, 8000 V step and hold for 8 hours and 1000 V step and hold for 24 hours.. The strips were then removed and placed in plastic tubes, with the plastic backing against a wall. The strips were washed briefly in running buffer, consisting of 6.06 g Tris, 28.8 g Glycine, 4 g SDS in 2 litres of water. The strips were then covered with 10 ml of equilibration buffer, with the addition of 100 mg DTT, and left to rock for 15 minutes. Equilibration buffer consisted of Tris HCl (50 mM), Urea (6 M), Glycerol (30 % v/v), SDS (2 % w/v) and bromophenol blue (0.002 % w/v). The strips were removed and washed in running buffer. They were again covered with 10 ml of equilibration buffer, with the addition of 250 mg of iodoacetamide, and left to rock for 15 minutes. The strips were then applied to the pre-cast 4-12 % gradient polyacrylamide gels (Invitrogen), with the plastic backing in contact with the backplate, and sealed with agarose gel. The gel tanks were filled with running buffer (as previous). A voltage was applied to the gels of 200 V for 2 hours to separate the samples. Finally, the gels were stained using a coomassie blue stain, scanned on the G-Box scanner from Syngene and evaluated visually.

The same samples were also all tryptically digested and analysed using a Dionex 3000 Nano-flow Liquid Chromatography (LC) system and a Bruker High Capacity Trap (HCT). The LC separation was performed on a 15 cm C18 PepMap reverse phase column with a 75 µm internal diameter, from Dionex. This was stored in a 37 ºC oven for the duration of the separation. Solvent A was a 2 % acetonitrile, 0.1 % formic acid solution and solvent B was an 80 % acetonitrile, 0.1 % formic acid. The loading solution used was a 2 % acetonitrile, 0.5 % tri-fluoroacetic acid solution. The loading flow, onto the trap was maintained at 30 µl/min. The column flow was maintained at 0.3 µl/min. The sample was loaded onto the pre-column trap and washed for 15 minutes, before the valves were switched to allow the sample to flow onto the column. Initially the ratio of solvents was 95 % solvent A and 5 % solvent B. Between 15 and 70 minutes, this ratio was increased to 50 % of each solvent. At 70.1 minutes, the ratio was increased to 90 % solvent B, and maintained there till 75 minutes. This was then decreased to 5 % at 80 minutes, and maintained there till 90 minutes. This gives a total LC method of 90 minutes. The mass spectrometer was set in standard (enhanced) mode, with positive polarity, scanning from m/z 100 – 2000, a maximum accumulation time of 100 ms and 8 average spectra.

The MS-MS data was converted to an mgf file in Data analysis (Bruker). It was then searched using MASCOT version 2.1, against database NCBInr 20090424 (8300899 sequences; 2858238196 residues).

MASCOT search settings were as follows:

Species: Rodentia, Enzyme: Trypsin, Report top hits: Auto, Variable modifications: Oxidation of Methionine, Charge: 1+, 2+ and 3+, Peptide tolerance: 1.2 Da, MS/MS tolerance: 0.8 Da, Instrument: ESI-Trap

4.3.4 LC-MS of Plasma samples for Label Free Quantitation using MicroTOFq, HCT and Q-Star

Initially, plasma samples were analysed using a Dionex Ultimate Nano-flow Liquid Chromatography system and an Applied Biosystems Q-Star. Repeatability and reproducibility of this system was evaluated using standard BSA digest samples and plasma samples. Base peak chromatograms were created for each sample, and overlaid for comparisons. This was repeated for:

- the same sample analysed on the same day
- the same sample analysed on different days
- different samples, from the same sample set, analysed on the same days

The LC method used was identical to that in the previous section, 4.3.3, although the 15 minute de-salting wash on the trap was not included in the method. MS data was collected from m/z 150-2000, in positive ion mode, with an accumulation time of 3 seconds.

Plasma samples were also analysed using a Dionex 3000 Nano-flow Liquid Chromatography system and a Bruker High Capacity Trap (HCT). Again repeatability and reproducibility of this system were evaluated using standard BSA digest samples and plasma samples. Base peak chromatograms were created for each sample, and overlaid for comparisons. This was repeated, as described previously, for:

- the same sample analysed on the same day

- the same sample analysed on different days

- different samples, from the same sample set, analysed on the same days

Plasma sample de-salting was also evaluated. A comparison was made of the use of previously mentioned (section 4.3.3) Pierce de-salting spin columns, and an automated de-salting step, incorporated into the liquid chromatography. The LC method used was identical to that in the previous section, 4.3.3, although initially the 15 minute de-salting wash on the trap was not included in the method. The mass spectrometer was set in standard (enhanced) mode, with positive polarity, scanning from m/z 100 – 2000, a maximum accumulation time of 100 ms and 8 average spectra.

For the label free quantitation analysis plasma samples were analysed using the same aforementioned Dionex 3000 Nano-flow Liquid Chromatography system, coupled to a Bruker MicroTOFq. Every plasma sample was analysed, using the automated de-salting step. The LC method used was identical to that in the previous section, 4.3.3.

The sample order (of all 38 samples) was randomised, to help prevent bias (as discussed in the introduction section 4.1) in the results. MS data, only, was collected on the microTOFq. Mass range used was m/z 150-3000. Each MS data file was converted to an 'mzxml' file which is compatible for import into label free quantitation software.

4.3.5 Label Free Relative Quantitation

Label free relative quantitation, as discussed in the main introduction, is a technique used to identify differences between samples. In this study two different types of software are evaluated for this purpose. These are Progenesis LC-MS from Non-Linear Dynamics and Decyder MS from GE Healthcare. The outcome of this is discussed in the results section 4.4.4.

Each MS data file was input into the label free quantitation software for analysis, therefore for each feature there exists a m/z ratio, a retention time and an intensity signal. This information is used to create a 3 dimensional map for each sample.

Peptide detection settings were optimised for the particular dataset used here, in order to minimise false positive and false negative detections. Settings for Peptide Detection in Decyder MS were as follows:

Typical peak width: 0.5 minutes, TOF, resolution: 8000, Charge states: 1-10, Uniform Background model type, Signal-to-background detection threshold: 4.0, Background subtracted quantitation, Enable charge assignment where possible, Charge assignment from two peaks: Limited, LC peak shape tolerance: 20 % , m/z shift tolerance: 0 u, m/z shape tolerance: 5 %, Remove peptides below S/N of 4.0, Remove peptides of unspecified charge below S/N of 4.0, Remove peptides with low quality LC peaks, Remove overlapping peptides, Spectrum vs model tolerance: 15 %

Settings for spectrum alignment in Decyder MS were as follows:

Max/stretch compress: 2, Max leader and trailer: 10 %, Stretch compress penalty: 0.1

Settings for peptide matching in Decyder MS were as follows:

Time tolerance: 1 minute, Mass tolerance: 0.5 Da, Use aligned retention times, Cross detection: Add peptides to intensity maps if present in 1 map

Normalisation was performed using a measured intensity distribution, as in chapter 3. This means that the entire peptide population is used for normalisation. It is assumed that most samples contain peptides that do not vary in intensity and the normalisation is performed without any further specifications. This is the recommended method of normalisation, by the maufacturers of the software, unless very few peptides are present in the samples.

4.3.6 Targeted MS-MS for Protein Identification

Following identification of features of interest, these were targeted for identification using an MS-MS method. A Dionex 3000 Nano-flow Liquid Chromatography system was used, in conjunction with an Applied Biosystems Q-Star. The targeted MS-MS was performed in two ways, initially targeting the features of interest via their masses alone, known as the ‘include list strategy’, and secondly using retention time windows in conjunction with the feature mass, known as the ‘retention time strategy’. The resulting MS-MS data was then searched against the MASCOT database for possible identifications.

The LC method used for this experiment was identical to that used in the initial MS experiment.

Settings on the Q-Star for the Include List Strategy were as follows:

Acquisition time: 90 minutes, Accumulation time: 3 seconds, TOF masses: 200 – 1500 m/z, Polarity: Positive, Cycles: 360, Cycle time: 15 seconds, Product ion masses: m/z 50 – 2000

The masses for the include lists were taken from the results of the statistical analysis for label free relative quantitation. Retention times were not specified. Samples analysed included one from each sample group.

Settings on the Q-Star for the retention time strategy were the same as above, with the exception that the product ion scans were divided into 10 minute windows with 4 product ion masses in each. The method was set to scan through each mass in turn.

All MS-MS data was searched using MASCOT version 2.1, against NCBInr database 20090424 (8300899 sequences; 2858238196 residues). The search settings were as follows:

Species: Rodentia, Enzyme: Trypsin, Report top hits: Auto, Variable modifications: Oxidation of Methionine, Charge: 1+, 2+ and 3+, Peptide tolerance: 1.2 Da, MS/MS tolerance: 0.8 Da, Instrument: ESI-TOF

4.3.7 Western Blots as a Validation Technique

In order to confirm the quantitation and identification achieved by mass spectrometry and label free quantitation western blots were used as a validation technique for two of the proteins identified.

For the western blot analysis plasma samples from each group were pooled together. 50 µl of each plasma sample was taken and combined with the relevant samples, giving 6 samples WKY no salt, WKY salt, Congenic no salt, Congenic salt, SHRSP no salt, SHRSP salt. From each samples 1 µl of plasma was taken, and diluted to 10 µl with water. 1 µl of this diluted sample was then made up to 10 µl with protein loading buffer. The plasma samples were heated to 99 ºC for 5 minutes, to denature the protein. The samples, along with a protein molecular weight ladder, were loaded onto a 4-12 % pre-cast polyacrylamide gel, from Invitrogen. This was replicated 4 times, to allow for 2 western blots, and two coomassie stained gels. The gels were immersed in pre-mixed SDS running buffer, also from Invitrogen. A voltage of 200 V, and the samples were left to separate on the gel for 2 hours. Following the separation, the gels were extracted from the plastic

casing, and the top part of the gel cut away. Two of the gels were immersed in Coomassie stain for 2 hours, followed by Coomassie de-stain for 30 mins, or until bands were visible. The gels were then scanned and visually evaluated. The second two gels were retained for western blot transfer. These gels were immersed in transfer buffer for 15 minutes at room temperature. The gel size was measured, and filter papers and western blot membrane cut to the same size (1 piece of membrane per gel, and 6 pieces of filter paper per gel). The filter paper was soaked in transfer buffer for 5 minutes. The membrane was wet in methanol for 15 seconds, soaked in water for 2 minutes, and then in transfer buffer for 5 minutes. The transfer stack was assembled in the Semi-Phor western blot apparatus. This consisted of 3 pieces of filter paper, the western blot membrane, gel and 3 further pieces of filter paper. 27 mA of current was applied, per gel, for 1 hour. The membranes were then removed and left to air dry overnight. The gels were discarded at this stage. The membranes were wet in methanol and then immersed in blocking solution for 1 hour. Blocking solution consisted of 100 ml PBS with 5 g of dried milk powder. The membranes were then removed and washed in PBS with 0.05 % Tween. The membranes were then each immersed in the primary antibody wash for 1 hour, consisting of 20 ml PBS with 0.05 % Tween, 5 % milk and 10 µl primary antibody. The membranes were then washed 6 times, for 5 minutes each time, in PBS with 0.05 % Tween, and immersed in secondary antibody solution for 1 hour. Secondary antibody solution consisted of of 20 ml PBS with 0.05 % Tween, 5 % milk and 10 µl secondary antibody. The six 5 minute washes were repeated. Each membrane was then covered in Pierce ECL reagent for 60 seconds. Membranes were dabbed dry and exposed and scanned using the G-Box scanner, from Syngene. Five 30 second exposures were used initially, followed by five 2 minute exposures. The primary antibodies used were chicken anti-rat hemopexin and chicken anti-rat fibrinogen alpha-1-isoform. The secondary antibody used was chicken IgY horse-radish peroxidase conjugate.

Quantitative information was extracted from the western blots by measuring the total raw pixel volume from each band. This was background corrected by subtracting an identical total pixel volume, taken from a background region of the blot. The data was normalised against the total pixel volume intensities from one band, taken from each sample, from the appropriate coomassie blue stained gel. These were then plotted in an expression profile style, for comparison with the expression profiles obtained from the label free relative quantitation experiment.

4.4 Results

4.4.1 Protein Concentration in Plasma Samples

Protein concentration was determined in all plasma samples. The results from the protein concentration assay of all rat plasma samples are shown in figure 4.4 below. The overall average protein concentration is 55.2 mg/ml with a standard deviation of 8.3 mg/ml.

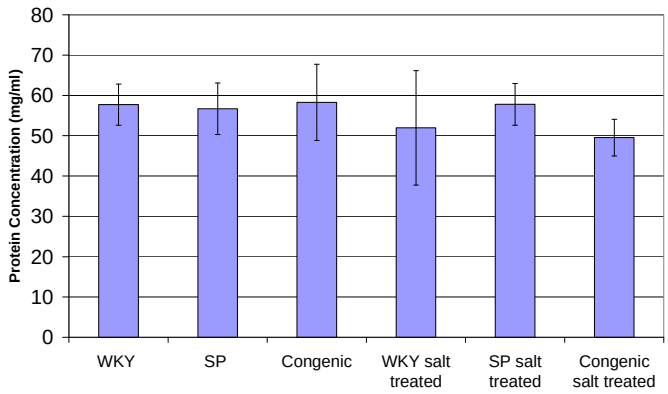


Figure 4.4 Results of protein concentration assay, across all plasma samples. The mean protein concentration within the labelled group is shown. Error bars represent ± one standard deviation.

As expected there is a high concentration of protein in the plasma samples. As previously mentioned, in section 4.1, plasma contains a wide dynamic range of proteins. Depletion of the most abundant proteins should be considered, in order to investigate the low concentration proteins which may be masked. Otherwise, the plasma samples must be diluted significantly prior to LC-MS analysis. As mentioned in the main introduction (section 1.4) nano-LC columns are easily blocked and overloaded, and the concentration of the plasma samples must be considered carefully prior to analysis to prevent damage.

4.4.2 Evaluation of the Effectiveness of Plasma Partitioning using 2D mini gels and LC-MS analysis

Plasma partitioning spin columns from Beckman Coulter (ProteomeLab IgY Spin Columns) were evaluated using 2D mini-gel electrophoresis in conjunction with LC-MS analysis.

The results of the 2D mini-gel analysis are shown in figures 4.5 – 4.8. The different samples separated and visualised on the 2D gels were an untreated plasma sample, plasma

with abundant proteins removed, the wash step from the partitioning column and the removed abundant proteins.

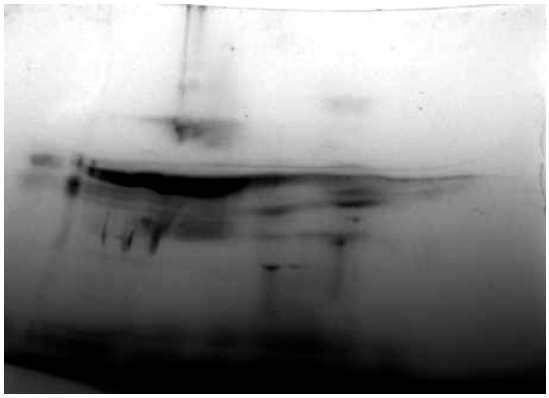


Figure 4.5 2D mini gel of untreated plasma sample, stained with Coomassie blue

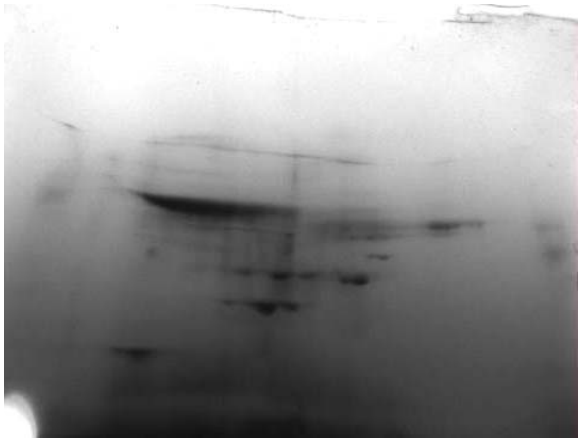


Figure 4.6 2D mini gel of plasma sample with seven abundant proteins removed using the Beckman Coulter IgY plasma partitioning spin column, stained with Coomassie blue.

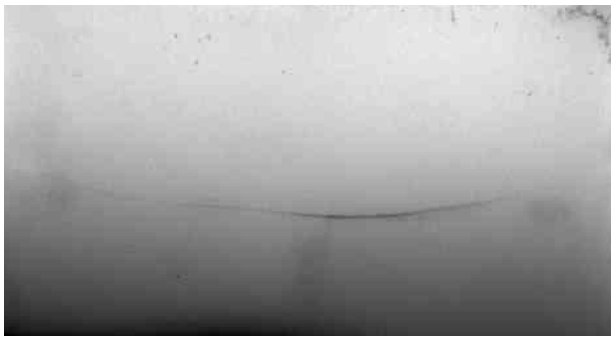


Figure 4.7 2D mini gel of wash step from plasma sample when treated with the Beckman Coulter IgY plasma partitioning spin column, when treated with Coomassie blue.

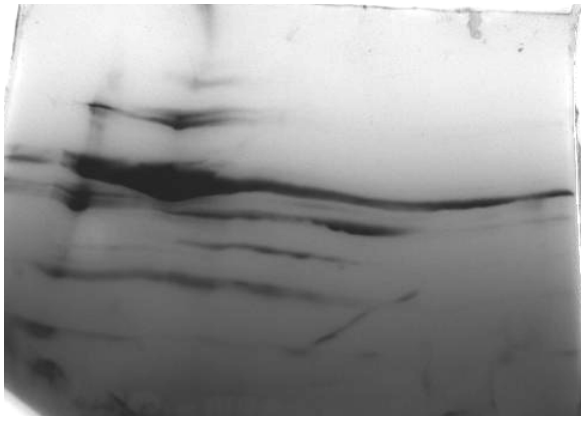


Figure 4.8 2D mini gel of the seven abundant proteins removed from plasma using the Beckman Coulter IgY plasma partitioning spin column, stained with Coomassie blue.

The 2D mini gels emphasize the complexity of the plasma samples, and also the high abundance of certain proteins, seen as larger smears on the gels. The large black smear seen in the centre of all the gels is presumed to be albumin, as it is seen in such high abundance. It is present even in the gel of the plasma sample treated with partitioning spin columns, where in theory it should be removed. This demonstrates that although the columns undoubtedly remove a proportion of the abundant proteins from the plasma samples, it is clear this is not a completely efficient process. From examination of the gel, of the sample consisting of the abundant, removed, proteins it is clear that there are more than seven spots present on the gel. Some of these may be due to modified versions of the abundant proteins, but it is proposed that some of these spots may be proteins that have been associated with the abundant proteins, and were unintentionally removed from the plasma samples. This type of pre-treatment causes problems when the samples are to be analysed quantitatively. If differing amounts of low abundance proteins, or indeed the 7 high abundance proteins, are removed from plasma samples, this could indicate artificial quantitative differences in the samples. It is suggested here that this type of plasma partitioning is not appropriate to be used in a quantitative biomarker discovery study, due to the differing amounts of protein that is removed from each sample, and also due to unintentional removal of low abundance proteins which may be of relevance.

Each fraction from the depletion column separation was also analysed using LC-MS-MS analysis and the results searched against the MASCOT database. Table 4.2 shows in which fraction each of the 7 abundant proteins was identified as being present:

Protein	Fractions Identified
Albumin	Wash fraction, Abundant proteins
IgG	None
Transferrin	None
Fibrinogen	All fractions
IgM	None
A1-antitrypsin	Wash fraction, Abundant proteins
Haptoglobin	None

Table 4.2 Seven abundant proteins as partitioned by the Beckman Coulter IgY plasma partitioning spin column. Shown here is a summary of which fractions the proteins were identified in, to help evaluate the efficiency of the column.

It can be seen from table 4.2 that the only three proteins identified in the 'abundant proteins' fraction are albumin, fibrinogen and a1-antitrypsin. Fibrinogen is also identified in all other fractions, indicating that it has certainly not been efficiently removed from the plasma sample. IgG, Transferrin, IgM and Haptoglobin were not identified in any of the fractions. It is proposed here that this indicates that these proteins are all in fact present in the abundant protein fraction, as indicated by the 2D mini gel, but are unable to be identified due to the high abundance of albumin and fibrinogen also present in this fraction. That they are not identified in the depleted plasma fraction may indicate that they have been efficiently removed from this sample, however this is a rather tenuos conclusion as no identifications were made of these proteins. The results presented here only emphasise the difficulties and challenges encountered when attempting to analyse plasma samples in this method.

It is concluded from the study and results presented in this section, that this type of plasma partitioning is not appropriate for use in a quantitative biomarker discovery study. This is predominantly due to the differing amounts of abundant proteins that are removed from each sample, which would undoubtedly distort any attempt at a quantitative analysis. Additionally, the unintentional removal of interesting low abundance proteins is a significant disadvantage to this type of experimental procedure.

A second, simpler type of plasma partitioning was also evaluated using a 2D mini gel. This was a molecular weight separation strategy, using Microcon Nominal Molecular Weight Filters. This separates the plasma sample into two separate fractions, one above 30 kDa in size and one below. The fraction above 30 kDa can be seen in the figure, 4.9, below:

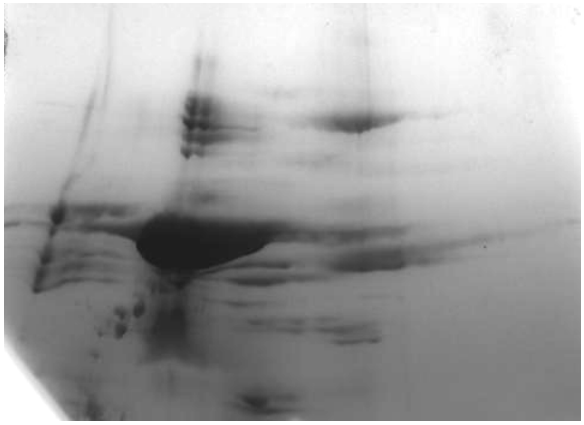


Figure 4.9 2D mini gel of plasma sample. (>30 kDa) after being size fractionated using Microcon Nominal Molecular Weight Filters

It can be seen, if compared with the untreated plasma gel in figure 4.5, that the majority of the proteins spots seen in the untreated plasma gel are also present in the above 30 kDa fraction gel. If low abundance, small proteins are to be targeted for identification, this could be an effective way of removing the majority of the protein contingent from the plasma sample. Additionally, this method is much cheaper and less labour intensive than the partitioning columns evaluated above.

However, it has been decided, for this study, that this is not an appropriate sample preparation method. The research presented here is an unguided, biomarker discovery project and therefore the removal of potentially interesting proteins could be detrimental to the project. Additionally, it has also been decided that the samples should be manipulated, and treated, as little as possible for this work. This will minimise the chance of artificial differences and bias being introduced into the samples. However, it is suggested that if depletion is a sample pre-treatment strategy of choice, then size fractionation may be a cheaper and more efficient method than antibody affinity partitioning.

4.4.3 LC-MS - Repeatability and Reproducibility Evaluations

A high level of repeatability and reproducibility in the LC-MS system is of vital importance for a label free quantitation experiment, due to the need to compare samples analysed at slightly different times. The reproducibility must be evaluated and kept to as high a level as possible throughout the biomarker study, to avoid introducing bias, or artificial, distorted differences into the results. The repeatability of the protocol was evaluated on different LC systems and compared, prior to the main sample analysis.

The repeatability of the Dionex Ultimate LC system and the Applied Biosystems Q-Star was evaluated, by analysing the same sample on three separate days. A base peak chromatogram was created for each sample and these were overlaid, as can be seen in the figure, 4.10, below. A base peak chromatogram monitors the most intense peak in each spectrum. Therefore when a base peak chromatogram is plotted it represents the intensity of the most intense peak at every point in the analysis. Base peak chromatograms are often used as they focus on a single analyte at each point in the analysis, meaning the background signal is reduced.

Examination of figure 4.10 shows that the repeatability of this system is not ideal. This is presumed to be due to the fact that the LC system does not have a good level of flow control. The column is also not temperature controlled, meaning analyte retention times are unlikely to be highly reproducible.

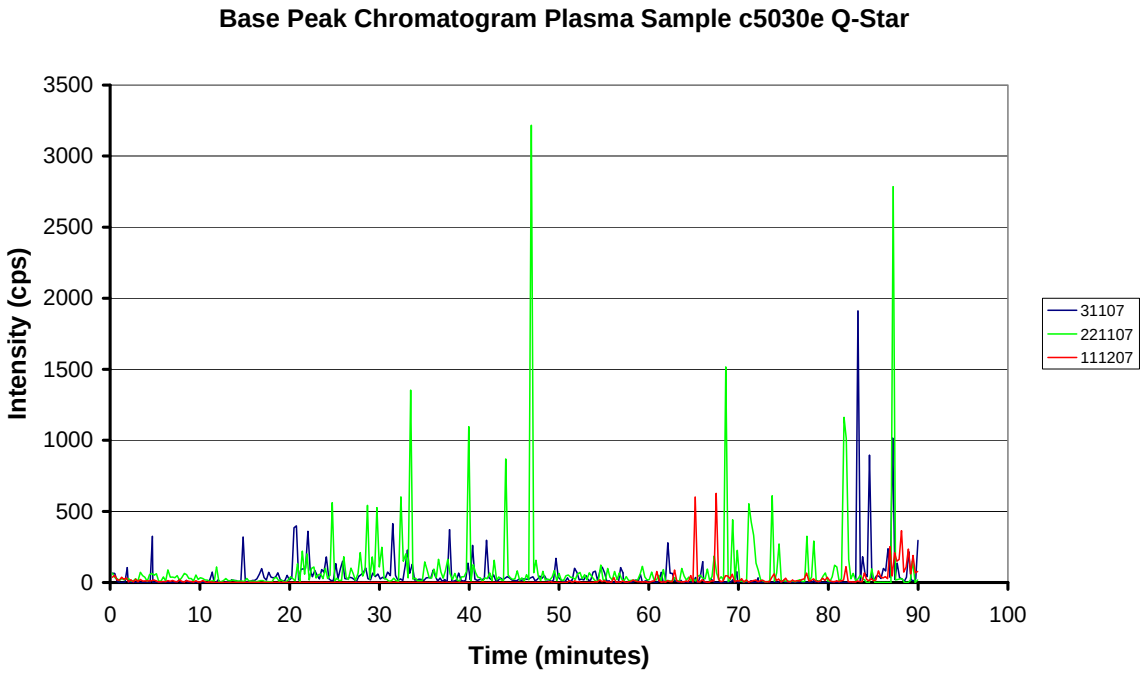


Figure 4.10 Base peak chromatogram from plasma sample analysed using Dionex Ultimate LC system and Applied Biosystems Q-Star. The different colour represent data collected on different days, with the dates depicted in the legend to the right. It can be seen that the day-to-day repeatability of this analysis is poor. This is thought to be due to due to the poor flow control and lack of temperature control on this LC system.

The repeatability of a Dionex 3000 LC system coupled to a Bruker High Capacity Trap (HCT) was also evaluated. Initially this was performed using a standard BSA digest analysed on the same day. The base peak chromatogram created from this experiment is shown in figure 4.11. On examination of figure 4.11 it can be seen here that there is a much higher level of repeatability when compared with figure 4.10. This is due to the fact

that this LC system has a high level of flow control and, in addition, the LC column is contained in a temperature controlled oven, minimizing temperature differences between samples.

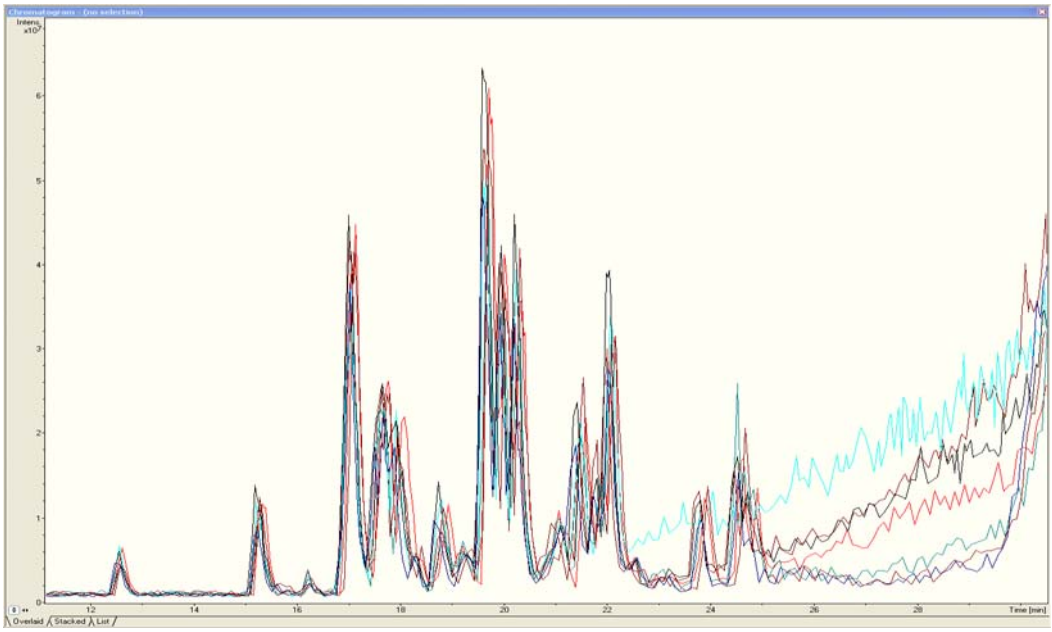


Figure 4.11 Base peak chromatogram from BSA digest (used as a standard sample) analysed using a Dionex 3000 LC system and a Bruker High Capacity Trap (HCT). Again the different colours represent data collected on different days. A higher level of repeatability can be seen with this LC system, as is has better flow control, as well as a temperature controlled oven for the column. This combination gives a higher level of repeatability.

Also shown, in the figures 4.12 and 4.13, is an enlarged extracted ion chromatogram from a particular peptide (421.5 Da) from the BSA digest, from two separate, and also six separate samples analysed on the same day. In an extracted ion chromatrogram a single analyte is monitored throughout the entire analysis. The total intensity (within a mass tolerance window) around the mass of interest is plotted at each point in time.

Again, this demonstrates that this system has a highly reproducible retention time for a particular peptide.

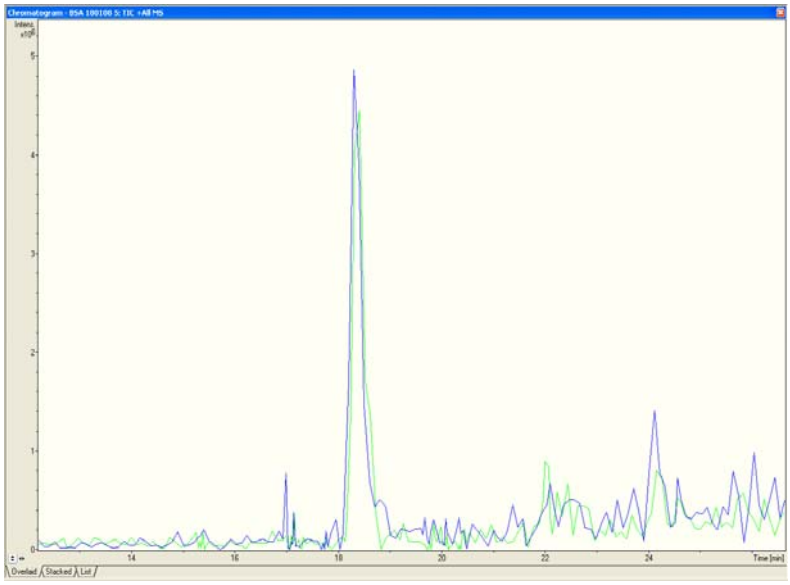


Figure 4.12 XIC of Trypsin peak 421.5 Da, from a BSA digest sample analysed using a Dionex 3000 LC system and a Bruker High Capacity Trap (HCT), on one day. The different colours represent the same analysis repeated twice on the same day. This shows the high level of repeatability of retention time from a particular peptide.

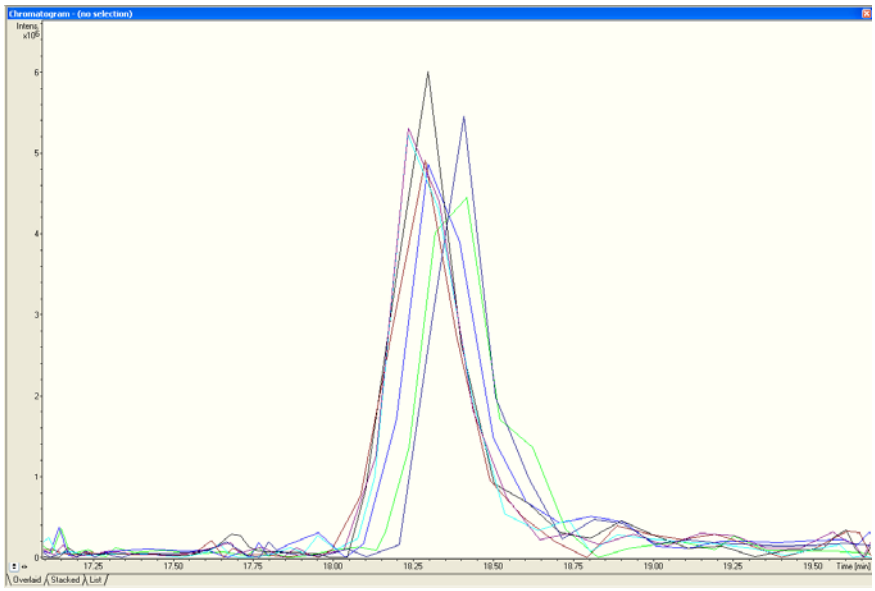


Figure 4.13 XIC of Trypsin peak 421.5 Da, from a BSA digest sample analysed on the same day. The different colours represent the same analysis repeated on the same day. This shows the high level of repeatability of retention time from a particular peptide.

In the next repeatability test, the BSA digest was analysed using the Dionex 3000 LC system coupled to a Bruker High Capacity Trap, on a day-to-day basis. Again, the base peak chromatogram of two samples can be seen in figure 4.14. On examination of this base peak chromatogram, it can be seen that the day-to-day repeatability is slightly lower than that of the same day repeatability. The retention times seem to be slightly shifted between days. This, however, is not a big problem for the label free relative quantitation performed

in this work, as the software can be used to align small retention time shifts between samples.

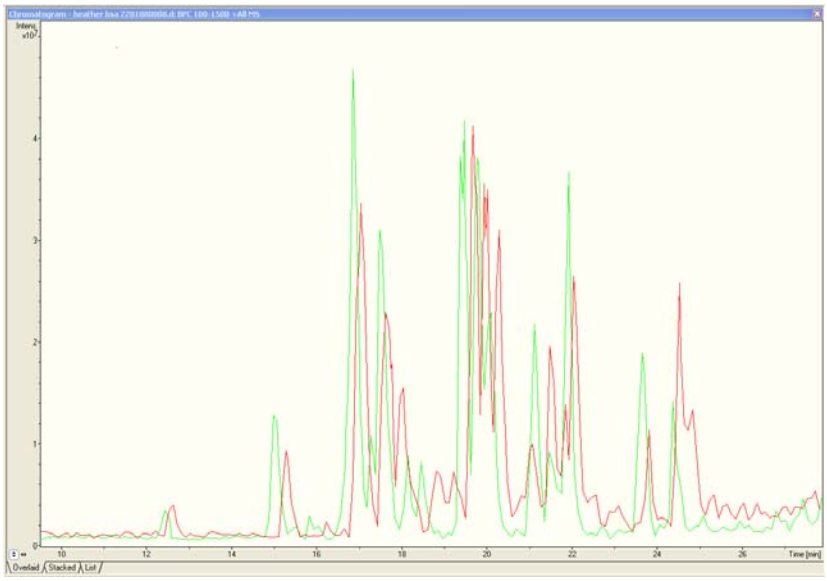


Figure 4.14 Base peak chromatogram from BSA digest analysed on a Dionex 3000 LC and Bruker HCT, showing day-to-day repeatability. The two different colour represent data collected on different days.

Following this, the repeatability of plasma samples was analysed, again using a 3000 LC system coupled to a Bruker High Capacity Trap. The results of overlaid base peak chromatogram from this can be seen in the figures 4.15 and 4.16. To evaluate the reproducibility of the sample preparation protocol, one plasma sample was separated and treated with the Pierce de-salting columns, and again the base peak chromatograms were overlaid. Examination of figure 4.15 shows that this process has a slightly lower level of reproducibility. This is thought to be due to differing protein sample loss from the de-salting columns. This is not acceptable for this type of study, due to possible artificial differences being introduced into the samples. Following this, the same experiment was repeated, however, as an alternative to the de-salting columns, an automated de-salting step was used, incorporated into the LC method. This involved loading the digested plasma sample onto the LC-MS trap, and washing for 15 minutes, prior to separation on the column. It is demonstrated by the base peak chromatogram (shown in figure 4.16) from this experiment, that this process improves the reproducibility of the sample preparation protocol to an acceptable level for this study.

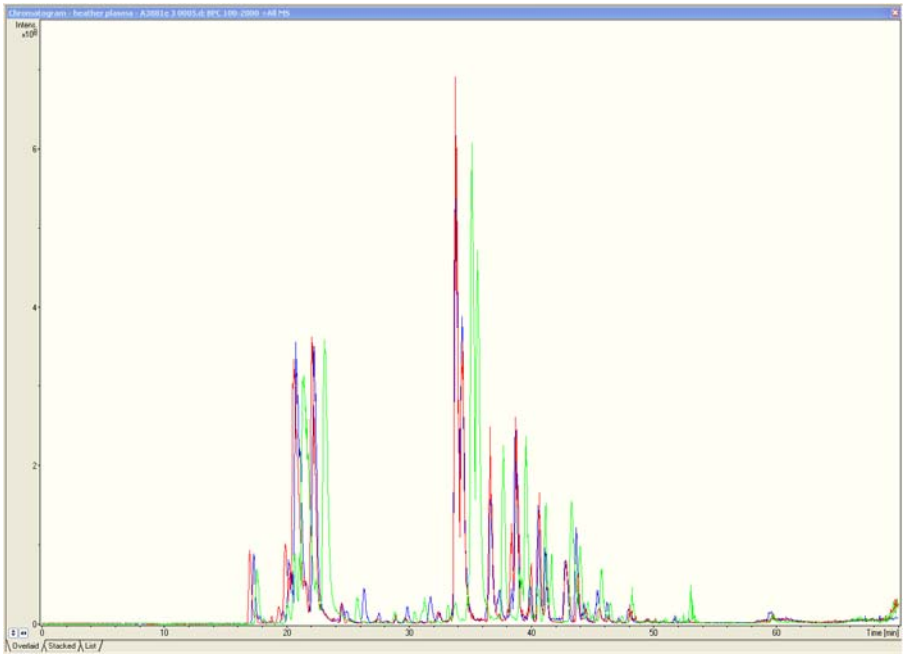


Figure 4.15 Base peak chromatogram from one plasma sample prepared and analysed three times using the Dionex 3000 LC system and Bruker HCT, showing a lower level of reproducibility, thought to be due to sample loss during the de-salting step of the protocol. The different colours represent data collected from one sample multiple times on the same day.

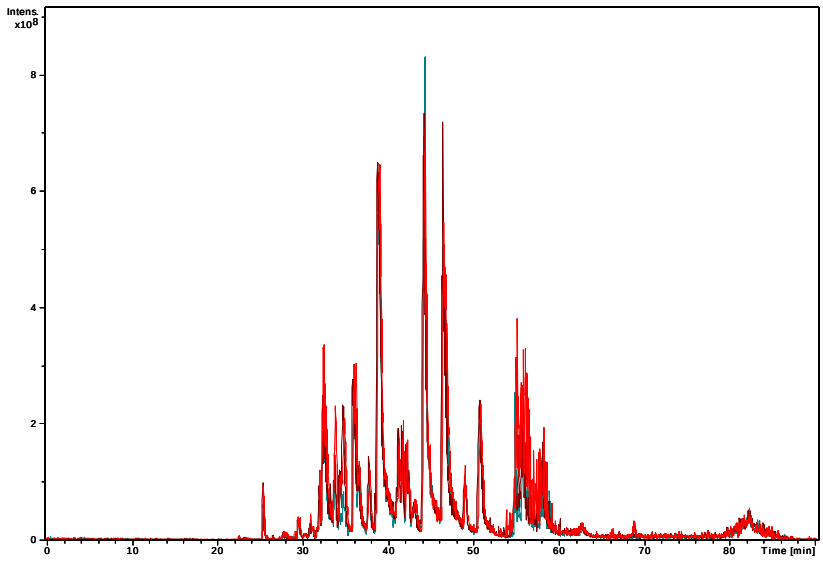


Figure 4.16 Base peak chromatogram from one plasma sample prepared, without de-salting columns and with automated de-salting LC method, and analysed three times using the Dionex 3000 LC system and Bruker HCT, showing a higher level of reproducibility, thought to be due removal of the de-salting step of the protocol. To replace this, a de-salting step is included in the LC separation. The different colours represent data collected from one sample multiple times on the same day. It can be seen that the chromatograms overlay closely on top of one another.

It is concluded from this study, that the Dionex 3000 LC system used here has an acceptable level of repeatability for further work of this type. Additionally, it is also

concluded that using an automated de-salting step with the LC system is preferable to de-salting spin columns, which seem to alter the samples somewhat.

It should be noted at this point that the mass spectrometer used is not important for the repeatability and reproducibility of the experiment. This LC system can be coupled to any electrospray mass spectrometer, and should still give the same high level of repeatability and reproducibility. For the MS and label free quantitation experiment the Dionex 3000 LC system is used coupled to a Bruker MicroTOFq mass spectrometer, and for the targeted MS-MS experiment the same LC system is used, this time coupled to the Applied Biosystems Q-Star.

4.4.4 Evaluation of Label Free Relative Quantitation Software

As previously mentioned, in the method section 4.3.5, in this study two different types of software were initially evaluated for the label free relative quantitation study. These were Progenesis LC-MS from Non-Linear Dynamics and Decyder MS from GE Healthcare. However, all data presented in this work is obtained from Decyder MS. The reasons for this are discussed in the following paragraphs.

There are significant differences to the way these two softwares work, although the basic principle behind both of them is the same. Label free relative quantitation is based on the principle of comparing identical proteolytically digested peptides, which elute with the same retention time using LC-MS analysis, from different samples. The relative ratios of the protein can then be calculated.

Progenesis LC-MS was the first software to be evaluated in this project. It is user friendly software, designed to be simple to use, even for a user with no previous experience of label free quantitation. It has an appealing user interface, and the statistical output is straight-forward to interpret. However the disadvantage of this software is that it is very inflexible. Settings are all standardised and cannot be adjusted. It is also difficult to export data for further analysis.

Decyder MS is a more specialised tool for label free quantitation, designed for use by a user with specific needs. The mass spectrometer type used is definable, for instance a trap, time-of-flight or FT-ICR mass spectrometer. If MS-MS has been performed on the data set, the sampling can be easily evaluated visually, as this is displayed in a 2D LC-MS plot. Detection parameters can be set by the user, for instance typical peak width, resolution and

desired range of charge states. Decyder MS also has an advanced detection parameters tab allowing advanced parameters to be defined, for instance, background correction algorithm, signal to background detection threshold, whether charge assignment is required, increased sensitivity for small peptides, charge state matching parameters, m/z shift tolerance, m/z shape tolerance and filtering parameters. Following peptide detection manual editing can be performed, meaning charge assignments can be manually modified and comments can be added. MS-MS identification can be performed through Decyder MS with either Mascot or Sequest databases. Time alignment parameters are also user defined. Time alignment parameters refer to maximum stretch/compression allowed, maximum leader and trailer and stretch/compress penalty. Following peptide matching cross-detection can be performed. This adds matched peptides present in at least a user specified number of intensity maps to all other intensity maps. The intensity values for these peptides are then measured. Cross detection is essential for statistical evaluation of presence/absence experiments. Peptides detected can be ordered according to many more parameters in Decyder MS, than in Progenesis LC-MS, including number of profiles present in, mass and retention time. Expression profiles can be viewed in two ways, either as individual peptides or experimental groups. Decyder MS allows normalization to be performed in a user defined manner. This gives the option to use internal standards for normalization, or alternatively the entire peptide population. Decyder MS reports ANOVA and t-test statistics, along with average difference and average ratio of peptides. Decyder MS data is easily exportable into excel format for further statistical analysis.

For these reasons it was decided that Decyder MS is more suitable for use in this study and all biological replicates and final data was analysed with Decyder MS. The main steps in label free quantitation using Decyder MS are:

- Peptide detection in each sample
- Definition of experimental groups
- Alignment of intensity maps
- Matching of detected peptides
- Cross-detection of matched peptides
- Evaluation of matched result
- Statistical analysis

4.4.5 MS and Label Free Relative Quantitation for Biomarker Discovery in Plasma

MS data for all 38 samples was collected in a randomised manner, using the Dionex 3000 LC system, coupled to the Bruker MicroTOFq. The randomisation of the samples was used to prevent the introduction of bias into the quantitative results. The MS data from each sample was converted into an 'mzxml', compatible for import into Decyder MS. From this, a 2D map was created for each sample with the axes being m/z and retention time. The features in each map were automatically detected across each sample, before being retention time aligned. Cross-detection across all samples was then performed, and a summary of features detected in each sample group is shown in the figure 4.17. The average number of features detected in each sample is between 8880 with a standard deviation of 947.

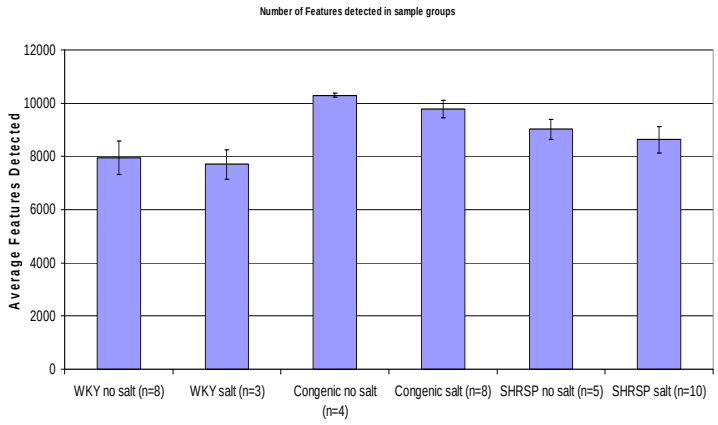


Figure 4.17 Graph demonstrating the average number of features detected in each sample group, with error bars representing ± one standard deviation. It is shown that between 8000 and 10,000 features are detected within each sample group.

The large number of features detected in the plasma samples is as expected, due to the high level of complexity of the samples and the wide dynamic range of proteins present. It can be seen from examination of figure 4.17 that a higher number of features were detected in the congenic plasma samples, when compared with the wild type samples. In addition the SHRSP samples show a higher number of detected features when compared with the wild type samples. The reason for these differences is unclear at this stage.

4.4.6 Statistical Analysis of LC-MS data for Label Free Quantitation

The LC-MS data from the different groups of plasma was analysed in several different ways. Initially a simple one-way analysis of variance (ANOVA) calculation was performed. The principle behind ANOVA analysis is described in detail in the previous chapter 'Assessment of Denator Heat Treatment for Prevention of Proteomic Sample Degradation using Label Free Relative Quantitation'.

In this analysis a decision must to be made regarding what level to set a significant p-value. As previously stated, a common method is to simply choose an arbitrary value of 0.01, 0.005 or 0.001. This would mean that the probability of obtaining the observed difference by chance would be 1 in 100, 0.5 in 100, or 0.1 in 100, respectively. This decision is somewhat further complicated by the number of features present, and therefore ANOVA calculations present in the dataset. In this case there are around 12,000 features present in the dataset, meaning that determining the significance of features detected can be somewhat complicated, as if enough tests are performed it is inevitable that a 'significant' feature will be found eventually, simply by chance. The researcher must be aware of attaching too much weight to a lone, or a few, significant results amongst a large amount of 'non-significant' results. One method of helping to determine true significance is reducing the significance level of the p-value, based on the number of tests performed. This is known as the Bonferroni correction. The Bonferroni correction is based on the reasoning (381-383) that if a null hypothesis is true, i.e. there is not a change in abundance between populations, then a significant difference, $p < 0.05$, will be observed by chance once in every 20 experiments. This is known as the type I error, or α. When 20 independent tests are performed, or in this case 20 different features are tested for significance, and the null hypothesis is true for all 20, then the chance of each test being determined as significant is no longer 0.05, but 0.64. The formula for this error rate is $1-(1-\alpha)^n$, where n is the number of tests performed. However, the Bonferroni adjustment reduces the α applied to each feature, so the error rate across the experiment remains at 0.05. The reduced significance level is $1-(1-\alpha)^{1/n}$, giving 0.00256 in this case. This is often approximated as α/n (0.0025). Therefore, for this dataset (with around 12,000 features) the reduced significance level could be approximated as $0.05/12000 = 4.16 \times 10^{-6}$. This would clearly give an extremely reduced number of 'significant' features, when compared with a standard p-value of 0.05. In this dataset if the p-value used is 0.05 there are over 5000 features deemed to be 'significant'. In comparison if the Bonferroni method is used as

described above this is reduced to around 500, which is clearly a much more manageable number to target for identification.

There is, however, an alternative school of thought, which suggests that the Bonferroni correction is not appropriate in many cases (384;385). The Bonferroni adjustment implies that any comparison will be interpreted in a different way, depending on the number of tests that are performed. Furthermore, type I errors cannot be decreased without the inflation of type II errors.

Type II errors refer to the probability of accepting the null hypothesis when it is not true. In this case, this would mean that the chance of not detecting a change in feature abundance when it is in fact present would be increased. Most advocates of the Bonferroni correction would include all statistical tests in a given report as a basis for the n-number by which to adjust p values. This in itself can cause further complications, as further tests may have been performed that were not included in the publication, or were included in other, related papers, meaning that a decision must be made regarding the value of the true n-number.

A further statistical method that has been developed in order to control the high false positive rate in microarray data is known as the False Discovery Rate (FDR) (386). The false discovery rate of a test is defined as the expected proportion of false positives within the declared significant results (387-389). Therefore the FDR can be a more useful statistical value than a p-value. The data set collected here, via LC-MS, has certain parallels which can be drawn with microarray data. Microarray data sets tend to test a large number of genes, which can be related to the high number of LC-MS features tested here. It is proposed here that using a statistical method designed for microarray data may be the most appropriate way of dealing with this type of LC-MS dataset.

FDR was originally described in 1995 by Benjamimi and Hochberg (387) and is a statistical method used in multiple hypothesis testing to identify the expected proportion of false positives among all significant hypotheses. For example if 100 observations were predicted to be different and a maximum FDR for these observations was 0.1 then the expectation would be that 10 of the observations were false positives.

Benjamimi and Hochberg then went on to demonstrate that the FDR can be developed into a simple and powerful procedure. It has since been used routinely for many microarray applications (386;390-393). More recently, this type of statistical analysis developed for use with microarrays has been applied to proteomics data (394-397).

In my research presented here, the false discovery rate calculation has been used in two separate ways. Both methods use multiple pairwise comparisons between the six groups, instead of the previous general comparison across all six groups. The pairwise comparisons can be visualised using a set of three Venn diagrams as shown in figures 4.18 to 4.20.

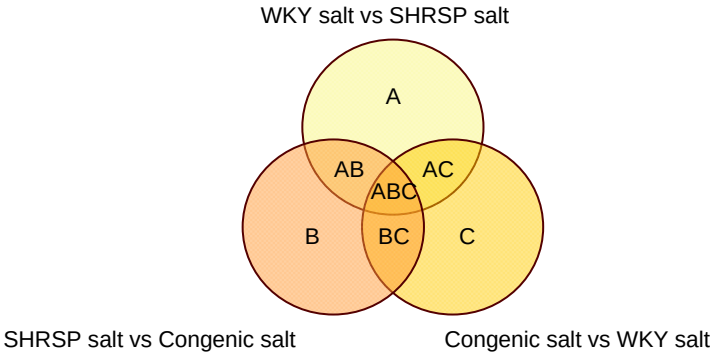


Figure 4.18 Venn Diagram number 1, showing pairwise comparisons of salt treated samples. Features are deemed to be significant between different sample groups based on a 5% FDR cut off value.

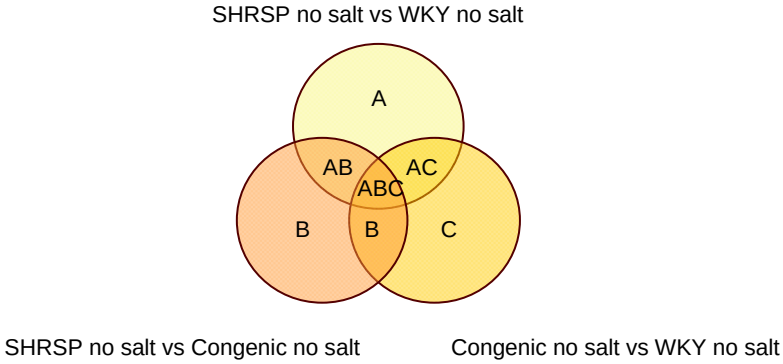


Figure 4.19 Venn diagram number 2 showing pairwise comparisons of untreated samples. Features are deemed to be significant between different sample groups based on a 5% FDR cut off value.

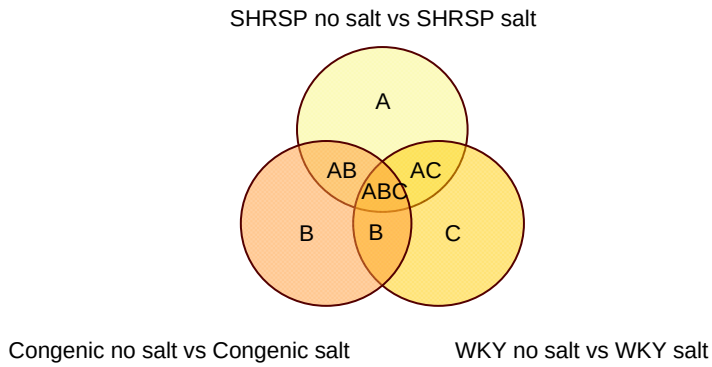


Figure 4.20 Venn diagram number 3, showing pairwise comparisons between salt treated and untreated samples. Features are deemed to be significant between different sample groups based on a 5% FDR cut off value.

Each part of the Venn diagram can be thought of as the statistically significant features from one pairwise comparison. The features are defined as significant in two separate ways, firstly using a pairwise Student's t-test and false discovery rate, and secondly with a method known as RankProducts analysis (RP) and false discovery rate.

The Student's t-test (henceforth referred to as t-test), is a commonly used statistical test. It is similar to the previously mentioned one-way ANOVA test, but it may only be used on two populations. It is used to test whether there is any significant difference between the means of the two populations. In this case a t-test is performed for each feature, for each pairwise comparison, as defined by the Venn diagrams. A small t-test p-value would indicate that there could be a significant difference between the two groups. The t-test used here is an equal variance two-tailed test, meaning it considers both increases and decreases in intensity. The t-test requires a minimum of two data points in each population, and in general the statistical significance will improve with the number of data points. The t-test null hypothesis is again similar to that of the one-way ANOVA, being that there is no change in feature abundance between the populations.

The RankProducts analysis was originally developed for use with microarray data (398;399). It is known as a 'primary' method for identification of differential genes between small groups of biologically replicated samples. It has been shown that the 'significant' lists of samples achieved using RP analysis are superior to those obtained using standard t-tests, if the number of biological replicates is less than ten.

In RP analysis it is assumed that only a minority of features will demonstrate expression changes, measurements are independent between replicate samples, most changes are independent of one another and measurement variance is roughly equal for all features.

The method used to identify differentially expressed features within the dataset is based on the assumption that it is extremely unlikely to observe a single feature with a higher intensity in one set of samples compared with another if none of the the features are in fact present in different intensities, i.e. if all null hypothesis are true. A list of features can be created with the top feature showing the largest intensity difference between the two sample sets, the second feature showing the second largest intensity difference and so on down the list. These RP values assigned to each feature can then be used to sort the features by the likelihood of observing them so high on the list by chance.

Both the pairwise t-tests and the RankProducts analyses were conducted using a false discovery rate with a 5 % cut off value. All RankProducts analyses was undertaken by Dr McClure (Institure of Cardiovascular and Medical Science, University of Glasgow) using a Perl script and a corresponding Windows executable system developed by R Breitling (Bioinformatics Research Centre, Institute of Biomedical and Life Sciences, University of Glasgow).

From this statistical analysis lists of significant features were composed. The most ‘interesting’ features were defined as those in which the congenic strain either behaved as the WKY strain, or showed an intermediate behaviour between the WKY and the SHRSP strains. The congenic strain has a background from the SHRSP strain with chromosome 2 from the WKY strain. These specific features should occur in the AB and ABC portions of Venn diagrams 1 and 2 (see figures 4.18 to 4.20). Therefore, the features occurring here were focussed upon. The individual expression profile for each feature was evaluated visually and a decision made regarding its interest. These feature lists were then used for targeted MS-MS analysis for peptide and protein identification.

As mentioned earlier, the initial statistical analysis of this dataset was to perform a simple ANOVA calculation across all six sample sets. This is shown in the figure, 4.21, below. It is seen here that, if even a fairly strict, p-value of 0.001 is set, there are still around 2000 features deemed to have significant changes across the six sample sets. 2000 features are far too many to target for further investigation and identification, in any sort of efficient manner. This prompted the move to investigate further statistical methods of analysis.As discussed above, this can be visualised in the form of three Venn diagrams. These three

Venn diagrams are repeated twice, using pairwise t-tests and 5 % FDR and also RankProducts analysis and 5 % FDR. The results of this analysis are shown in the Venn diagram format, with the numbers present in each section representative of the number of features deemed as having a significant difference between groups. This is shown in figures 4.22 – 4.27.

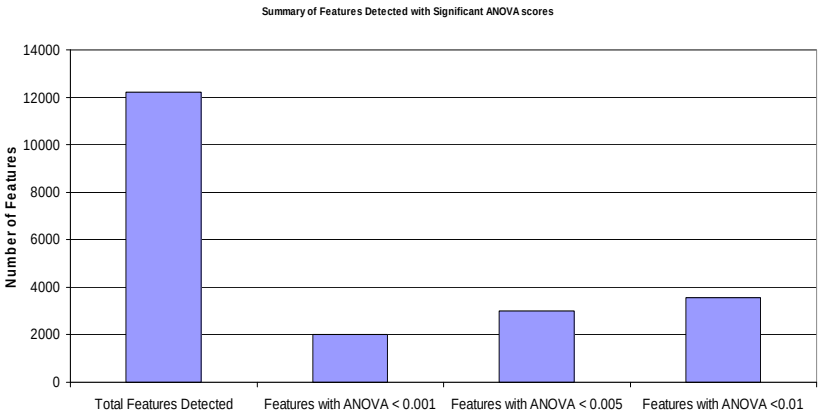


Figure 4.21 Graph displaying results from one-way ANOVA analysis of LC-MS analysis. This demonstrates that with a significant p value of 0.001 there are around 2000 features deemed to be significant.

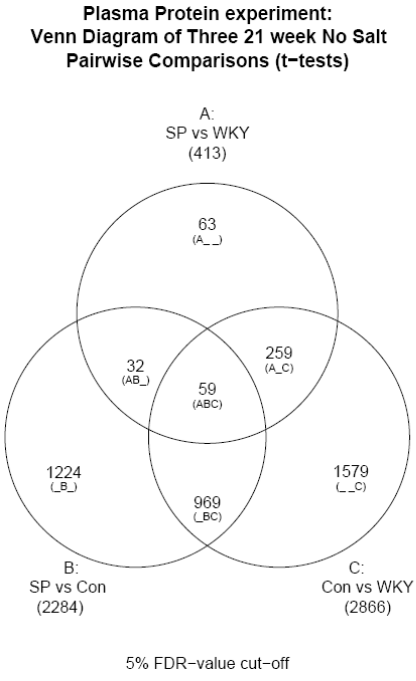


Figure 4.22 Venn diagram showing the no salt pairwise t-test statistical analysis with a 5 % FDR cut off value. Numbers represent the number of features determined as significant using each statistical test.

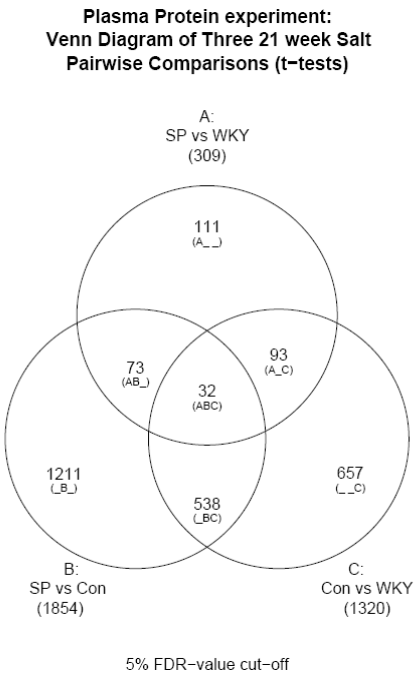


Figure 4.23 Venn diagram showing the salt pairwise t-test statistical analysis with a 5 % FDR cut off value. Numbers represent the number of features determined as significant using each statistical test.

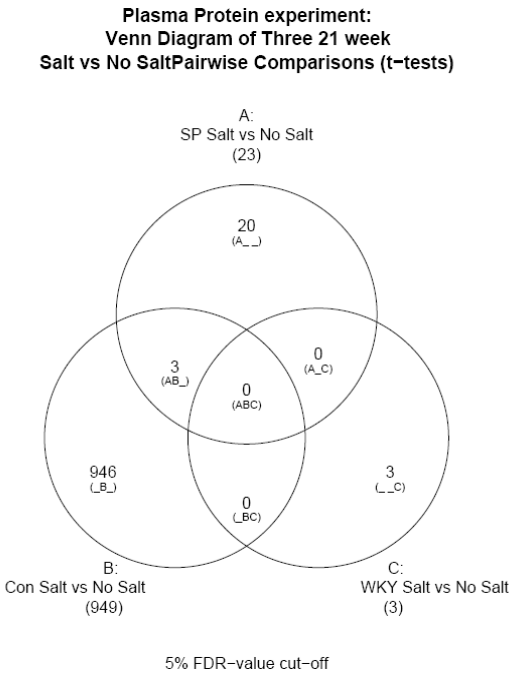


Figure 4.24 Venn diagram showing the salt vs no salt pairwise t-test statistical analysis with a 5 % FDR cut off value. Numbers represent the number of features determined as significant using each statistical test

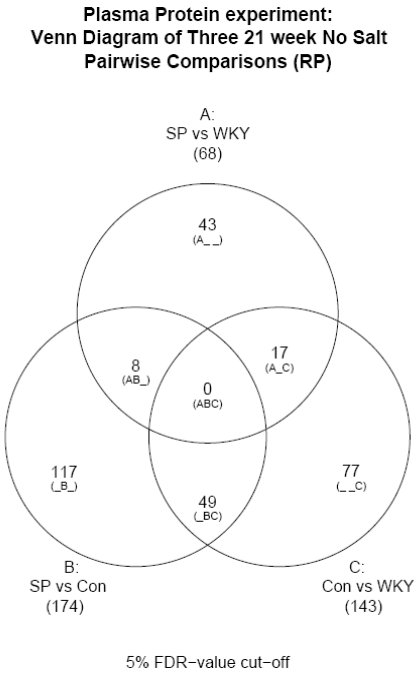


Figure 4.25 Venn diagram showing the no salt pairwise RP statistical analysis with a 5 % FDR cut off value. Numbers represent the number of features determined as significant using each statistical test.

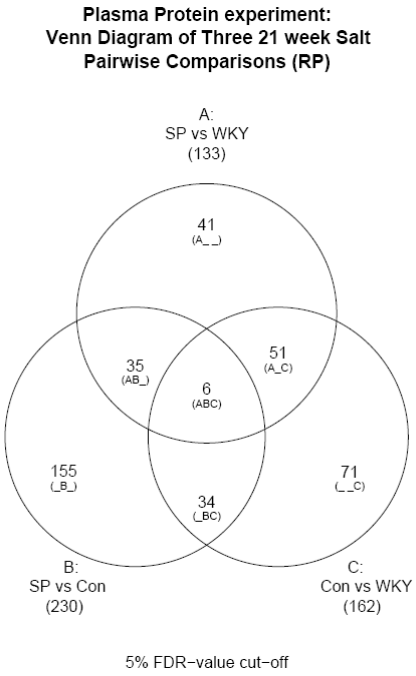


Figure 4.26 Venn diagram showing the salt pairwise RP statistical analysis with a 5 % FDR cut off value. Numbers represent the number of features determined as significant using each statistical test.

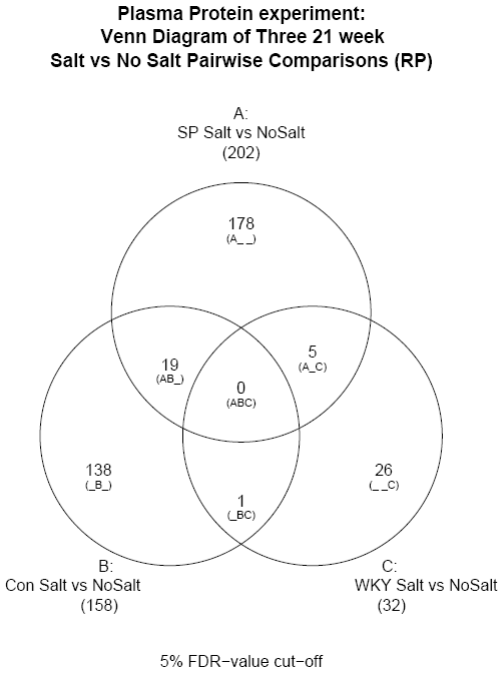


Figure 4.27 Venn diagram showing the salt vs no salt pairwise RP statistical analysis with a 5 % FDR cut off value. Numbers represent the number of features determined as significant using each statistical test.

From visual examination of any features falling into the AB or ABC parts of Venn diagrams 1 and 2, for the RP and t-test analysis, the following features were identified as possible interesting targets for LC-MS:

AB - Match Number	Mass (Da)	2+	3+	Retention Time (min)
1762	360.197	181.098	121.066	59.201
2212	1279.121	640.560	427.374	62.745
2348	931.457	466.729	311.486	56.532
3573	972.719	487.360	325.240	66.335
4263	1393.236	697.618	465.412	50.259
5007	440.231	221.115	147.744	58.908
5780	2232.123	1117.061	745.041	56.840
6441	1016.013	509.006	339.671	60.347
6699	444.273	223.136	149.091	57.399
6892	425.219	213.609	142.740	61.802
7065	1351.687	676.843	451.562	57.928
7195	790.450	396.225	264.483	59.945
7467	640.301	321.151	214.434	47.652
7539	2185.062	1093.531	729.354	47.977
7588	1247.683	624.841	416.894	58.977
7637	848.462	425.231	283.821	59.987
7854	769.444	385.722	257.481	60.702
7880	766.409	384.205	256.470	57.658
7957	1533.844	767.922	512.281	59.499
8259	408.207	205.104	137.069	62.690
8270	894.463	448.232	299.154	58.591
8770	475.207	238.604	159.402	58.913
8788	431.210	216.605	144.737	57.220
9193	1705.823	853.911	569.608	48.768

Table 4.3 Table of features of interest from AB Venn diagram untreated samples and pairwise t-test analysis

ABC - Match Number	Mass (Da)	2+	3+	Retention Time (min)
2304	1508.827	755.414	503.942	62.720
6668	447.230	224.615	150.077	56.086
7679	2150.106	1076.053	717.702	59.419
7840	1307.587	654.793	436.862	60.932

Table 4.4 Table of features of interest from ABC Venn diagram untreated samples and pairwise t-test analysis

AB - Match Number	Mass (Da)	2+	3+	Retention Time (min)
2620	1119.178	560.589	374.059	58.947
2749	1127.779	564.889	376.926	73.518

Table 4.5 Table of features of interest from AB Venn diagram salt treated samples and pairwise t-test analysis

ABC - Match Number	Mass (Da)	2+	3+	Retention Time (min)
2797	946.619	474.309	316.540	56.569

Table 4.6 Table of features of interest from ABC Venn diagram salt treated samples and pairwise t-test analysis

AB - Match Number	Mass (Da)	2+	3+	Retention Time (min)
1696	700.542	351.271	234.514	55.445
2258	1044.964	523.482	349.321	51.490
3992	2154.372	1078.186	719.124	67.930

Table 4.7 Table of features of interest from AB Venn diagram untreated samples and pairwise RP analysis

AB - Match Number	Mass (Da)	2+	3+	Retention Time (min)
1951	1086.758	544.379	363.253	55.152
2698	2109.122	1055.561	704.041	54.492
2024	1807.877	904.939	603.626	61.047
2369	1790.967	896.483	597.989	54.570
1615	1274.231	638.116	425.744	70.261
2969	2148.369	1075.184	717.123	58.640
2056	2268.298	1135.149	757.099	61.475
4167	3700.054	1851.027	1234.351	53.897
2981	1636.020	819.010	546.340	56.122
1346	685.709	343.855	229.570	62.257
1243	811.537	406.769	271.512	68.022
4280	1451.477	726.738	484.826	58.977
4012	946.110	474.055	316.370	68.698
1600	942.445	472.222	315.148	61.127
2220	843.127	422.564	282.042	60.506

Table 4.8 Table of features of interest from AB Venn diagram salt treated samples and pairwise RP analysis

ABC - Match Number	Mass (Da)	2+	3+	Retention Time (min)
4371	677.063	339.531	226.688	57.400

Table 4.9 Table of features of interest from ABC Venn diagram salt treated samples and pairwise RP analysis

Expression profiles of selected features are shown in the figures 4.28 to 4.33. All of these features, in tables 4.4 to 4.10), were subsequently targeted for MS-MS identification.

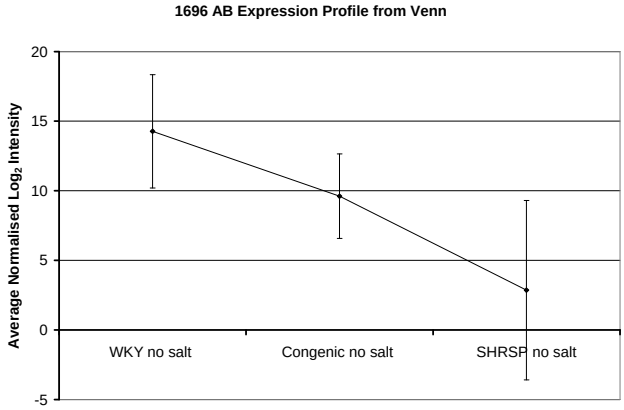


Figure 4.28 Expression profile of feature 1696. This shows the differing intensities across the three relevant sample groups

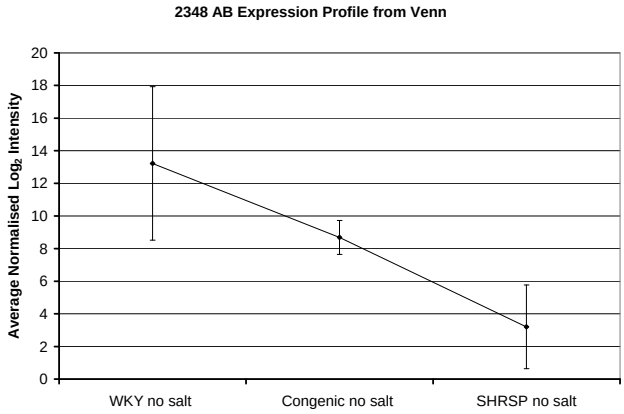


Figure 4.29 Expression profile of feature 2348. This shows the differing intensities across the three relevant sample groups

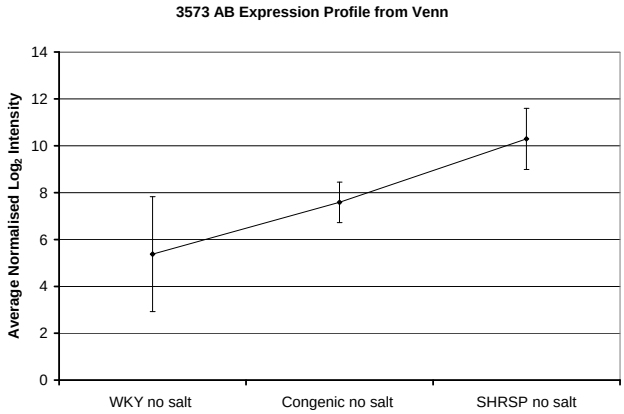


Figure 4.30 Expression profile of feature 3573. This shows the differing intensities across the three relevant sample groups

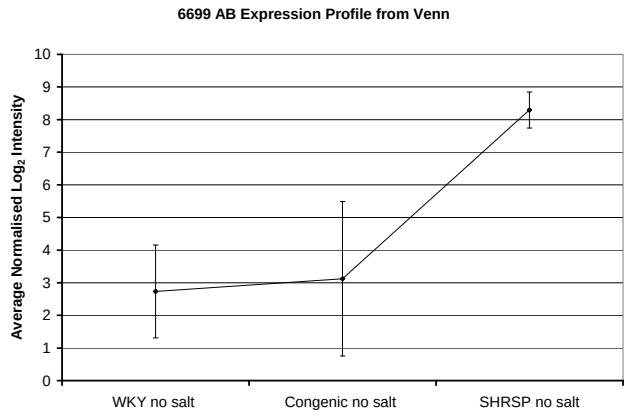


Figure 4.31 Expression profile of feature 6699. This shows the differing intensities across the three relevant sample groups

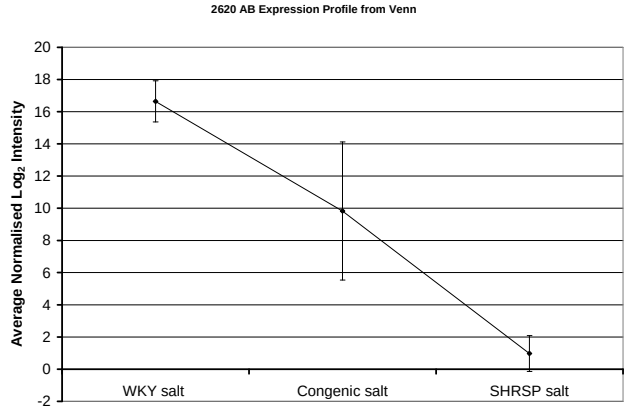


Figure 4.32 Expression profile of feature 2620. This shows the differing intensities across the three relevant sample groups

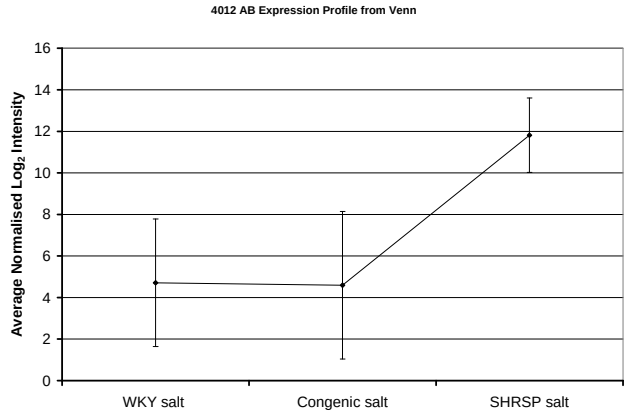


Figure 4.33 Expression profile of feature 4012. This shows the differing intensities across the three relevant sample groups

Examination of figures 4.28 to 4.33 demonstrates that the features selected show different expression profiles across the sample groups. The features shown in figures 4.28, 4.29, 4.30 and 4.32 show an intermediate intensity in the congenic samples, when compared to

the wild type and SHRSP samples. The features shown in figures 4.31 and 4.33 show a similar intensity in the congenic and wild type samples and a different intensity in the SHRSP samples. It is concluded that any of these features could be possible markers for disease. However, it is now important to gain identifications of these features. This will be discussed in the following section 4.4.7.

4.4.7 Targeted MS-MS for Protein Identification in Plasma

Targeted MS-MS analysis was performed, on the Dionex 3000 LC-MS coupled to the Applied Biosystems Q-Star. The MS-MS data from each sample was searched using MASCOT version 2.1 against NCBInr database 20090424. The identified peptides were then matched to the features of interest, using mass and retention time.

From this analysis, four features were identified as being from specific proteins. Their expression profiles and MASCOT identifications are shown in figures 4.34 to 4.41.

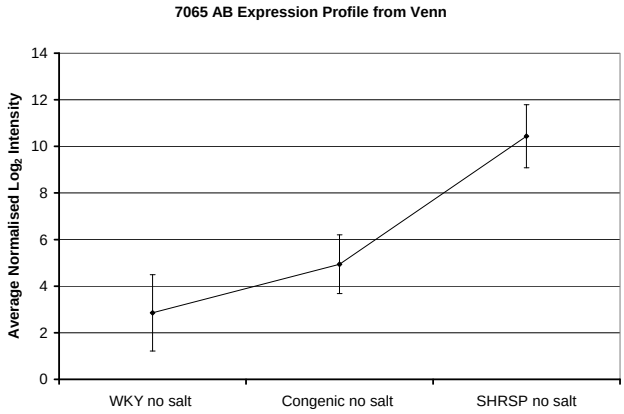


Figure 4.34 Expression profile from feature number 7065, 1351.687 Da (57.9 mins), from Pairwise t tests AB

15. gi|220698 **Mass:** 46448 **Score:** 49 **Queries matched:** 3 **emPAI:** 0.15
contrapsin-like protease inhibitor (CPi-21) [Rattus norvegicus]
Check to include this hit in error tolerant search or archive report

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
232	654.8949	1307.7753	1306.6405	1.1348	0	11	1.2e+02	4	K.IAELFSDLDER.T
241	451.5771	1351.7095	1351.6480	0.0615	1	(34)	0.57	1	R.DTLPHEDQGKGR.Q
247	676.9041	1351.7936	1351.6480	0.1456	1	44	0.062	1	R.DTLPHEDQGKGR.Q

Figure 4.35 MASCOt results showing feature number 7065, identified as contrapsin-like protease inhibitor. Analsyed using Dionex 3000 LC system and Applied Biosystems Q-Star

Examination of the expression profile in figure 4.34 shows that the selected feature, with mass 1351.687 Da, has an increased intensity in the SHRSP no salt samples, when

compared with the WKY no salt samples. Additionally, the feature has an intermediate intensity in the congenic no salt samples. Therefore, the selected feature is a potential marker for hypertension. Figure 4.35 shows the identification of this feature from targeted MS-MS analysis. The mass of 1351.7936 Da corresponds to a peptide from the protein contrapsin-like protease inhibitor.

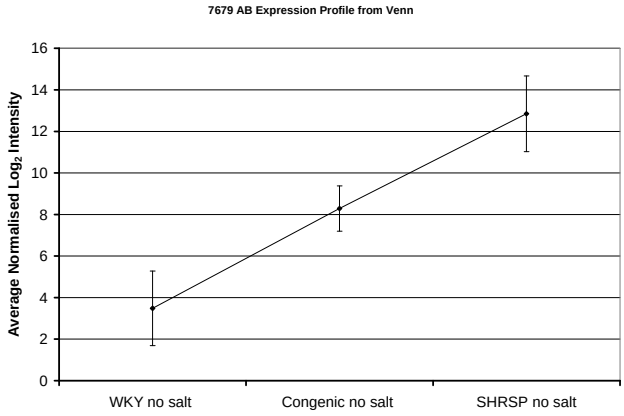


Figure 4.36 Expression profile from feature number 7679, 2150.106 Da (59.4 mins) from pairwise t tests ABC analysis

1. gi|56797757 **Mass:** 86603 **Score:** 137 **Queries matched:** 20 **emPAI:** 0.25
fibrinogen alpha chain isoform 1 [Rattus norvegicus]
Check to include this hit in error tolerant search or archive report

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
42	360.6963	719.3780	719.3636	0.0144	0	33	1	1	K.ALTEMR.Q
45	366.7054	731.3963	731.3813	0.0149	0	35	0.76	2	R.LDELSR.M
46	368.6934	735.3722	735.3585	0.0137	0	(21)	13	1	K.ALTEMR.Q + Oxidation (M)
76	408.7044	815.3943	815.3773	0.0170	0	12	1.2e+02	8	R.GDLPGDSR.G
127	469.7563	937.4981	937.4691	0.0290	0	16	31	1	R.NIMEYLR.G
128	471.2845	940.5545	940.5341	0.0203	1	43	0.069	1	K.ELLPAKDR.Q
133	480.2626	958.5106	958.4906	0.0200	0	27	3.5	1	R.MELERPGK.D
137	484.7592	967.5039	967.4836	0.0203	0	15	43	2	K.TVLGNDGHR.E
192	405.8991	1214.6754	1214.6506	0.0248	1	33	0.8	1	K.GDKELLIGNEK.V
206	649.8383	1297.6621	1297.6302	0.0318	0	56	0.0038	1	K.SQLQEGPPEWK.A
219	687.3749	1372.7353	1372.7061	0.0292	0	35	0.49	1	K.MSPVPDLVPGSFK.S
222	468.5642	1402.6706	1402.6437	0.0270	1	11	1.1e+02	2	R.GDLPGDSRGDSATR.G
225	482.5874	1444.7403	1444.7092	0.0310	1	(15)	49	1	R.MELERPGKDGASR.G
229	487.9167	1460.7282	1460.7041	0.0240	1	18	23	1	R.MELERPGKDGASR.G + Oxidation (M)
270	573.2505	1716.7296	1716.7009	0.0287	0	50	0.013	1	K.MADEAASEAHQEGDTR.T
271	578.5822	1732.7246	1732.6958	0.0288	0	(37)	0.26	1	K.MADEAASEAHQEGDTR.T + Oxidation (M)
272	578.5871	1732.7396	1732.6958	0.0438	0	(28)	2.1	1	K.MADEAASEAHQEGDTR.T + Oxidation (M)
319	717.8000	2150.3782	2151.0055	-0.6273	1	15	24	1	R.LDELSRMHPELGSFYDSR.F
393	905.0990	2712.2752	2712.2039	0.0713	0	66	0.00021	1	K.TSDSDIFTDIENPSSHVPEFSSSSK.T
394	905.4467	2713.3184	2712.2039	1.1145	0	(25)	3	1	K.TSDSDIFTDIENPSSHVPEFSSSSK.T

Figure 4.37 MASCOT results showing feature number 7679, identified as fibrinogen alpha chain isoform 1. Analsyed using Dionex 3000 LC system and Applied Biosystems Q-Star

Examination of the expression profile in figure 4.36 shows that the selected feature, with mass 2150.106 Da, has an increased intensity in the SHRSP no salt samples, when compared with the WKY no salt samples. As for figure 4.34, the feature has an intermediate intensity in the congenic no salt samples. Therefore, as before, the selected feature is a potential marker for hypertension. Figure 4.37 shows the identification of this

feature from targeted MS-MS analysis. The mass of 2150.3782 Da corresponds to a peptide from the protein fibrinogen alpha chain isoform 1.

It should, however, be noted that the identification of this peptide is questionable. The difference between the predicted and observed mass is rather large, at 0.6273, and in addition the score of 15 is lower than the expected score.

As several other peptides have also been identified from this protein, it would be possible to examine their expression profiles and compare them with the expression profile shown in figure 4.36. This could be a potential avenue for future research.

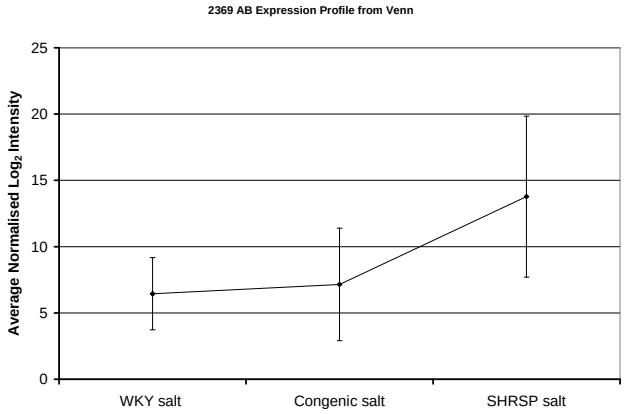


Figure 4.38 Expression profile from feature number 2369, 1790.967 Da (54.5 mins) from pairwise RP AB analysis.

10. gi|21955142 **Mass:** 167053 **Score:** 83 **Queries matched:** 9 **emPAI:** 0.09
pregnancy-zone protein [Rattus norvegicus]
Check to include this hit in error tolerant search or archive report

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
89	430.2393	858.4641	858.4195	0.0446	0	24	8.5	1	K.NVEENVR.E
93	433.8016	865.5886	865.5385	0.0501	0	18	19	1	K.NLKPAPVK.V
173	551.3179	1100.6213	1100.5574	0.0639	0	39	0.22	1	K.LQDQSNIQR.T
191	580.8490	1159.6835	1159.6237	0.0598	0	44	0.074	1	K.VNTLPLNFDK.A
211	410.5793	1228.7162	1228.6524	0.0638	1	21	14	1	K.KLQDQSNIQR.T
235	463.6271	1387.8594	1387.7897	0.0696	0	24	5.5	1	K.MVSGFIPVKPSVK.K
281	836.9970	1671.9794	1671.8832	0.0962	0	37	0.24	1	R.TEVNTNHVLIYIEK.L
295	597.6832	1790.0276	1789.9223	0.1053	1	38	0.21	1	K.VNTLPLNFDKAEHHR.K
461	913.2785	3649.0851	3648.8610	0.2241	1	33	0.26	1	R.SFSYKPRAPSAEVEMTAYVLLAYLTSASSRPTR.D

Proteins matching the same set of peptides:
gi|81872093 **Mass:** 167019 **Score:** 83 **Queries matched:** 9
RecName: Full=Alpha-1-macroglobulin; Short=Alpha-1-M; AltName: Full=Alpha-1-macroglobulin 165 kDa subunit; Contains: RecName: Full=Alpha-1-macroglobul

Figure 4.39 MASCOT results showing feature number 2369, identified as pregnancy zone protein/alpha-1-macroglobulin. Analsyed using Dionex 3000 LC system and Applied Biosystems Q-Star

Examination of the expression profile in figure 4.38 shows that the selected feature, with mass 1790.967 Da, has an increased intensity in the SHRSP salt samples, when compared with the WKY salt samples, however the error bars from the SHRSP samples do overlap

with the lower intensity signals from the WKY and wild type samples, which could mean the higher intensity observed is questionable. In this example the feature has similar intensities in the congenic salt samples and the WKY samples. This feature has still been included here as a potential marker for hypertension, as as mentioned previously the congenic strain has a background from the SHRSP strain with chromosome 2 from the WKY strain. Figure 4.39 shows the identification of this feature from targeted MS-MS analysis. The mass of 1790.0276 Da corresponds to a peptide from the pregnancy zone protein, although the difference between the predicted and observed mass of this peptide is rather larger than the ideal.

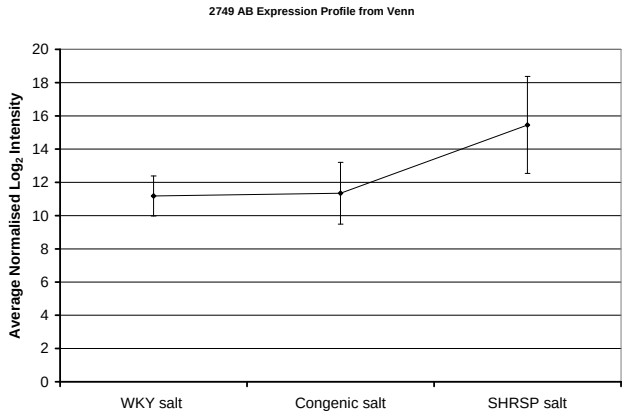


Figure 4.40 Expression profile from feature number 2749, 1127.779 Da (73.5 mins) from pairwise t tests AB analsyis.

8. gi|122065203 **Mass:** 51318 **Score:** 89 **Queries matched:** 12 **emPAI:** 0.18
RecName: Full=Hemopexin; Flags: Precursor
Check to include this hit in error tolerant search or archive report

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
63	397.2274	792.4402	792.3919	0.0483	0	20	17	1	K.GEFVWR.G
65	398.2457	794.4768	794.4286	0.0482	0	23	7.1	1	K.LYVTSGR.R
147	509.7554	1017.4963	1016.5331	0.9632	0	10	2e+02	3	K.VWVYPPEK.K
174	553.3170	1104.6195	1104.5564	0.0632	0	26	4.7	1	K.GGNNLVSGYPK.R
180	564.8000	1127.5854	1127.5189	0.0666	0	14	67	1	K.WFWDFATR.T
194	588.8365	1175.6585	1175.5935	0.0650	0	58	0.0027	1	K.NPVTSVDAAFR.G
205	606.8573	1211.7001	1211.6299	0.0702	0	50	0.014	1	R.FNPVTGEVPPR.Y
255	497.6236	1489.8490	1489.7678	0.0812	1	24	5.7	1	R.WKNPVTSVDAAFR.G
262	506.6296	1516.8670	1516.7926	0.0744	2	17	30	1	K.EDKVWVYPPEKK.E
440	804.4124	3213.6205	3213.4482	0.1723	0	42	0.047	1	R.DGWHSWPIAHHWPQGPSAVDAAFSWDEK.V
470	945.0036	3775.9852	3776.7967	-0.8115	1	7	1.3e+02	2	K.SLGPYSCSSNGPNLFFIHGPNLYCYSSIDKLNAAK.S
481	1049.3011	4193.1754	4193.9728	-0.7974	2	10	51	1	K.LFQEEFPGIPYPPDAAVECHRGECQSEGVLFFQGNRK.W

Figure 4.41 MASCOT results showing feature number 2749, identified as Hemopexin. Analsyed using Dionex 3000 LC system and Applied Biosystems Q-Star

Examination of the expression profile in figure 4.40 shows that the selected feature, with mass 1127.779 Da, has an increased intensity in the SHRSP salt samples, when compared with the WKY salt samples. As for the previous example the feature has similar intensities

in the congenic salt samples and the WKY samples. This means that the selected feature is a potential marker for hypertension, as for the feature in figure 4.38. Figure 4.41 shows the identification of this feature from targeted MS-MS analysis. The mass of 1127.5834 Da corresponds to a peptide from the hemopexin protein, however it should be noted that the score of 14 for this identification is rather lower than the expected score of 67, which is not ideal for a peptide identification.

The possible biological significance of these proteins is discussed in section 4.4.9.

From these four identifications of possible biomarkers, antibodies were available for two of them, hemopexin and fibrinogen alpha-1-isoform.

4.4.8 Western Blotting for Confirmation of Protein Identifications

The western blots of hemopexin and fibrinogen alpha-1-isoform are shown in figures 4.42 and 4.43. Lanes 1 to 6 represent the six different sample groups WKY no salt, WKY salt, Congenic no salt, Congenic salt, SHRSP no salt and SHRSP salt. It was expected that the results from these western blots would mimic the quantitation data from the label free relative quantitation expression profiles.

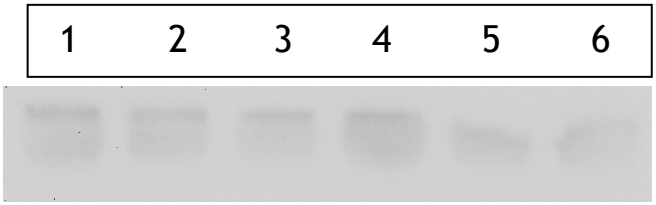


Figure 4.42 Quantitative western blot for fibrinogen alpha-1-isoform. Lane 1 is WKY no salt, lane 2 is WKY salt, lane 3 is congenic no salt, lane 4 is congenic salt, lane 5 is SHRSP no salt and lane 6 is SHRSP salt.

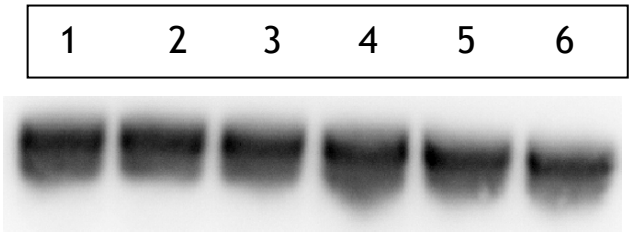


Figure 4.43 Quantitative western blot for Hemopexin. Lane 1 is WKY no salt, lane 2 is WKY salt, lane 3 is congenic no salt, lane 4 is congenic salt, lane 5 is SHRSP no salt and lane 6 is SHRSP salt.

Visual examination of the western blots confirms that the two proteins, fibrinogen alpha 1 isoform and hemopexin, are present in all 6 samples. It is difficult to draw any further conclusions from the blots using only a visual analysis. Quantitative information was extracted from these western blots, using the total pixel intensity from each band. This was then plotted in an expression profile style for comparison with the label free relative quantitation results. This is shown in figures 4.44 and 4.45, error bars shown ± one standard deviation.

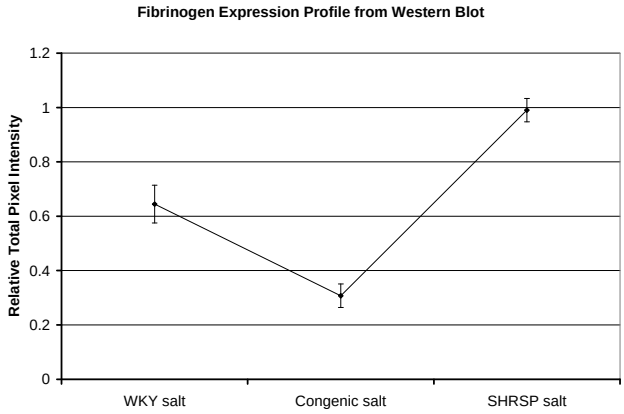


Figure 4.44 Expression profile from quantitative western blot for Fibrinogen

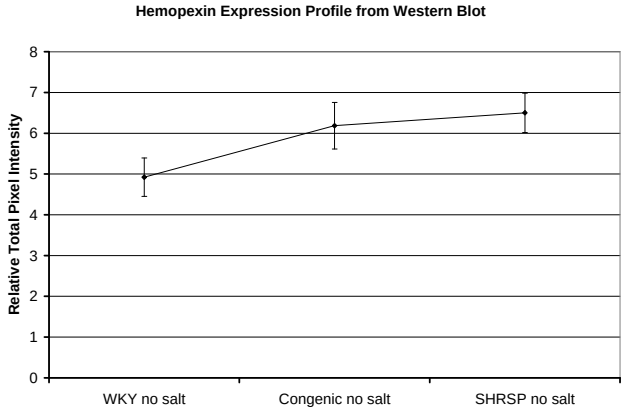


Figure 4.45 Expression profile from quantitative western blot for Hemopexin

It can be seen that the pattern of the expression profiles from the quantitative western blots does not directly mimic the expression profiles seen from the label free quantitation data. However, in both western blots, the intensity is higher in the SHRSP samples, relative to the WKY and congenic samples. This is as predicted by the relative label free quantitation experiment. The profiles may not be exactly the same as those seen in the label free

quantitation experiment for several reasons. The quantitation from the western blots may not be entirely accurate due to the fact that pooled samples were used as opposed to western blots being performed against every sample available, and average results being taken. It is also possible that the label free relative quantitation may not be entirely accurate, due to the relatively low number of replicate samples available. Unfortunately this could not be avoided in this case, but it should be noted for future work that a higher number of replicate samples in each sample set would be preferable.

The western blot serves as a good validation method for the quantitative data previously collected. However, it should be taken into consideration that if the results do not exactly mimic each other, it can be challenging as to what conclusions to draw.

4.4.9 Biological Significance of the Identified Proteins

4.4.9.1 Hemopexin

The main function of hemopexin is known to be binding heme and transporting it to the liver, for breakdown and iron recovery. Following this the free hemopexin is returned to circulation. Hemopexin is expressed by the liver and secreted in the plasma. It is a plasma glycoprotein, and is has the highest known heme affinity. It is also known to be an acute phase protein, with levels being increased in correlation with inflammation (400;401).

4.4.9.2 Fibrinogen alpha-1-isoform

Fibrinogen is known to have two functions, yielding monomers that polymerize into fibrin and also acting as a co-factor in platelet aggregation (UniProt Protein Knowledgebase, http://www.uniprot.org).

4.4.9.3 Pregnancy zone protein (Alpha-1-macroglobulin)

Pregnancy zone protein is known to inhibit all four classes of proteinases using a unique 'trapping' mechanism. It has a peptide stretch, called the 'bait region' which contains specific cleavage sites for different proteinases. When a proteinase cleaves the bait region, a conformational change is induced in the protein trapping the proteinase. The entrapped enzyme remains active against low molecular weight substrates (although activity against high molecular weight substrates is greatly reduced). Following cleavage in the bait region a thioester bond is hydrolyzed and mediates the covalent binding of the protein to the

proteinase. It is found in plasma, and is found in high abundance during the late stages of pregnancy (UniProt Protein Knowledgebase, http://www.uniprot.org).

4.4.9.4 Contrapsin-like-protease inhibitor

Contrapsin-like-protease inhibitor (alternative name Serine protease inhibitor A3K) is known to bind to and inhibit kallikreins. It inhibits trypsin but not chymotrypsin or elastase. It is found in the liver and in plasma. Its level is known to be reduced by inflammation. Its level is also known to be higher in male rats, than in female (402-405).

4.5 Discussion

This work has had several aims, all of which have been achieved. Firstly, the efficiency of two different plasma partitioning strategies have been evaluated, using 2D mini gel and LC-MS analysis. It was concluded that although plasma partitioning has its place as a sample preparation method, neither antibody affinity partitioning or size separation depletion are appropriate for use in a quantitative biomarker discovery project, such as this one. This is for several reasons, the prime one being the possibility of the introduction of artificial differences into the sample set. This has the possibility of causing misleading results, and also may mask true, biological differences between the samples. A second reason that plasma partitioning is not appropriate for biomarker discovery projects is due to what may be lost with the removed fractions. The removal of abundant proteins is likely to also remove interesting, low abundance proteins, which are associated with the more abundant proteins. Therefore, it was concluded that plasma partitioning, or depletion, would not be used in this project.

Secondly, the repeatability and reproducibility of two nano-flow liquid chromatography systems were evaluated. This is a step which cannot be under-estimated when label free quantitation is performed. If the liquid chromatography is not reproducible, none of the quantitation results will be accurate. It was concluded from this, that the Dionex 3000 nano-flow liquid chromatography system, with a 15 cm C18 column was the most appropriate system for this work available, and the Dionex Ultimate nano-flow liquid chromatography system was not appropriate, due to poor flow control and lack of temperature control. It was also concluded that automated de-salting gave better sample reproducibility than manual de-salting spin columns. Ensuring reproducibility of results is a vital step in this type of analysis, and should perhaps consume the majority of experimental time. Following assurance of reproducibility biological replicates can be analysed relatively quickly, with minimal changes to the equipment to help ensure valid results are obtained. Additionally, degradation of samples should be closely monitored to help prevent mis-leading results.

Thirdly, LC-MS data was collected for all rat plasma samples available. This was analysed using Label Free Relative Quantitation software, Decyder MS. This software was also thoroughly evaluated and it was decided to be the most appropriate software available for this project.

One of the main novel aspects of this research is the new application of statistical methods to a relatively new type of dataset. LC-MS and label free relative quantitation of biological samples, such as plasma, produce a wealth of information. There have been many different ways of analysing datasets such as this reported, ranging from simple ANOVA analysis (406), statistical tests such as the Mann-Whitney U test (40) or partial least-squares discriminant analysis (407) to more complicated statistical analysis involving different ways of filtering the dataset and techniques such as FDR (41).

One of the largest challenges here is extracting data in a useful and meaningful way, in order to conduct biomarker research. It is certainly not as straightforward as may initially be imagined, and mining this data for useful changes in abundance needs careful experimental planning and additionally, the extremely careful application of the correct statistical techniques. Simply using a certain technique because it is widely accepted in the field is not an acceptable selection technique. Careful thought and research must be applied to statistics, and it should be one of the most important parts of the data analysis, and not simply an afterthought. Additionally, perhaps statisticians will begin to be incorporated further into biological research, to help enable a fresh view point on these challenging datasets, and more importantly an expert application of the techniques, and possibly even development of new, tailor-made statistical analysis.

In the work presented here a great deal of thought, discussion, expert advice and trialling of different methods has gone into the statistical analysis performed. It is proposed here that the most appropriate method available has been chosen to attempt to identify interesting changes in abundance from this complicated dataset. Different statistical analysis including ANOVA and sets of pairwise t-tests and RankProducts with false discovery rate were evaluated. It was decided that a simple ANOVA calculation was not useful with this dataset, even though it is the most widely accepted and used type of analysis in this field, mainly because it did not specifically target the interesting features of the dataset. Both the pairwise t tests and RankProducts analysis with a 5% FDR helped to target these features in a more specific manner.

For future work it would be possible to investigate the use of two or three-way ANOVA to identify features of interest. A significant p-value resulting from a one-way ANOVA test indicates that a feature is present at a significantly different level in at least one of the groups analyzed. However, if there are more than two groups being analyzed, the one-way ANOVA does not specifically indicate which pair of groups exhibits statistical differences. Two-way ANOVA tests (also called two-factor analysis of variance) can be used to

measure the effects of two factors simultaneously, for example the effect of the disease and the salt treatment. Two-way ANOVA is able to assess both disease and salt treatment in the same test, and also whether there is an interaction between the parameters. Three-way ANOVA works in a similar way for three factors. These methods were not applied in this work, in order to keep the statisical analysis used simple and, in addition, to investigate the effect of using RankProducts and FDR anaylsis on the dataset. The two or three-way ANOVA could be a possible future avenue for investigation.

Following the statistical analysis, a targeted MS-MS strategy was used to attempt to identify the features of interest. This resulted in four identifications of possible biomarkers. This was then followed up by two separate quantitative western blots to validate these results; however the quantitation could not be exactly replicated. Validation of these results using traditional biochemical techniques is a vitally important part of the experiment and is widely used (40;41). It will take some time, if ever, before label free relative quantitation is accepted as a stand-alone technique through which to discover and monitor changes in peptide abundance. It is likely, due to expense and the expert operation required that it will always remain as a tool for the research laboratory, and will not spread into the clinical field. However, these methods can still play an important role in the future of proteomic research into diseases such as hypertension and many others. It will not be possible to investigate the levels of certain molecules in biological samples through other techniques, and the application of proteomics to biomarker discovery is expected to continue to grow over the coming years. Following the perfection of techniques, such as the one described here, it will be simple to apply these to a wide variety, if not all, human diseases. There is a huge amount of knowledge to be obtained in this way, and the continuation of focussed research is vitally important to our continued improvements in treatment of disease.

It is concluded that label free relative quantitation is a valid tool for biomarker discovery, for a disease such as hypertension. Different types of statistical analysis can be used to target the results in a specific way, by identifying potential markers showing certain types of expression. This is a particularly useful type of analysis when there is more than two sample sets, as in this case. This method could be applied to any disease state, where multiple samples are available. As this is an unguided biomarker discovery method, no previous disease knowledge is required, making it a particularly useful and powerful technique.

5 Conclusions and Discussion

The work presented here has been broken down into three main chapters, namely:

- Biomarker Identification in Stroke Brain guided by MALDI-imaging
- Assessment of Denator Heat Treatment for Prevention of Proteomic Sample Degradation using Label Free Relative Quantitation
- Discovery and Identification of Biomarkers for Hypertension

Between these three chapters runs a joint theme, with all the work presented aiming to better proteomic research relating to biomarker discovery. This is with reference to sample preparation, the importance of experimental design (especially when working with limited samples) and the application of relevant and useful statistical methods to help mine into rich datasets.

The importance of investigating and learning about ischemic stroke was emphasised, and the current challenges facing researchers when attempting to gain identification of interesting up or down-regulated biomolecules seen in diseased tissue, using a fairly new technique known as MALDI-MSI. It is certainly not a straight forward process and obstacles were encountered. One of the important challenges in this work was the changes the disease made to the physicality of the tissue, causing global changes to protein levels in the samples used. The only way to overcome this issue is to use strict controls and meticulous experimental design. A further challenge in this part of the research was severe lack of sample availability, resulting in the work not being fully completed. This signifies the importance of precise experimental design, prior to practical experimentation.

Following on from this was the investigation into a new heat treatment device, aiming to prevent sample degradation. This is an important issue in any sort of research conducted using biological samples, as the researcher will wish to analyse the samples with as few artificial changes as possible. Enzymatic action causes large changes in samples following collection and storage to point of analysis. This degradation is important as it can give mis-leading results in any qualitative or, more importantly, quantitative proteomic analysis. In the past, the way to avoid this has been to store samples at a low temperature and to apply protease inhibitors, causing problems with mass spectrometric interference and possible

sample loss. Controlling the degradation processes will be of significant importance within many areas of proteomic and other types of biological research. A sample preservation step that is applied as early as possible after sampling was investigated as a possible solution to this problem. A denaturing device (Denator AB, Gothenburg, Sweden) was evaluated for sample preservation using mouse brain tissue as the subject. This instrument works using a combination of pressure and heat to rapidly and homogenously heat the tissue sample preventing proteolytic activity. It was proposed that this fast, homogenous heat distribution may give more effective, reproducible sample preservation. The use of heat-treatment on tissues immediately postmortem, to halt enzymatic degradation by denaturing both enzymes and proteins within tissue, was described here. To assess this device and process, in comparison with conventional snap freezing, LC-MS and Label Free Relative Quantitation were employed.

It was concluded that the heat treatment had little affect on the sub 10 kDa proteomic data. The features that did show significant changes between the frozen and heat treated groups showed a number of different types of changes, varying from showing feature preservation when the heat treatment is used, to sample degradation by the treatment. It was concluded that the heat treatment could potentially be a useful technique when particular targets cannot be detected using conventional snap freezing treatment. However, for analysis of the sub 10 kDa LC-MS proteomic data from mouse brain tissue sample studied here, it is not expected that the heat treatment system will become a popular sample preparation technique, due to the small change it demonstrates on the majority of the data and the additional expense it incurs to the research laboratory.

A further conclusion was that MS-MS identification of non-tryptically digested protein and peptide fragments can be challenging, as it was not possible to identify these features, using the techniques employed in this work. Additionally, it must be considered that the features under investigation were biomolecules such as lipids, and not proteins or peptides, as assumed. This could potentially be investigated in the future, along with the heat treatment affect on these types of molecule. Statistical analysis techniques were also investigated here, and the outcome of this was carried forward into further work conducted in the proceeding chapter regarding biomarkers for hypertension in rat plasma samples.

Hypertension was an interesting disease to study, due to its common occurrence and also its link with coronary heart disease, ventricular hypertrophy, heart failure, peripheral vascular disease, renal disease and atherosclerosis. To enable the prevention of these diseases, increased understanding of hypertension is paramount. Additionally, the

relationship with dietary salt is of interest, as the current opinion is that a high level of dietary salt is associated with increased blood pressure in later life. Previously conducted studies into biomarkers for hypertension have been limited in the fact that they have been guided by previous knowledge alone. As the occurrence of, and possible predisposition to, hypertension is not fully understood this must have limited possible biomarker discovery significantly. It was proposed here that biomarkers for hypertension are currently not particularly well characterised. In my research, biomarkers for hypertension were investigated in a completely unguided manner, which is not a common practice. A label free relative quantitation method was again used, this time to examine differences between diseased and healthy samples, mainly concentrating on plasma samples. In this work a rat model was used, and the effect of salt treatment was also investigated. The advantages and disadvantages of the use of an animal model were also discussed, with the main reason, in this case, for using an animal model being the reduction of sample heterogeneity. Additionally, the importance of experimental design was discussed, with emphasis put on the avoidance of introducing bias into an experimental study such as this. In this work plasma samples were used, and the challenges and issues of this type of sample were also discussed. The main challenge of analysing plasma samples was stated as being the the extremely high dynamic range of the samples and the high abundance levels of a small number of proteins. Different types of sample depletion and enrichment were considered and investigated. The efficiency of two different plasma partitioning strategies was evaluated, using two types of analysis, a traditional 2D mini gel and an LC-MS analysis. It was found that while plasma partitioning may have a place as a possible sample preparation method, neither antibody affinity partitioning or size separation depletion are particularly appropriate for use in a quantitative biomarker discovery project. The main problem with this type of sample preparation is the possibility of introducing artificial differences into the samples, in turn giving potentially mis-leading results. Additionally, important parts of the sample may be lost using this type of sample preparation. Hence, plasma depletion was not used in this work.

Repeatability and reproducibility was emphasised as being a vitally important part of the biomarker discovery process, and it was proposed that this should not be underestimated as, perhaps, one of the most important parts of the research. This was fully investigated prior to collection of biological replicate samples, and should ensure consistent results throughout the work. Another important part of this work is the selection of the most appropriate statistical method, and this was also fully discussed and investigated. The methods used were pairwise t-tests and RankProduct analysis, developed for use with microarray data.

The outcome of this work was that several features were found to be differentially regulated between diseased and healthy animals. These were subsequently identified as being from the following proteins:

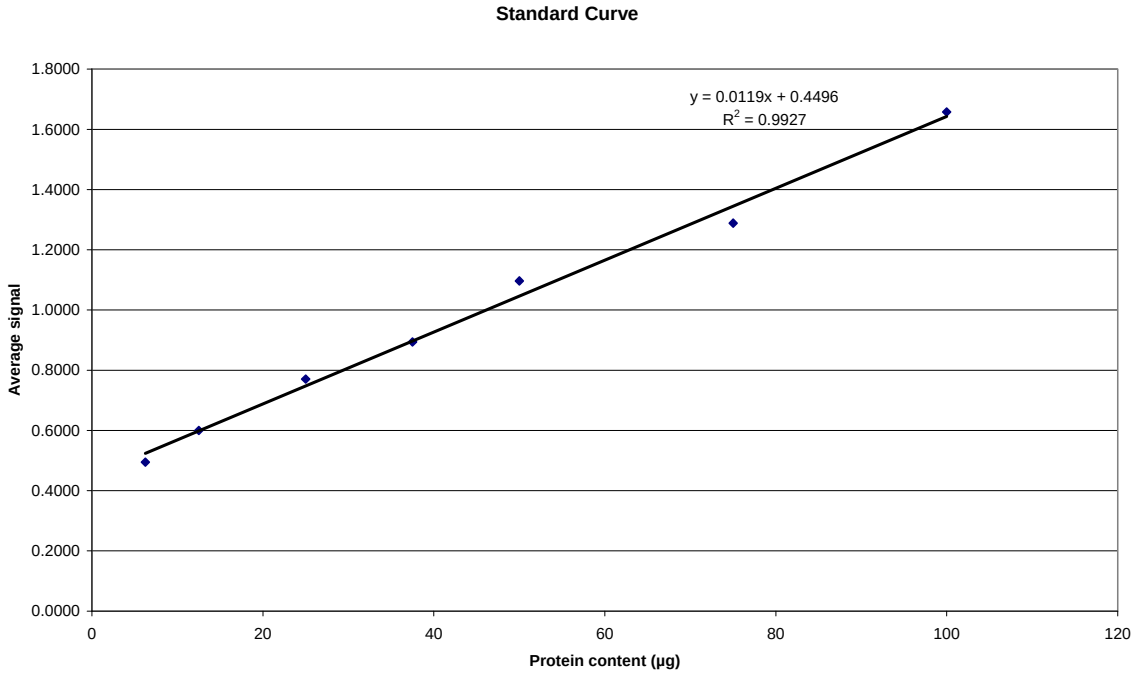
- Hemopexin
- Fibrinogen alpha-1-isoform
- Pregnancy zone protein (alpha-1-macroglobulin)
- Contrapsin-like-protease inhibitor

The biological significance of these markers was summarised in chapter four. A future possible project will be further investigation into the significance of these, and possibly even their occurrence in humans.

Ultimately, however, this project has not been a biological one. The development and perfection of techniques has been the main goal throughout, and the biological samples applied are somewhat irrelevant. The perfected proteomic biomarker discovery and identification project could therefore be applied to any disease of interest in the future.

Appendix 1 – Biomarker Identification in Stroke Brain Guided by MALDI imaging

The graph below shows the standard curve used for determination of protein concentration in mouse brain tissue, as discussed in section 2.4.2.



Appendix 2 – Biomarker Identification in Stroke Brain Guided by MALDI imaging

The following table shows 67 proteins identified from LC-MS-MS data searched using MASCOT version 2.1 against the NCBInr database 20070413 as detailed in section 2.3. In addition, selected mascot results for the identified proteins are shown below the table.

Proteins Identified	Gi Identification Number from NCBInr 20070413	Mass (Da)	Mowse Score
Beta 1 Globin	gi/4760586	15699	491
Hemoglobin alpha chain	gi/122421	15076	609
Complexin 1 Mouse	gi/31542416	15140	167
Microtubule associated protein 2, isoform 2	gi/68341935	49306	342
Predicted: Solute carrier family 25 (mitochondrial carrier, adenine nucleotide translocator)	gi/82917337	27960	119
ATP Synthase, H+ transporting, mitochondrial FO Complex, subunit F	gi/7949005	12489	66
FABP-I, fatty acid binding protein (N-terminal)	gi/404383	3733	65
Hemoglobin beta chain complex	gi/ 158969	15969	144
Ig Heavy Chain V region	gi/90771	12993	65
Peptidylproyl Isomerase A	gi/8394009	17863	91
Hemoglobin alpha 2 chain	gi/60678292	15275	240
Myelin Basic Protein	gi/293725	3956	82
Myelin Basic Protein	gi/4454311	14184	230
Cytochrome C Oxidase Subunit Viic	gi/6680991	7328	86
Eukaryotic translation initiation factor 5A	gi/4503545	16821	65
Polyglutamine containing protein	gi/60223071	81444	86
Atp5b protein	gi/17390926	10152	65
37kd Protein	gi/205964	7428	125
Mitochondrial malate dehydrogenase 2, NAD	gi/89574115	31657	190
DJ – 1 Protein-Rat	gi/16924002	19961	81
H(+) – ATP Synthase subunit e (N terminal)-rat	gi/258788	4964	79
Migration Inhibitory Factor	gi/191491	11877	68
MHC Class II Antigen – rat	gi/18958188	6937	68
Atp2b3 protein	gi/111600317	121813	52
Aspartate aminotransferase	gi/871422	46203	149
Predicted Similar to Ubiquitin – protein ligase EDDI – mouse	gi/94399459	333979	64
Chain A, Macrophage Migration Inhibitory Factor Y95f Mutant	gi/5822092	12349	102
Phosphoglycerate mutase, p29	gi/235503	1328	84
Gln Synthetase	gi/228136	40411	78
Actin, gamma, cytoplasmic	gi/123298588	16723	65
Profilin 1	gi/56206029	11813	98

Actin, gamma, cytoplasmic	gi/123298588	16723	87
Triosephosphate Isomerase	gi/538426	26904	70
Ubiquitin carboxyl-terminal hydrolase isozyme L1 (UCH-L1)	gi/1717867	1815	67
Rod Outer segment membrane protein 1 (ROSP1)	gi/417693	37242	65
Claudin 11	gi/6679186	22099	65
Predicted Similar to Glyceraldehyde-3-phosphate dehydrogenase	gi/109484558	35760	104
Predicted similar to NADH-ubiquinone oxidoreductase MLRQ subunit	gi/109471761	8984	82
Cytochrome c oxidase, subunit 7a 3 –rat	gi/11968072	9347	64
GAPDH protein	gi/66364570	14920	155
Ubiquitin carboxy-terminal hydrolase L1	gi/6755929	24822	155
Predicted similar to glyceraldehyde-3-phosphate dehydrogenase	gi/51769013	35807	149
Pyruvate kinase	gi/16757994	57781	145
Creatine Kinase-B	gi/203476	40598	141
Pgam1 Protein	gi/45767854	22401	138
Microtubule associated protein 2	gi/6981182	198446	131
Unnamed protein product	gi/12842521	9854	114
Enolase 2, gamma	gi/26023949	47111	83
Choline transporter-like protein 2 (solute carrier family 44 member 2, inner ear supporting cell a)	gi/73918927	79918	77
Glyoxylase 1 –rat	gi/46485429	20806	75
Beta Actin	gi/28950674	11856	140
Creatine Kinase, Brain	gi/31542401	42685	111
Actin Cytoplasmic, 1 (Beta actin)	gi/1351867	41711	99
Mlrq like protein	gi/1401252	8509	72
Macrophage migration inhibitory factor	gi/694108	12538	65
Myelin Basic Protein Isoform 6	gi/69885073	14202	85
ATP Synthase, H+ Transporting, mitochondrial F1 complex, beta subunit	gi/54792127	56318	143
DJ-1 Protein-rat	gi/16924002	19961	78
Pyruvate kinase muscle –rat	gi/16757994	57781	76
Macrophage Migration inhibitory factor homolog –rat	gi/285058	2904	74
VN Precursor – rat	gi/1334229	13072	66
Beta globin	gi/193767	5687	135
Calm2 protein	gi/14715123	11067	70
Myelin basic protein isoform 5	gi/8393759	14202	91

Ulip2 Protein	gi/1915913	62132	95
Alpha Globin	gi/30027750	9346	133

Selected mascot results for the identified proteins are shown below:

1. gi|4760586 **Mass:** 15699 **Score:** 491 **Queries matched:** 15

beta-1-globin [Mus musculus]

Check to include this hit in error tolerant search or archive report

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
59	439.1238	876.2330	876.4745	-0.2415	0	26	5.1e+02	2	L.LVVYPWT.Q
67	446.6028	891.1911	891.4490	-0.2580	0	27	3.9e+02	1	L.VVYPWTQ.R
141	503.1261	1004.2376	1004.5331	-0.2955	0	31	1.8e+02	1	L.LVVYPWTQ.R
225	606.1234	1210.2322	1210.6135	-0.3813	0	33	1.3e+02	1	L.VVYPWTQRY.F
238	414.1095	1239.3066	1239.6459	-0.3392	0	53	1.4	1	-.VHLTDAEKAAVS.G
273	662.6294	1323.2443	1323.6975	-0.4532	0	45	8.7	1	L.LVVYPWTQRY.F
274	662.6752	1323.3359	1323.6975	-0.3616	0	(30)	2.3e+02	7	L.LVVYPWTQRY.F
275	662.6752	1323.3359	1323.6975	-0.3616	0	(31)	2.1e+02	1	L.LVVYPWTQRY.F
276	662.6752	1323.3359	1323.6975	-0.3616	0	(33)	1.2e+02	4	L.LVVYPWTQRY.F
355	737.1499	1472.2853	1470.7659	1.5194	0	32	2e+02	1	L.LVVYPWTQRYF.D
393	515.1295	1542.3666	1542.8001	-0.4335	0	52	1.8	1	G.KVNSDEVGGEALGRL.L
415	793.6739	1585.3333	1585.7929	-0.4596	0	56	0.69	1	L.LVVYPWTQRYFD.S
450	552.8042	1655.3909	1655.8842	-0.4933	0	33	1.4e+02	1	G.KVNSDEVGGEALGRLL.V
499	596.1339	1785.3797	1785.9009	-0.5211	0	124	1.4e-07	1	L.WGKVNSDEVGGEALGRL.L
640	842.5000	2524.4782	2525.3600	-0.8819	0	95	0.00014	1	E.KAAVSGLWGKVNSDEVGGEALGRLL.V

gi|122441 **Mass:** 15076 **Score:** 609 **Queries matched:** 15

Hemoglobin subunit alpha (Hemoglobin alpha chain) (Alpha-globin)

Check to include this hit in error tolerant search or archive report

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
338	437.8012	1310.3817	1310.7194	-0.3376	0	51	1.2	1	L.ASVSTVLTSKYR.-
374	475.4711	1423.3915	1423.8034	-0.4119	0	21	1.3e+03	8	F.LASVSTVLTSKYR.-
484	564.1250	1689.3532	1689.8191	-0.4659	0	48	2.6	1	M.FASFPTTKTYFPHF.D
521	602.4494	1804.3265	1804.8460	-0.5195	0	48	2.3	1	M.FASFPTTKTYFPHFD.V
583	504.1343	2012.5079	2013.0068	-0.4989	0	32	1.1e+02	1	T.LASHMPADFTPAVHASLDK.F
616	554.1437	2212.5456	2213.1229	-0.5773	0	18	3e+03	9	L.VTLASHMPADFTPAVHASLDK.F
634	582.3889	2325.5265	2326.2069	-0.6805	0	54	0.75	1	L.LVTLASHMPADFTPAVHASLDK.F
669	652.2000	2604.7709	2605.2924	-0.5216	0	40	17	1	L.ASHMPADFTPAVHASLDKFLASVST.V
676	680.4029	2717.5824	2718.3765	-0.7941	0	40	19	1	T.LASHMPADFTPAVHASLDKFLASVST.V
683	730.4266	2917.6771	2918.4926	-0.8155	0	37	39	1	L.VTLASHMPADFTPAVHASLDKFLASVST.V
685	733.4471	2929.7591	2930.5290	-0.7698	0	59	0.22	1	T.LASHMPADFTPAVHASLDKFLASVSTVL.T
688	758.6805	3030.6931	3031.5767	-0.8836	0	32	1.2e+02	1	L.LVTLASHMPADFTPAVHASLDKFLASVST.V
694	803.9322	3211.6999	3212.6213	-0.9214	0	93	8.6e-05	1	M.VLSGEDKSNIKAAWGKIGGHGAEYGAEALER.M
696	811.7199	3242.8505	3243.7291	-0.8786	0	36	48	1	L.LVTLASHMPADFTPAVHASLDKFLASVSTVL.T
701	836.7000	3342.7709	3343.6618	-0.8909	0	103	1e-05	1	M.VLSGEDKSNIKAAWGKIGGHGAEYGAEALERM.F

gi|68341935 **Mass:** 49306 **Score:** 342 **Queries matched:** 5

microtubule-associated protein 2 isoform 2 [Mus musculus]

Check to include this hit in error tolerant search or archive report

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
569	666.1837	1995.5292	1996.0840	-0.5548	0	74	0.015	1	E.SPQLATLAEDVTAALAKQGL.-
583	690.1839	2067.5298	2068.1051	-0.5754	0	102	2.5e-05	1	L.LESPQLATLAEDVTAALAKQ.G
599	709.1851	2124.5335	2125.1266	-0.5931	0	70	0.035	1	L.LESPQLATLAEDVTAALAKQG.L
600	709.1851	2124.5335	2125.1266	-0.5931	0	52	2.7	1	L.ESPQLATLAEDVTAALAKQGL.-
611	746.8794	2237.6163	2238.2107	-0.5944	0	104	1.4e-05	1	L.LESPQLATLAEDVTAALAKQGL.-

gi|89574115 **Mass:** 31657 **Score:** 190 **Queries matched:** 4

mitochondrial malate dehydrogenase 2, NAD [Mus musculus]

☐ Check to include this hit in error tolerant search or archive report

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑ 339	724.1648	1446.3151	1446.7024	-0.3872	0	54	1.1	1	F.SLVDAMNGKEGVVE.C
☑ 384	762.7167	1523.4189	1523.8671	-0.4482	0	54	1.2	1	N.AKVAVLGASGGIGQPLS.L
☑ 441	546.4952	1636.4637	1636.9511	-0.4874	0	28	4.9e+02	1	N.AKVAVLGASGGIGQPLSL.L
☑ 655	740.4990	2957.9667	2958.7592	-0.7925	0	98	8.5e-05	1	N.AKVAVLGASGGIGQPLSLLLKNSPLVSRLT.L

gi|871422 **Mass:** 46203 **Score:** 149 **Queries matched:** 4

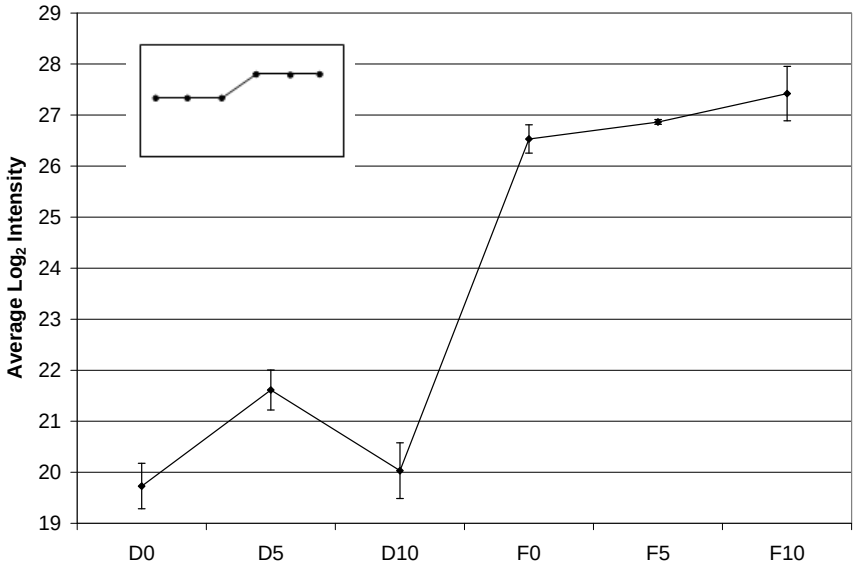
aspartate aminotransferase [Mus musculus]

☐ Check to include this hit in error tolerant search or archive report

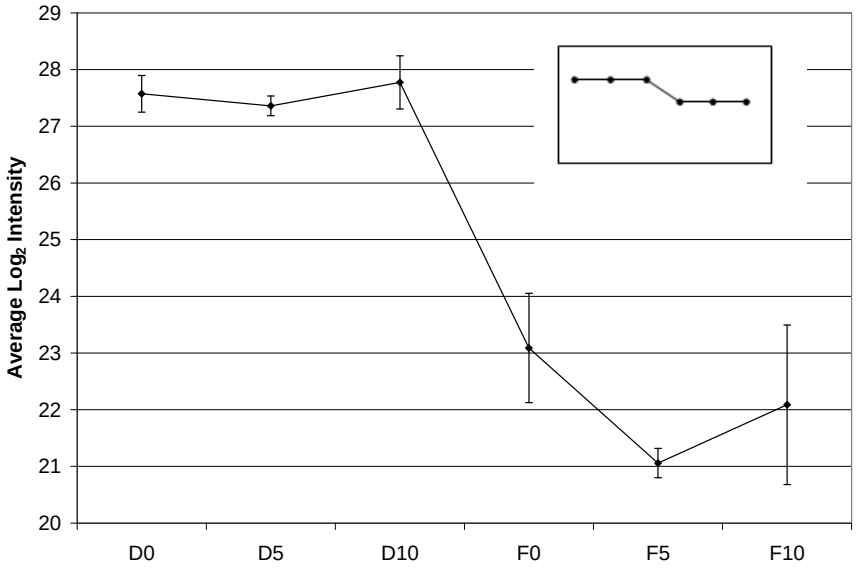
Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑ 434	539.8192	1616.4357	1616.8926	-0.4568	0	63	0.17	1	M.APPSVFAQVPQAPPVL.V
☑ 524	621.8510	1862.5311	1863.0294	-0.4983	0	31	2.8e+02	1	M.APPSVFAQVPQAPPVLVF.K
☑ 566	664.5315	1990.5727	1991.1243	-0.5517	0	39	46	1	M.APPSVFAQVPQAPPVLVFK.L
☑ 591	702.2136	2103.6189	2104.2084	-0.5895	0	68	0.064	1	M.APPSVFAQVPQAPPVLVFKL.T

Appendix 3 – Assessment of Heat Treatment for Prevention of Proteomic Sample Degradation using Label Free Relative Quantitation

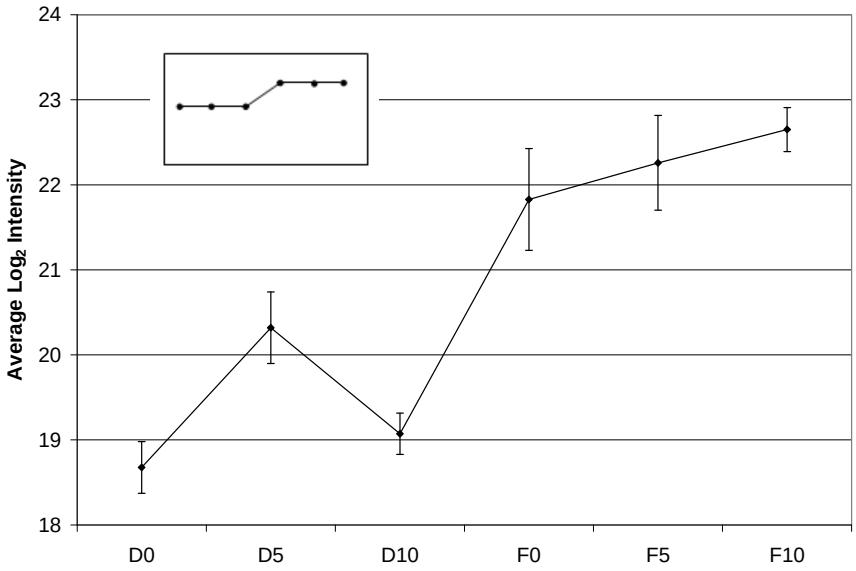
The figures shown below demonstrate the expression profiles exhibited by the 50 features with p-values ≤ 0.001 from the sub 10kDa intact samples, as discussed in section 3.4.2.1. From visual examination of the profiles it is demonstrated that 7 show expression profile as predicted by figure 3.2 C and 25 show expression profile as predicted by figure 3.2 D.



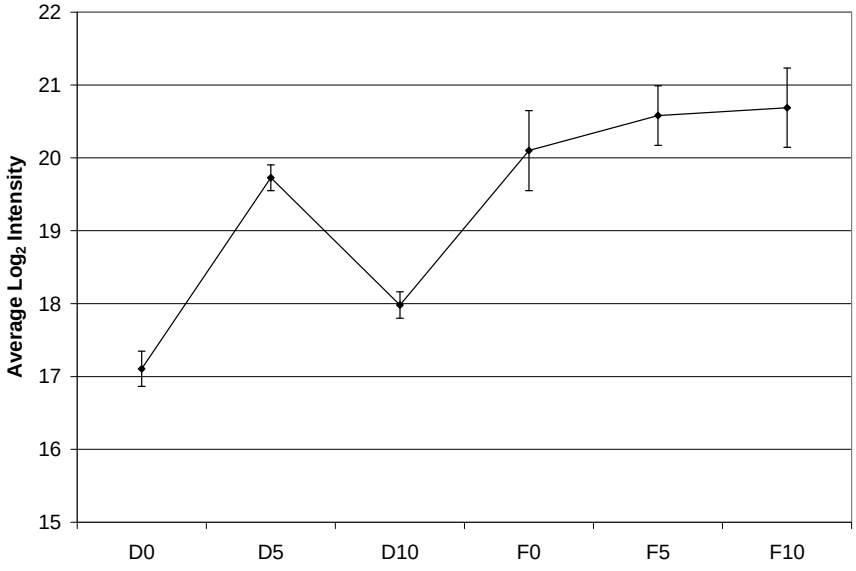
Expression profile from significant feature 1 from sub 10kDa intact samples



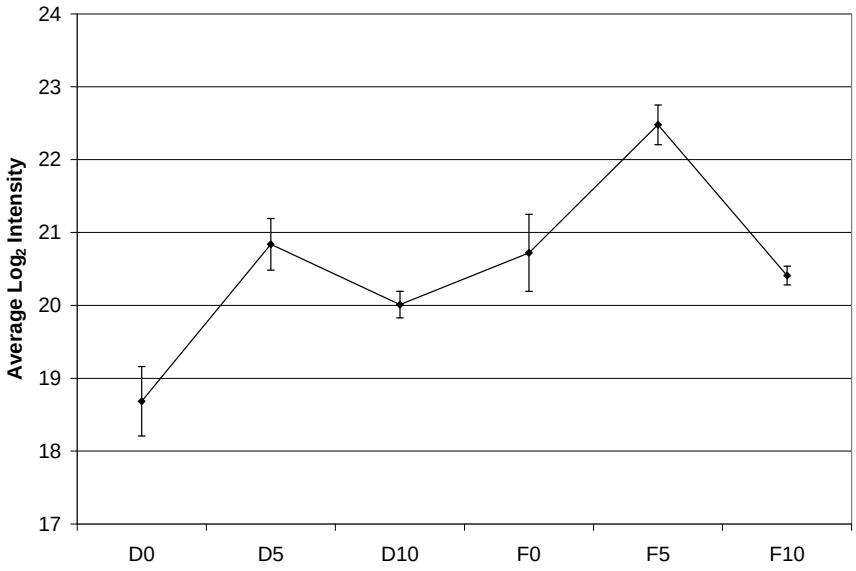
Expression profile from significant feature 2 from sub 10kDa intact samples



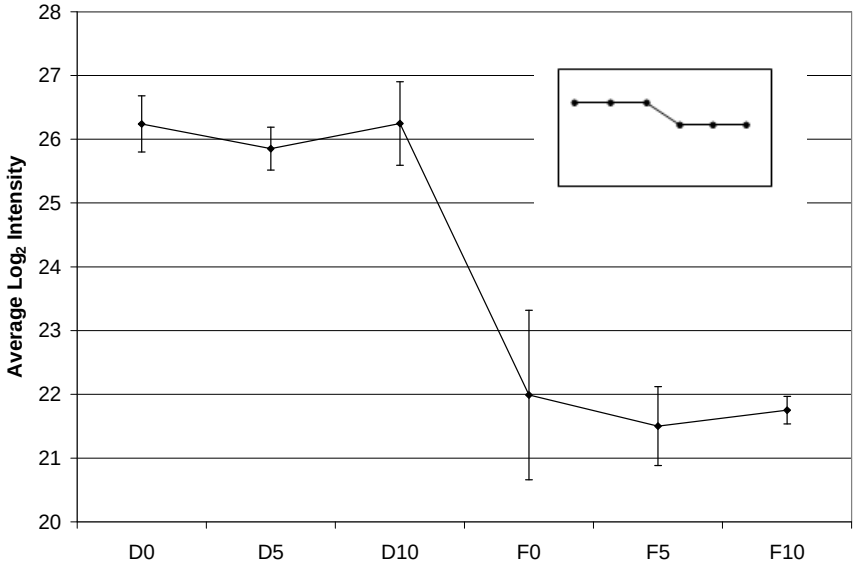
Expression profile from significant feature 3 from sub 10kDa intact samples



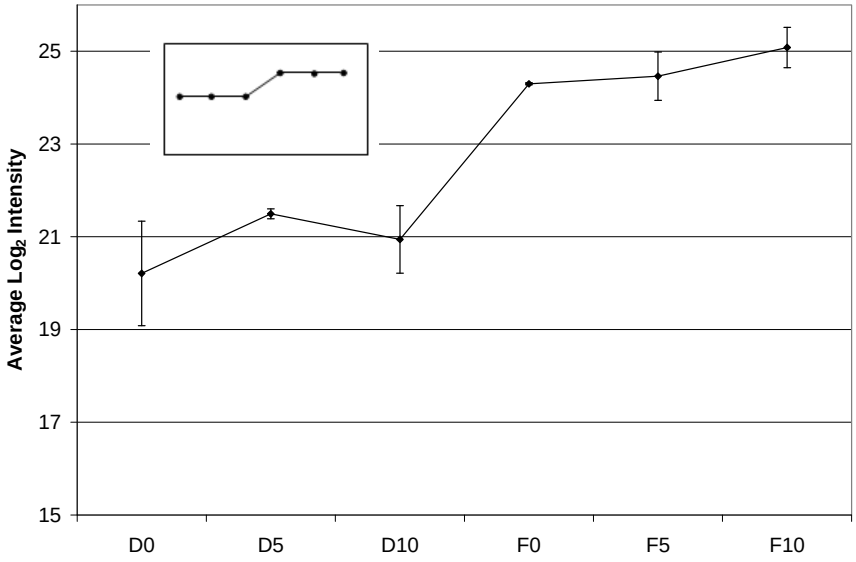
Expression profile from significant feature 4 from sub 10kDa intact samples



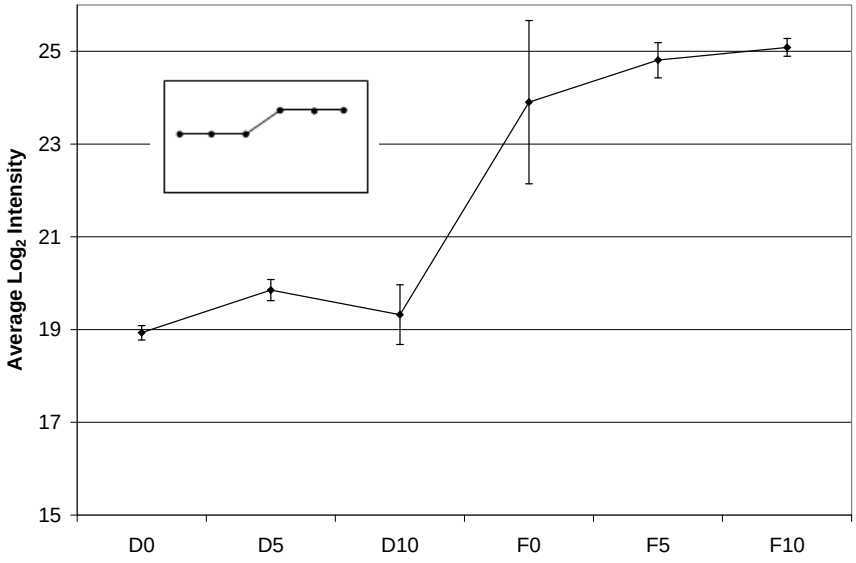
Expression profile from significant feature 5 from sub 10kDa intact samples



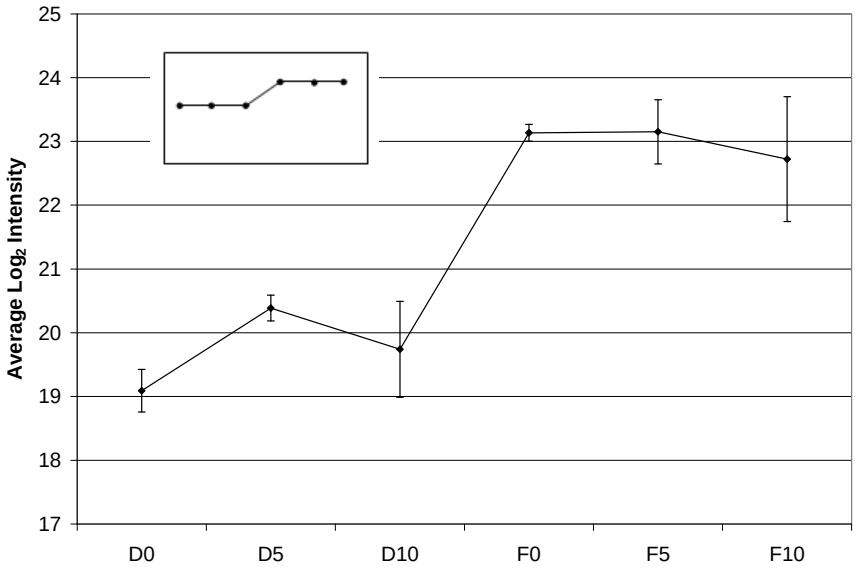
Expression profile from significant feature 6 from sub 10kDa intact samples



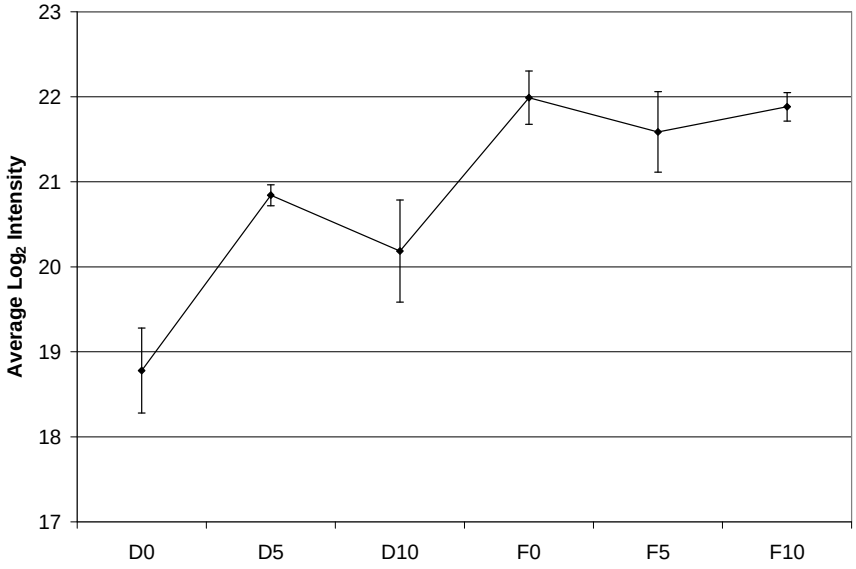
Expression profile from significant feature 7 from sub 10kDa intact samples



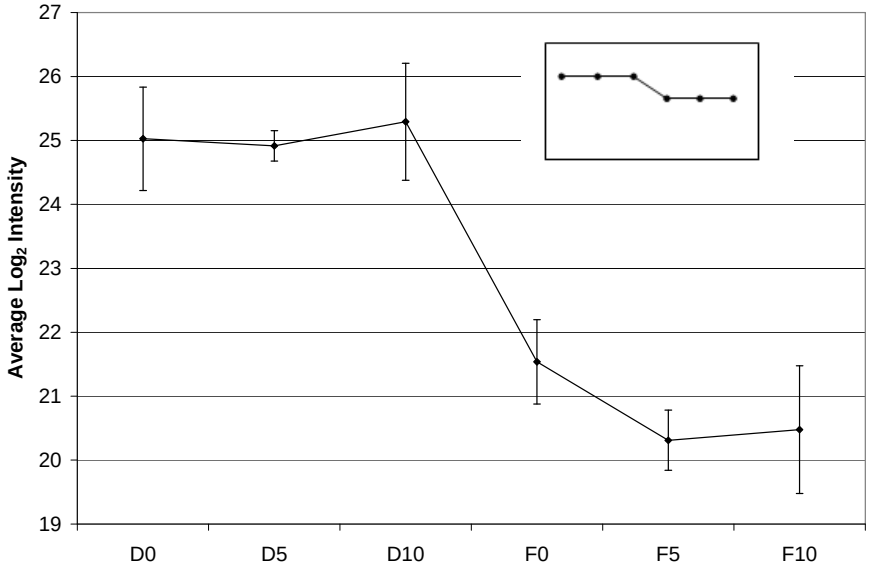
Expression profile from significant feature 8 from sub 10kDa intact samples



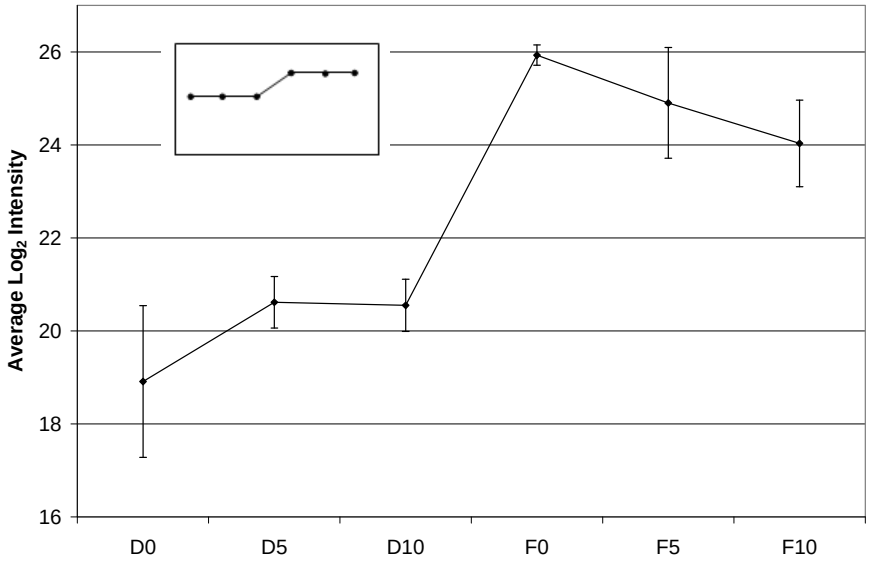
Expression profile from significant feature 9 from sub 10kDa intact samples



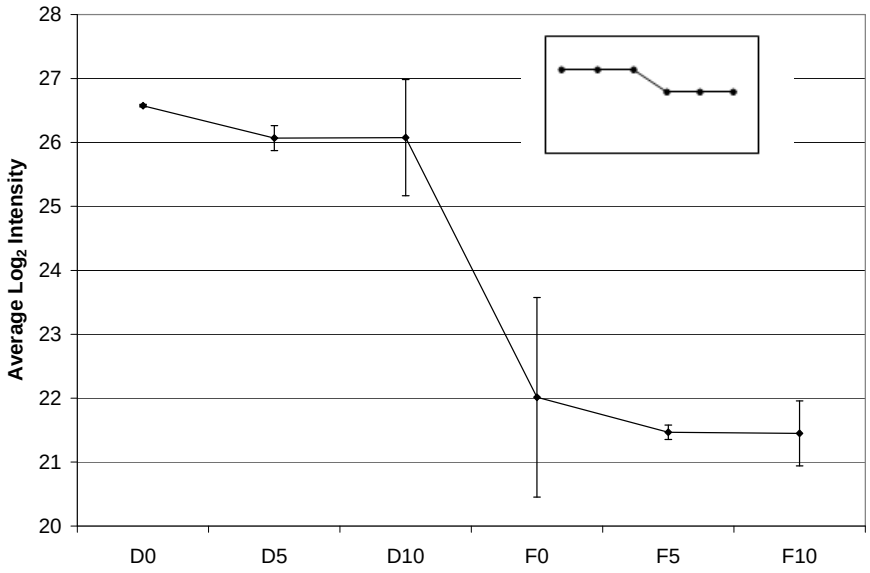
Expression profile from significant feature 10 from sub 10kDa intact samples



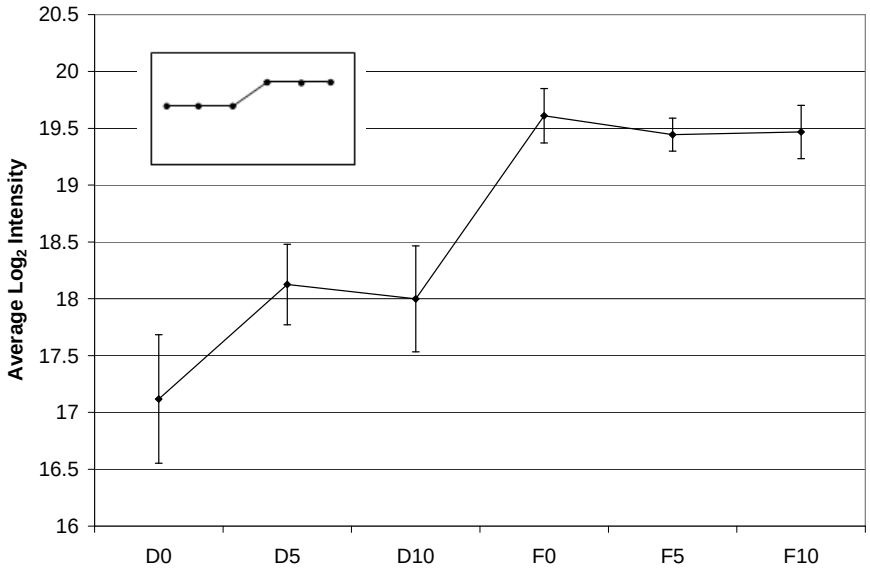
Expression profile from significant feature 11 from sub 10kDa intact samples



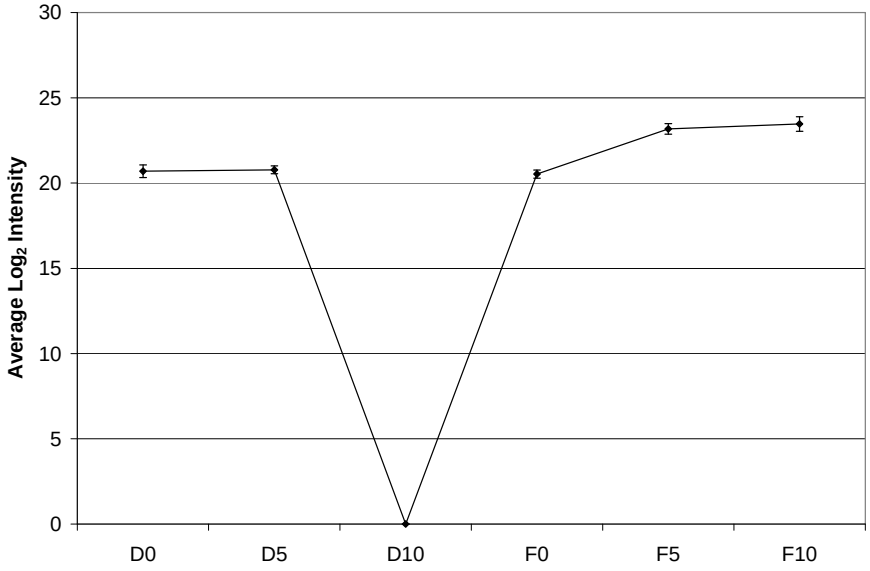
Expression profile from significant feature 12 from sub 10kDa intact samples



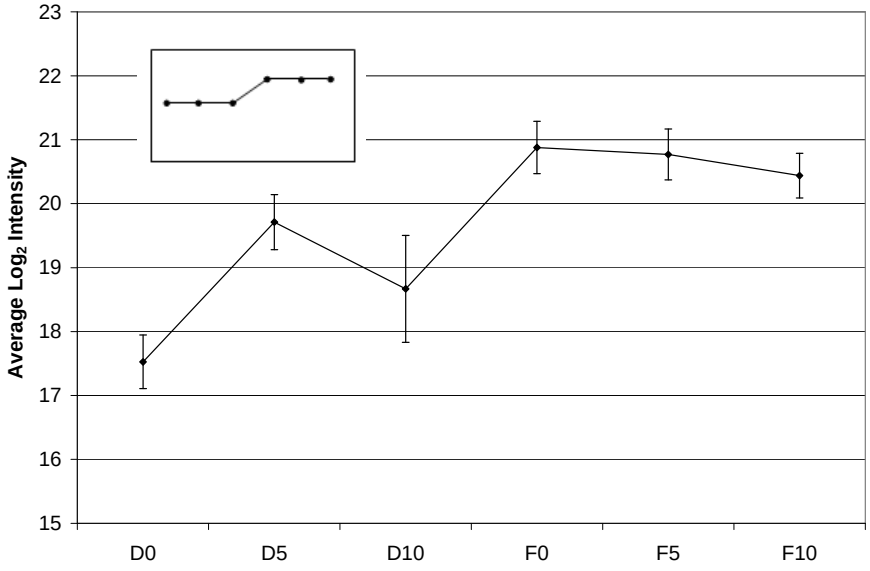
Expression profile from significant feature 13 from sub 10kDa intact samples



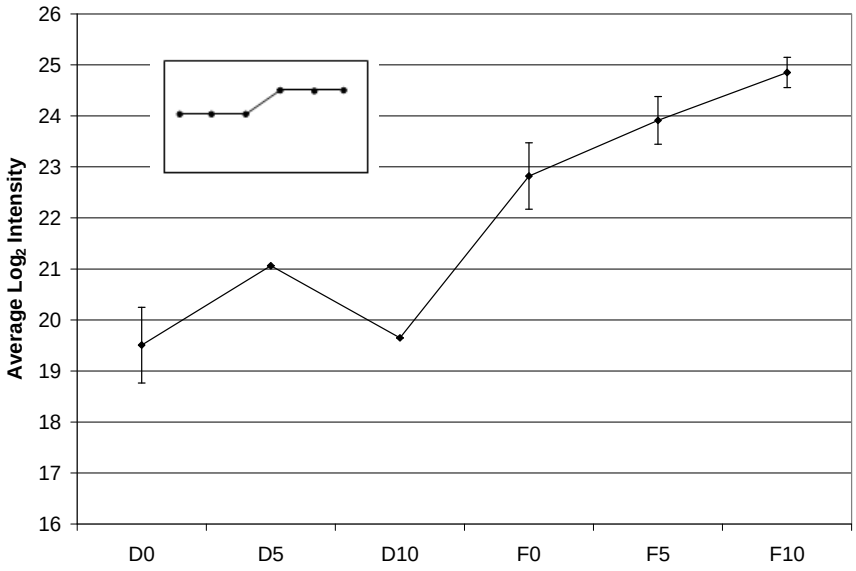
Expression profile from significant feature 14 from sub 10kDa intact samples



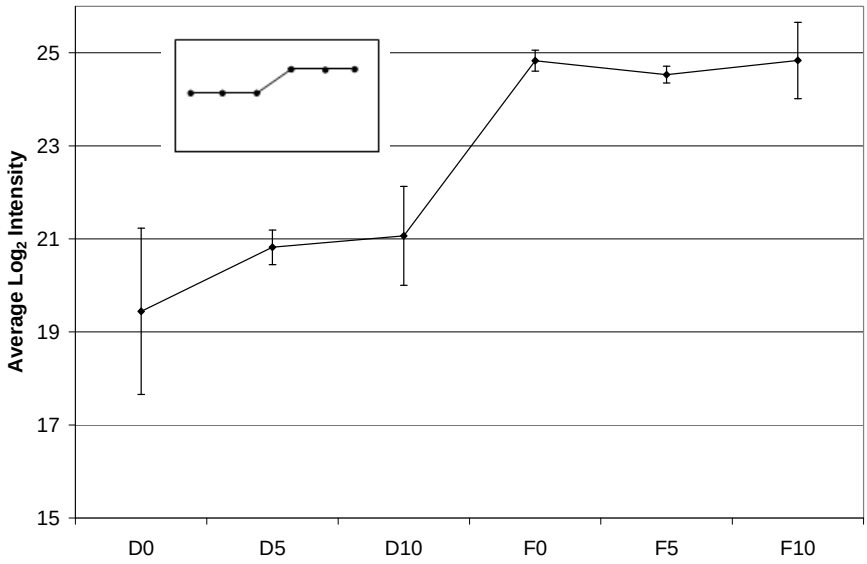
Expression profile from significant feature 15 from sub 10kDa intact samples



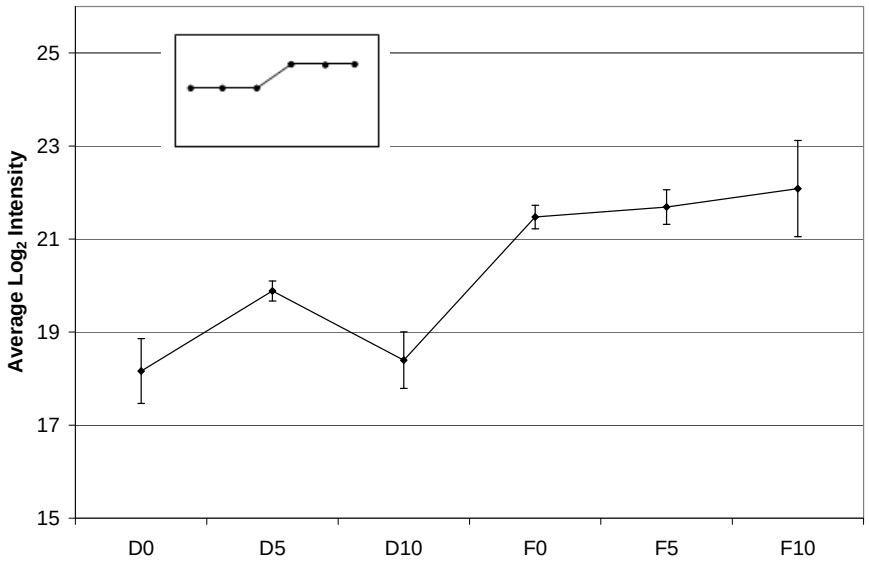
Expression profile from significant feature 16 from sub 10kDa intact samples



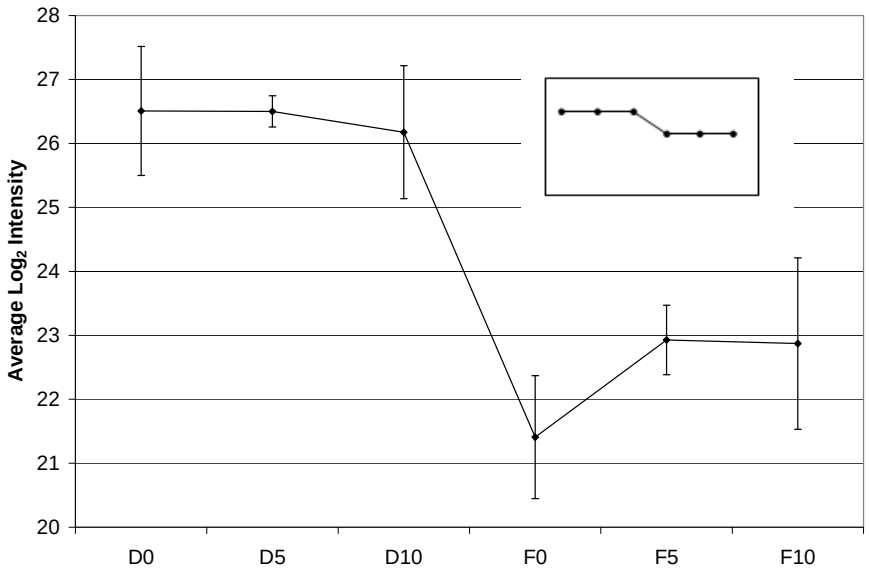
Expression profile from significant feature 17 from sub 10kDa intact samples



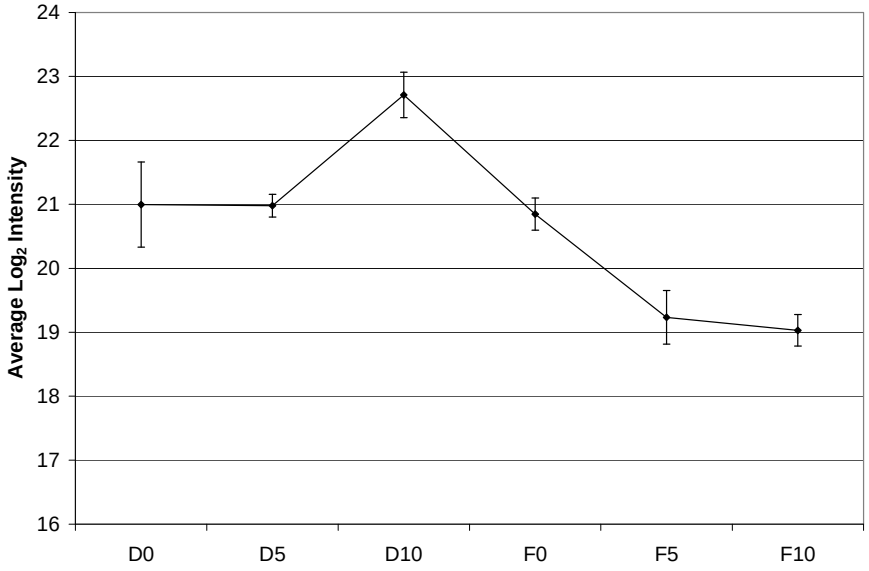
Expression profile from significant feature 18 from sub 10kDa intact samples



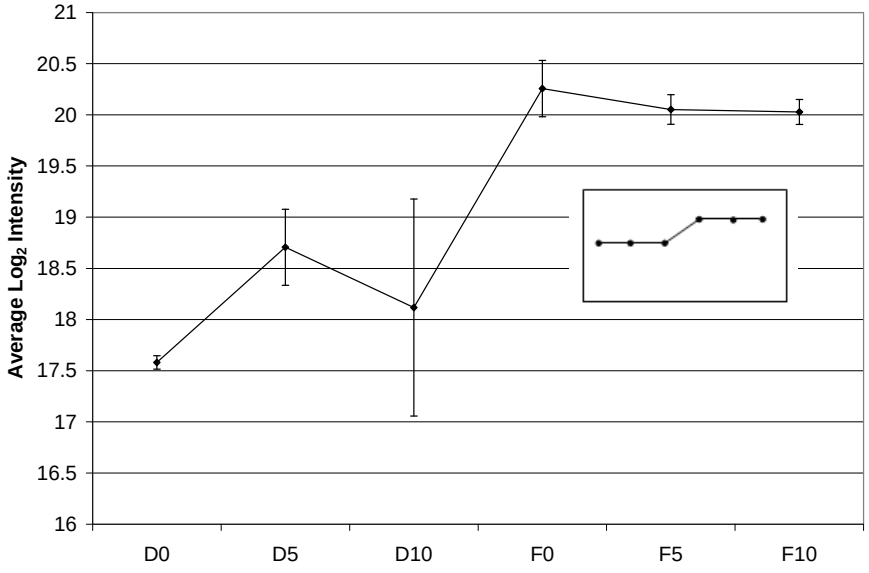
Expression profile from significant feature 19 from sub 10kDa intact samples



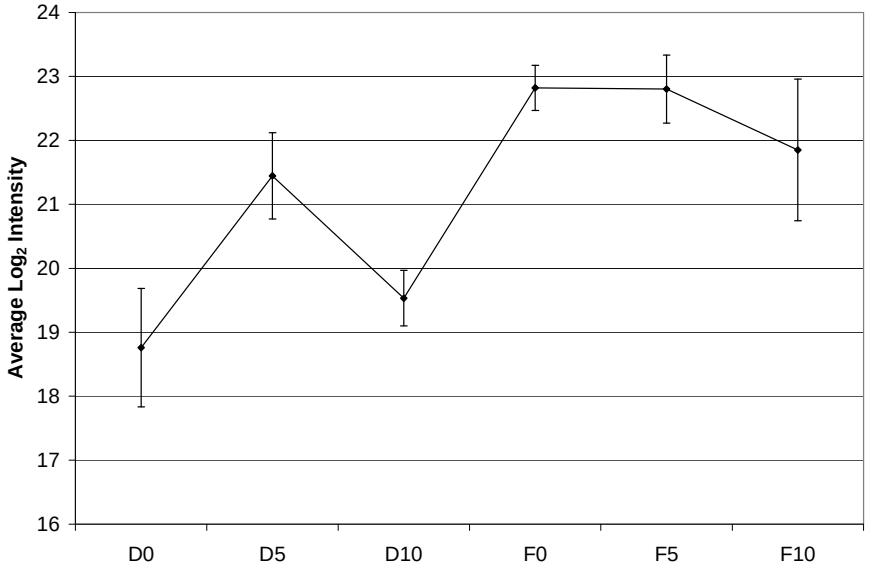
Expression profile from significant feature 20 from sub 10kDa intact samples



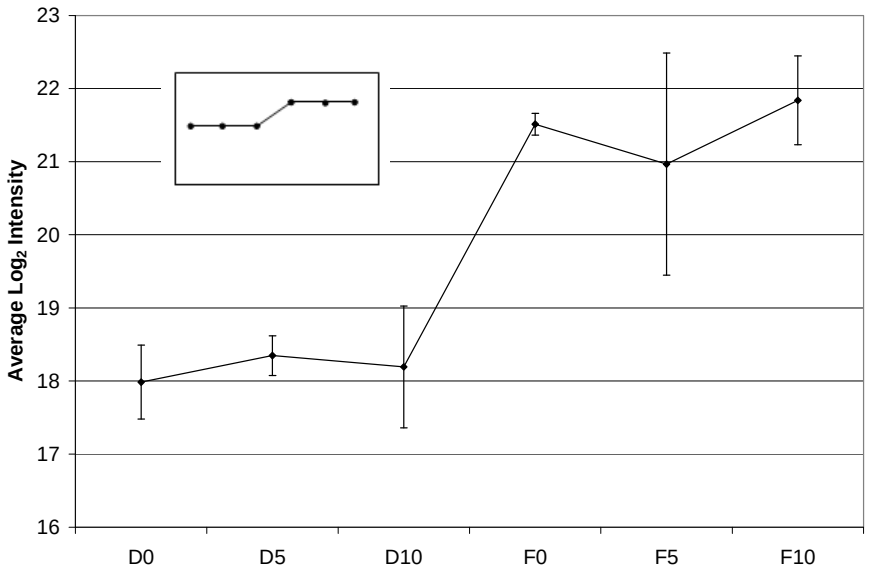
Expression profile from significant feature 21 from sub 10kDa intact samples



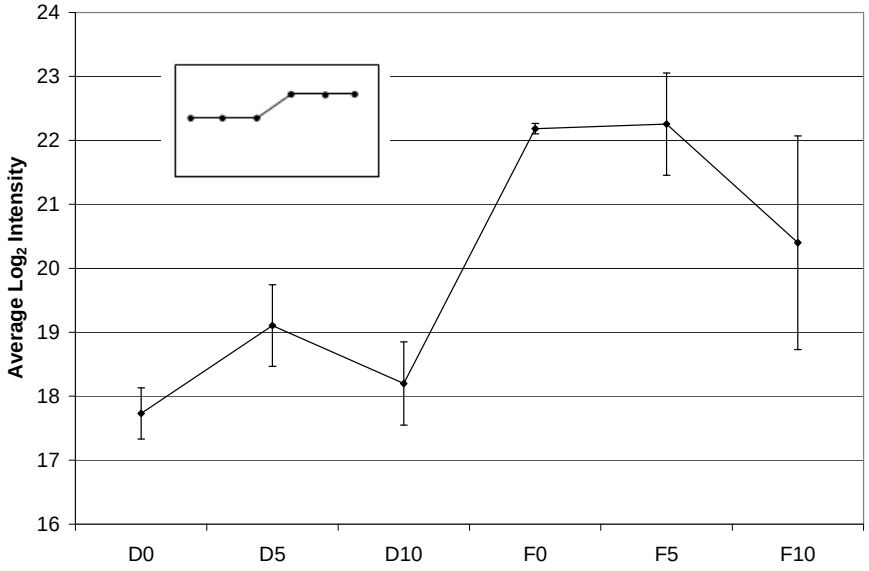
Expression profile from significant feature 22 from sub 10kDa intact samples



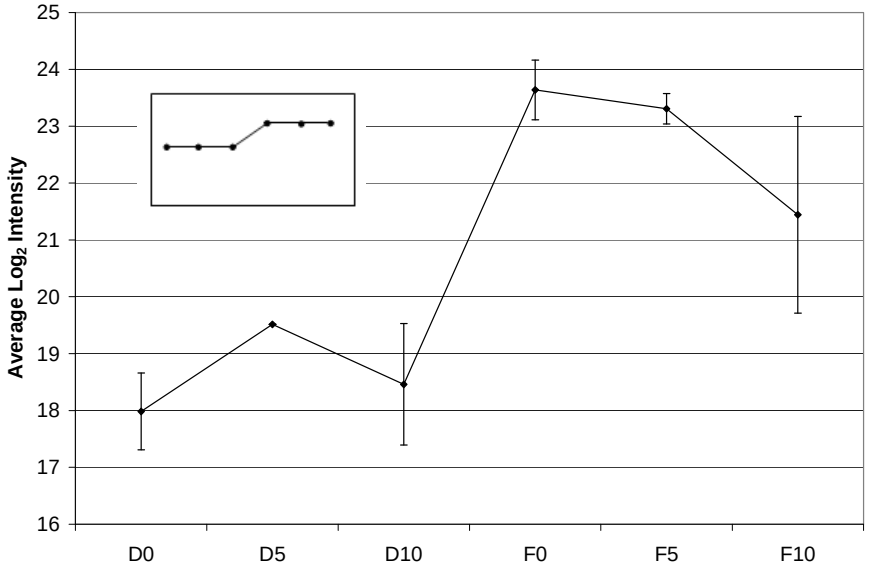
Expression profile from significant feature 23 from sub 10kDa intact samples



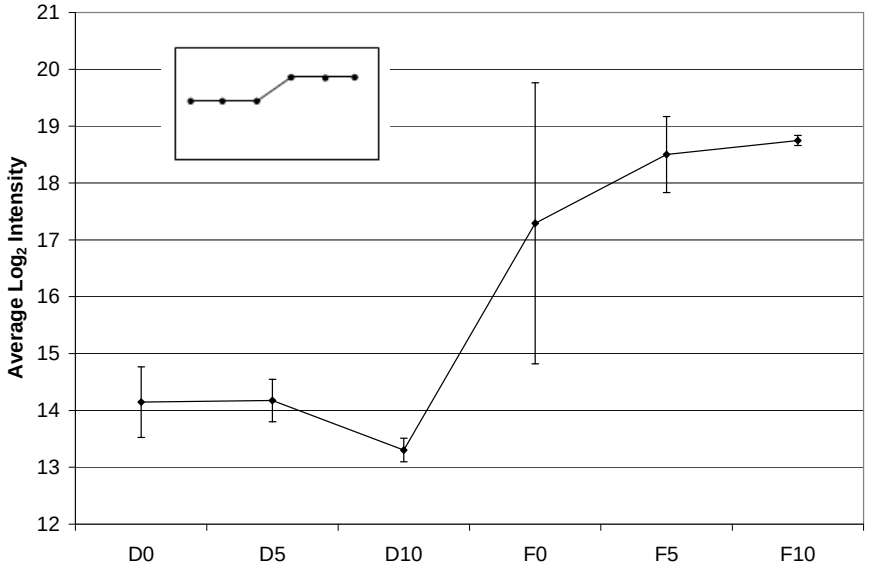
Expression profile from significant feature 24 from sub 10kDa intact samples



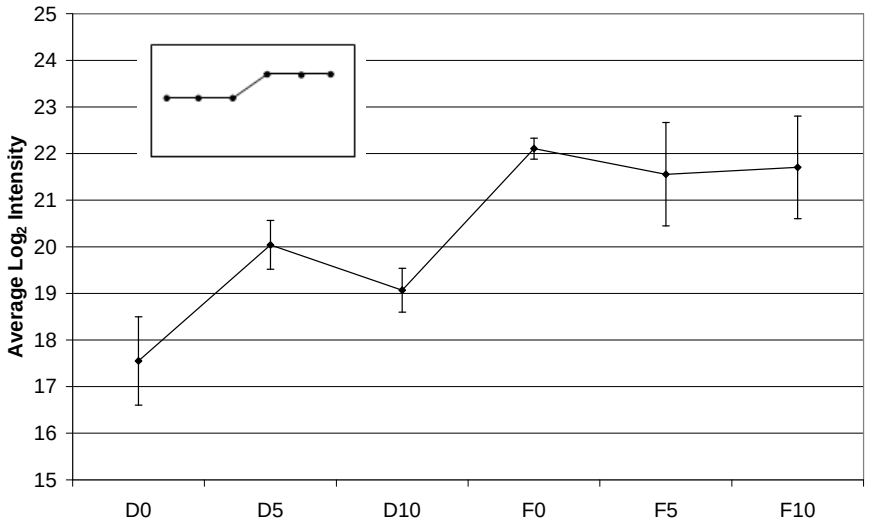
Expression profile from significant feature 25 from sub 10kDa intact samples



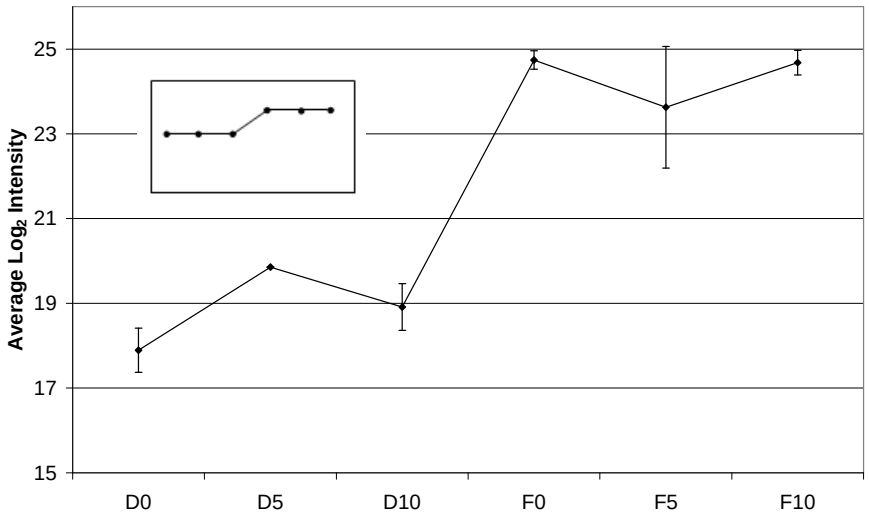
Expression profile from significant feature 26 from sub 10kDa intact samples



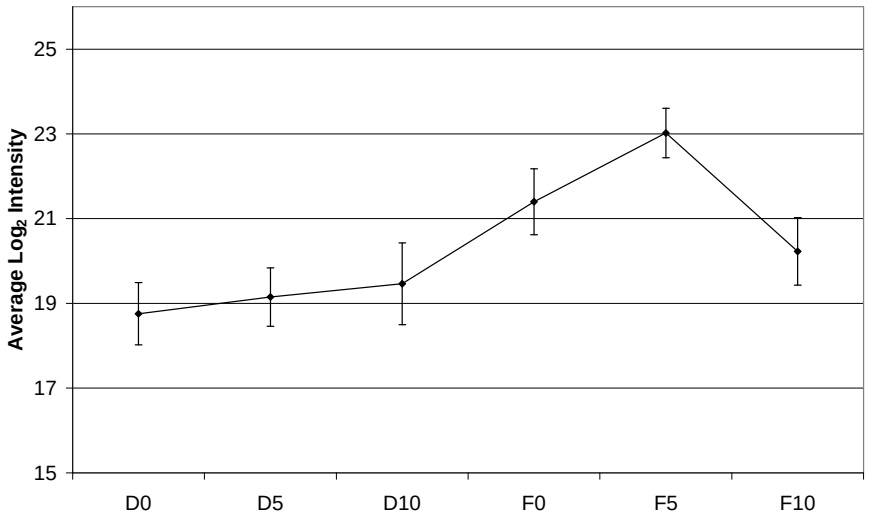
Expression profile from significant feature 27 from sub 10kDa intact samples



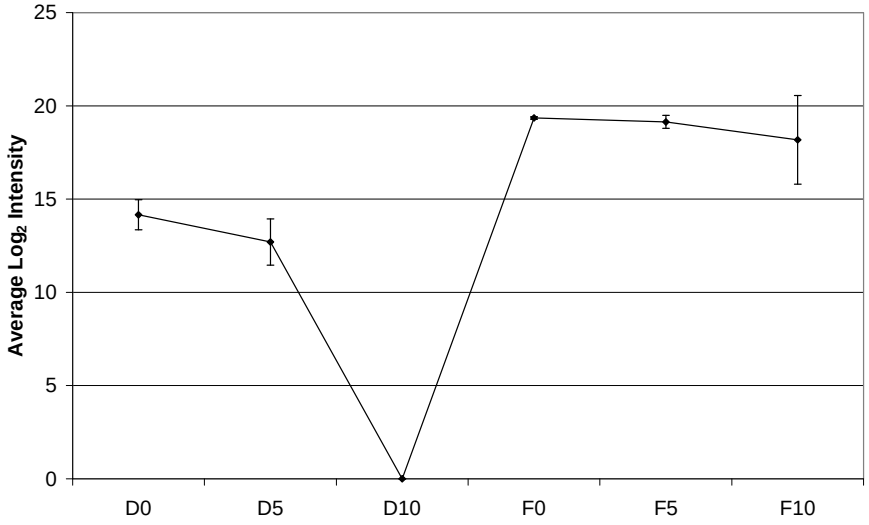
Expression profile from significant feature 28 from sub 10kDa intact samples



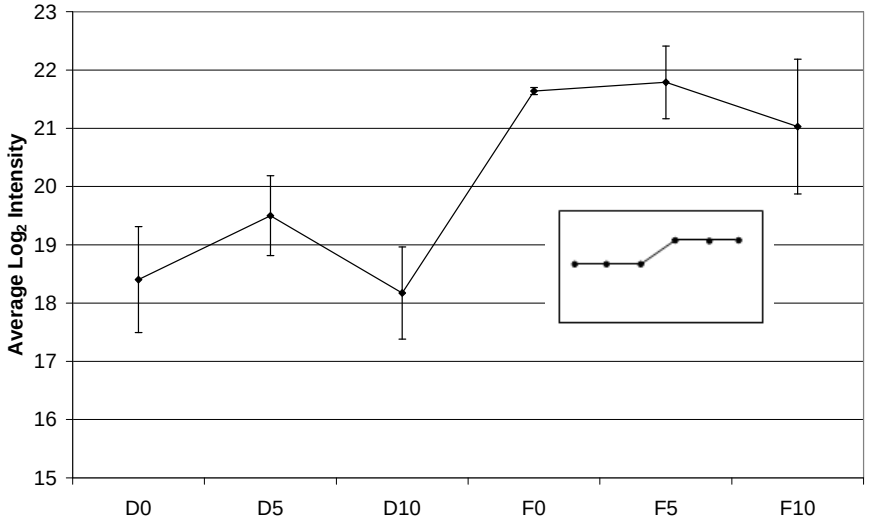
Expression profile from significant feature 29 from sub 10kDa intact samples



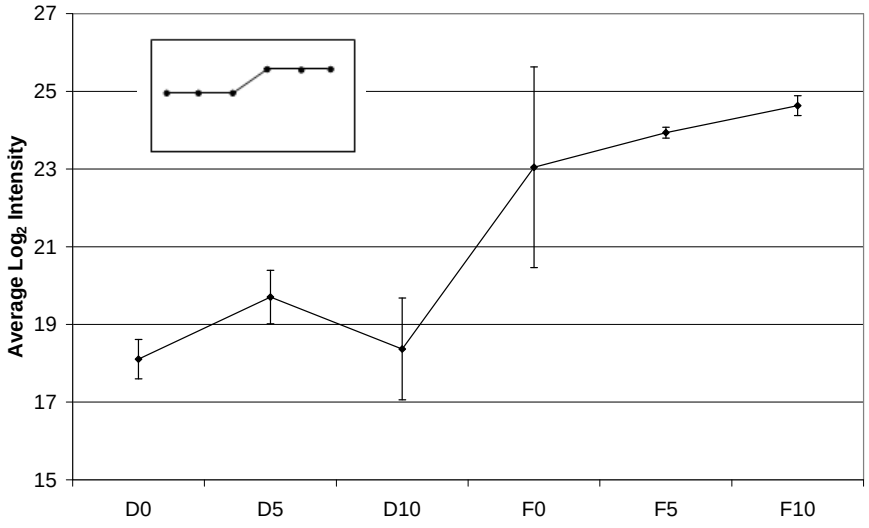
Expression profile from significant feature 30 from sub 10kDa intact samples



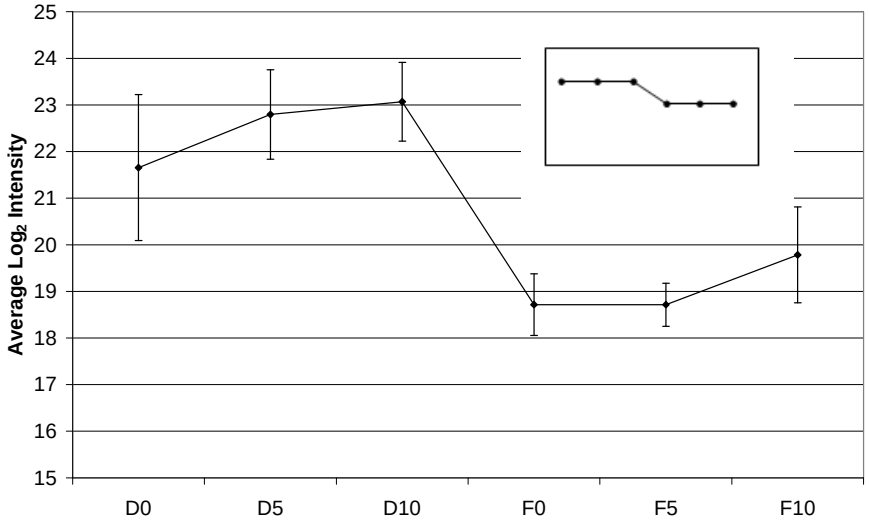
Expression profile from significant feature 31 from sub 10kDa intact samples



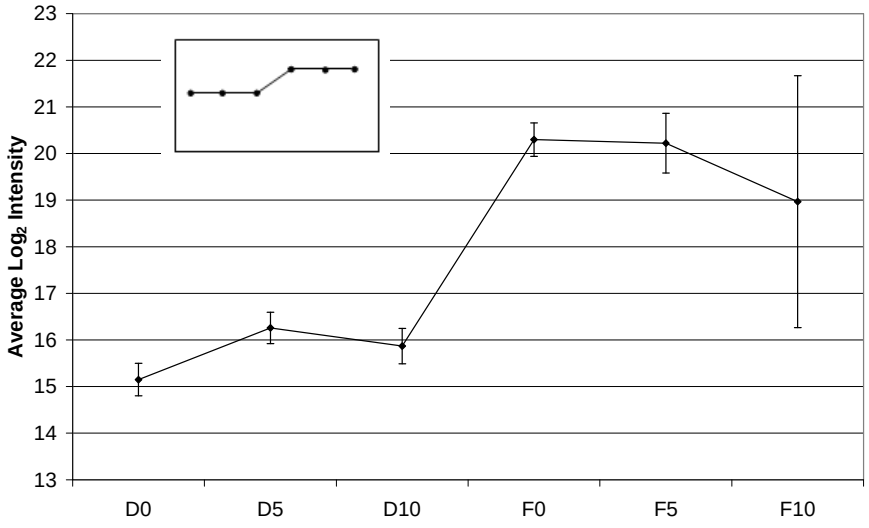
Expression profile from significant feature 32 from sub 10kDa intact samples



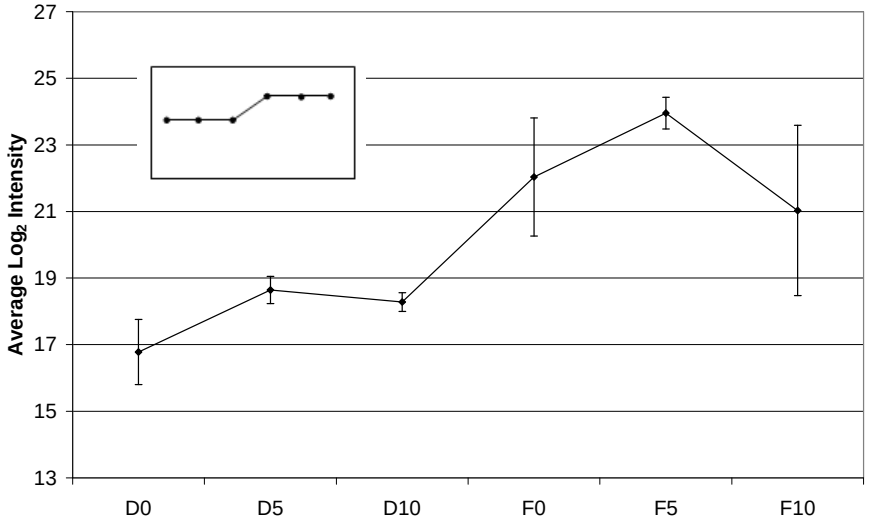
Expression profile from significant feature 33 from sub 10kDa intact samples



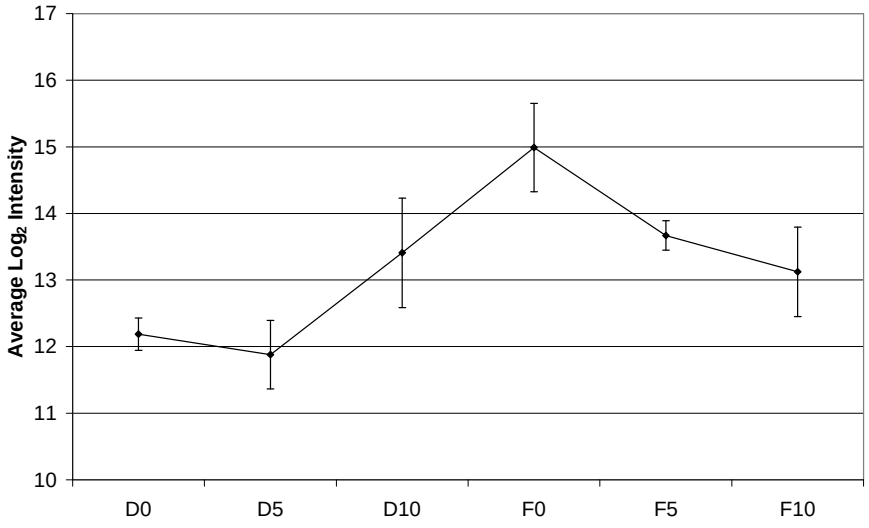
Expression profile from significant feature 34 from sub 10kDa intact samples



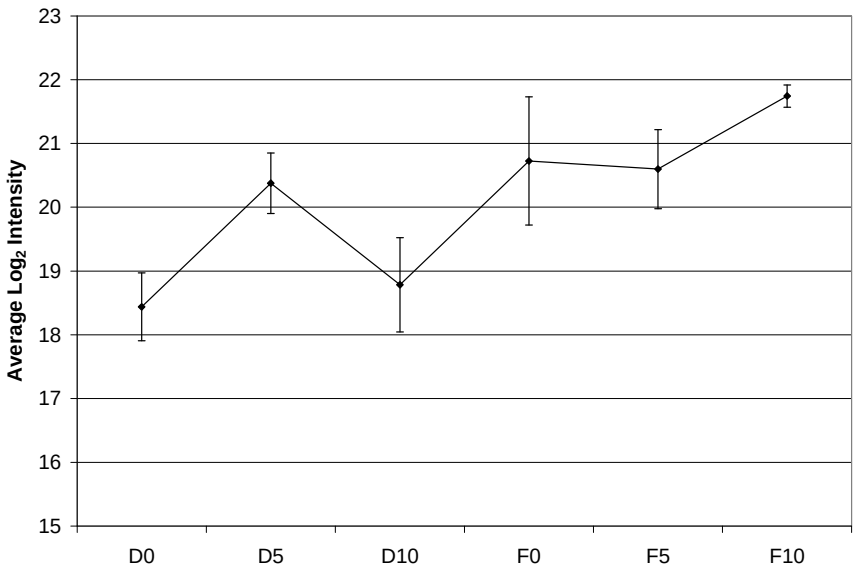
Expression profile from significant feature 35 from sub 10kDa intact samples



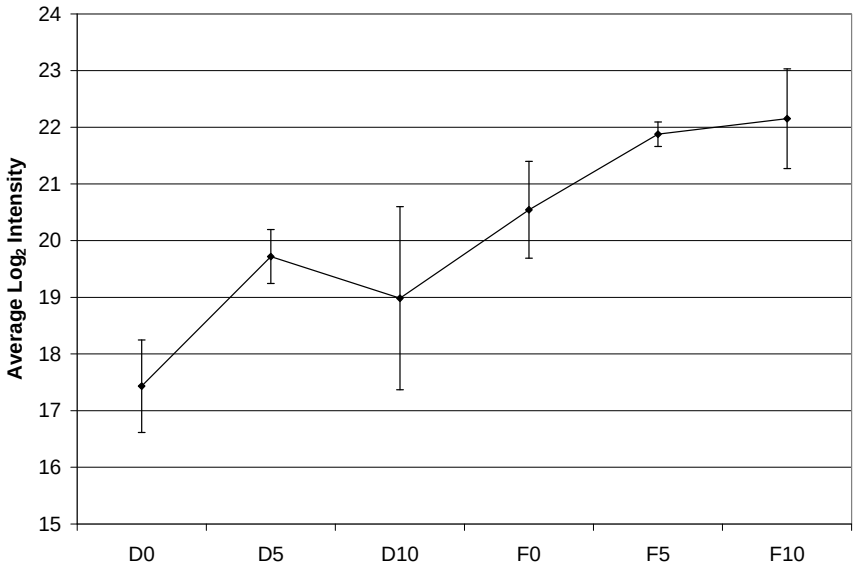
Expression profile from significant feature 36 from sub 10kDa intact samples



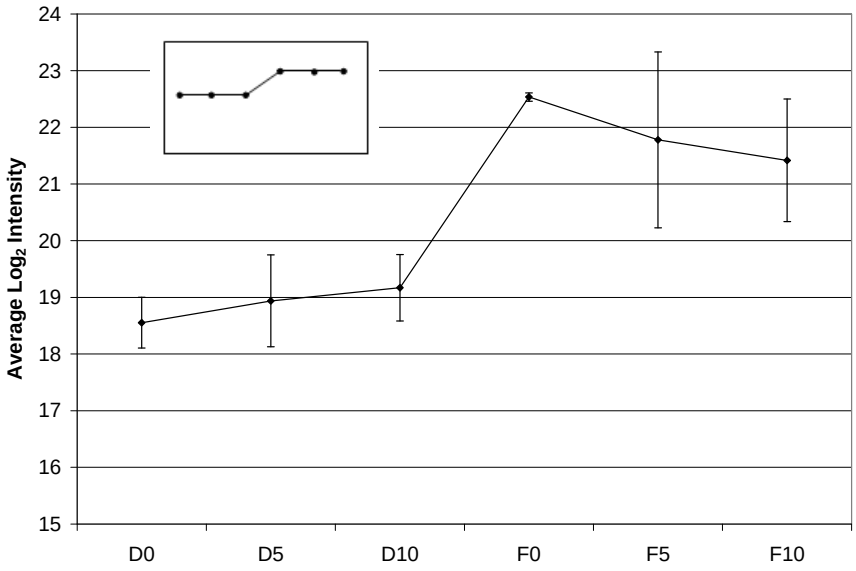
Expression profile from significant feature 37 from sub 10kDa intact samples



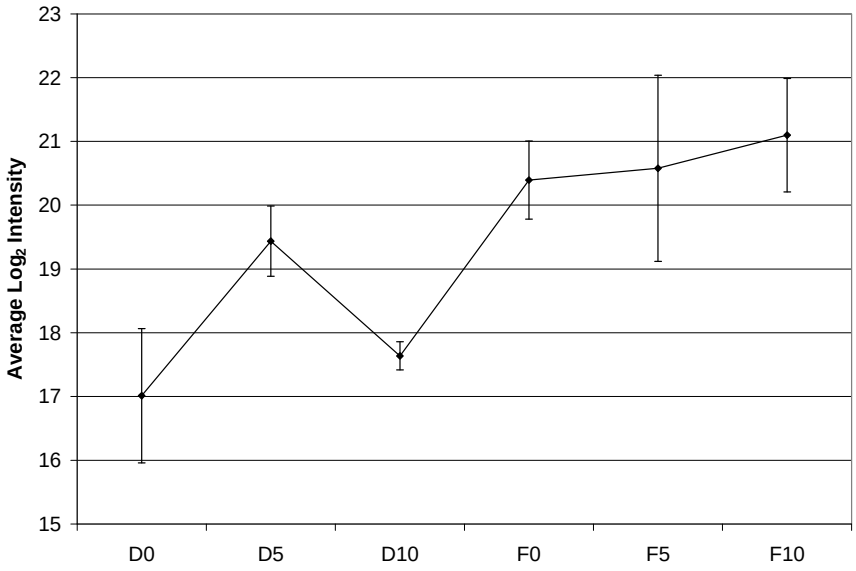
Expression profile from significant feature 38 from sub 10kDa intact samples



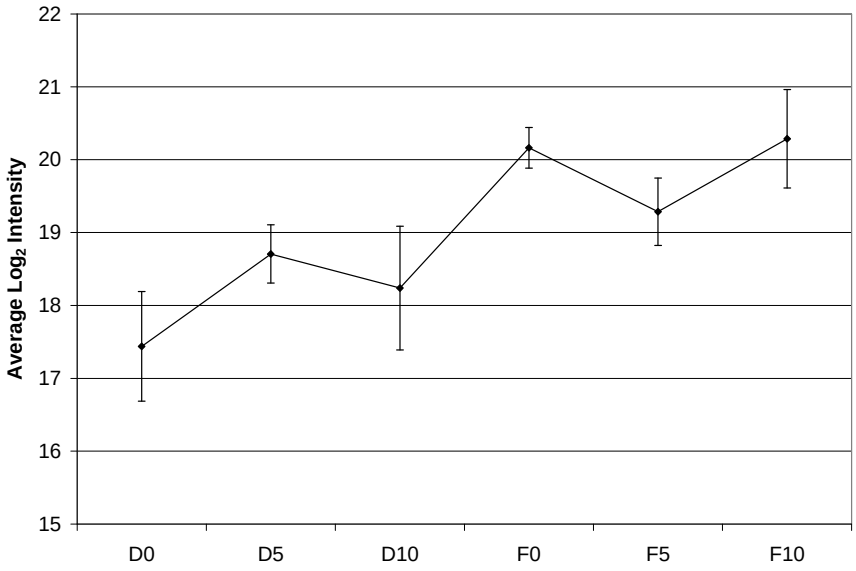
Expression profile from significant feature 39 from sub 10kDa intact samples



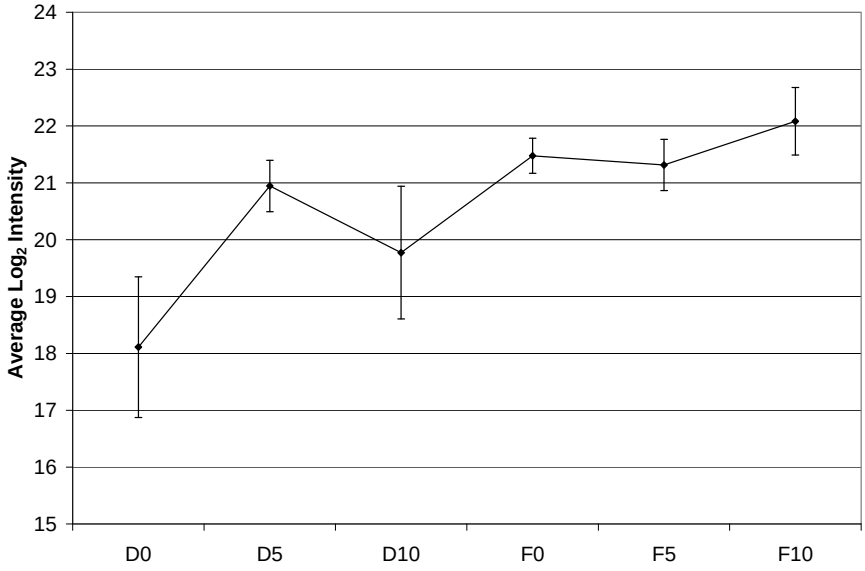
Expression profile from significant feature 40 from sub 10kDa intact samples



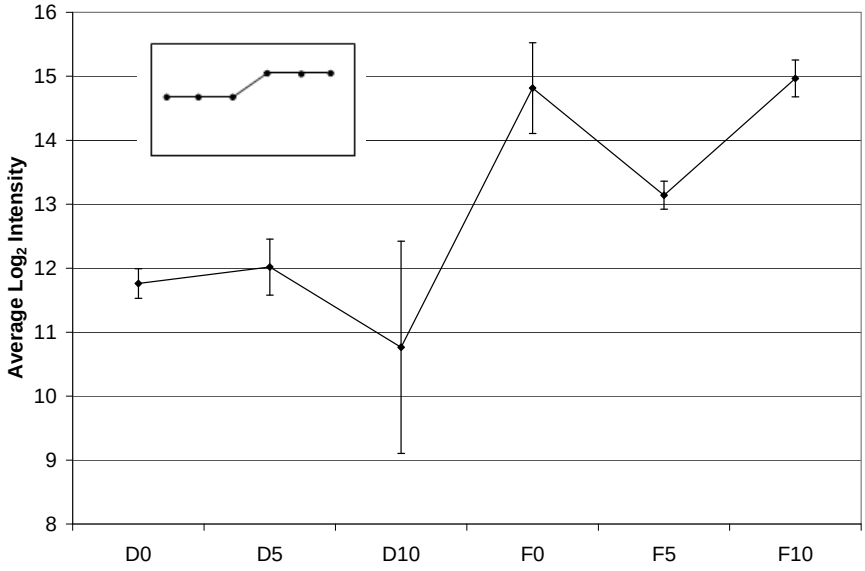
Expression profile from significant feature 41 from sub 10kDa intact samples



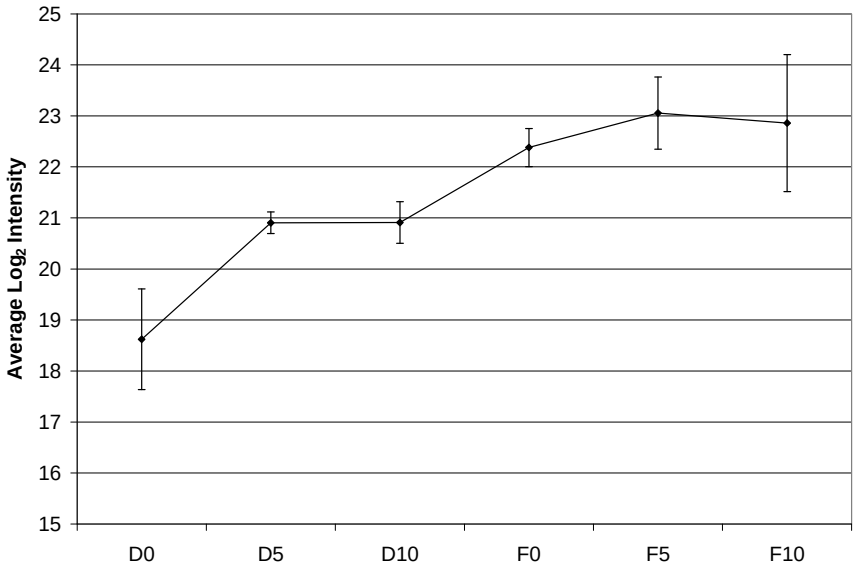
Expression profile from significant feature 42 from sub 10kDa intact samples



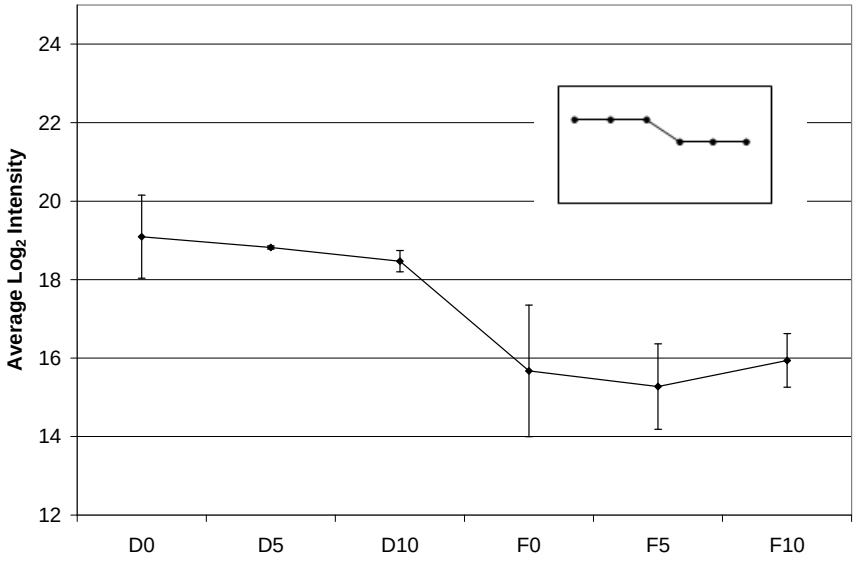
Expression profile from significant feature 43 from sub 10kDa intact samples



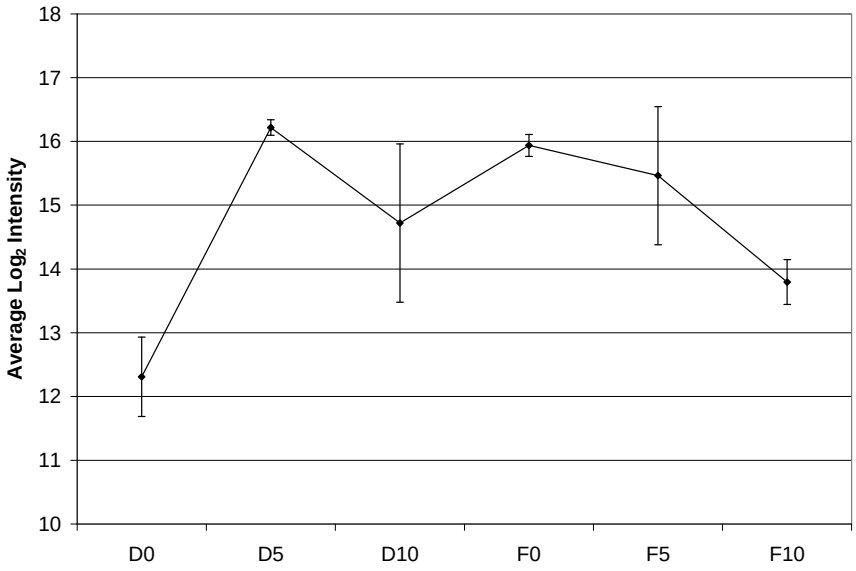
Expression profile from significant feature 44 from sub 10kDa intact samples



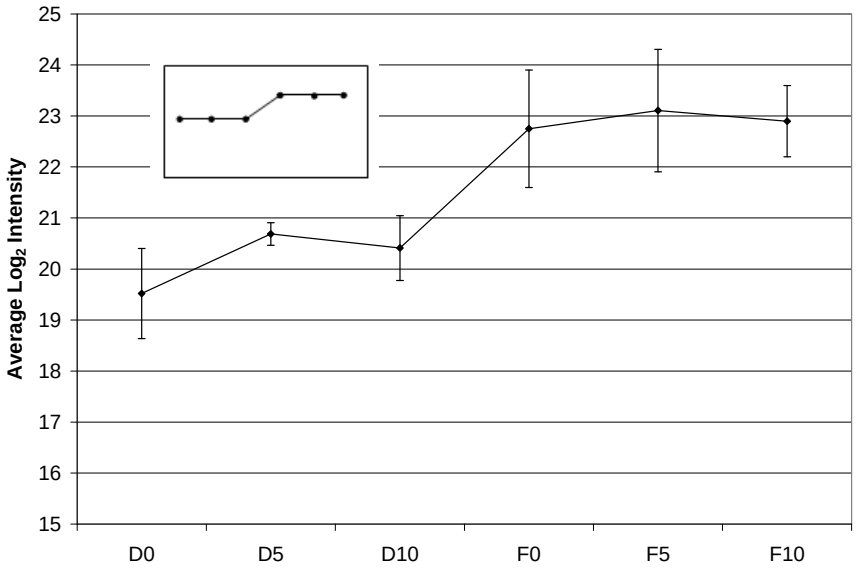
Expression profile from significant feature 45 from sub 10kDa intact samples



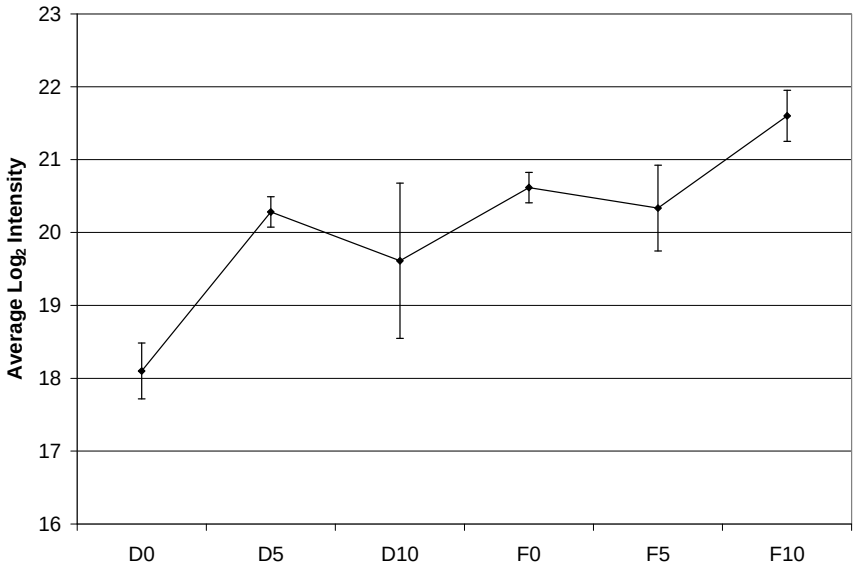
Expression profile from significant feature 46 from sub 10kDa intact samples



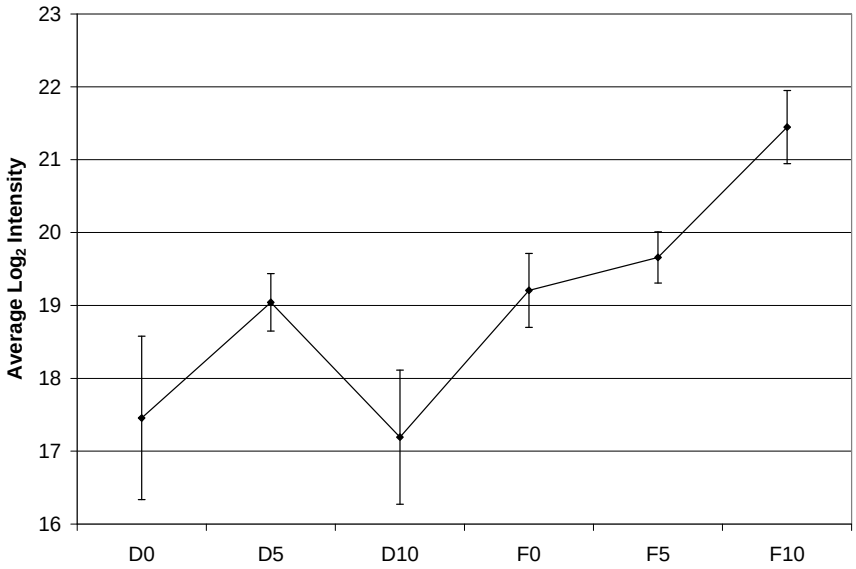
Expression profile from significant feature 47 from sub 10kDa intact samples



Expression profile from significant feature 48 from sub 10kDa intact samples



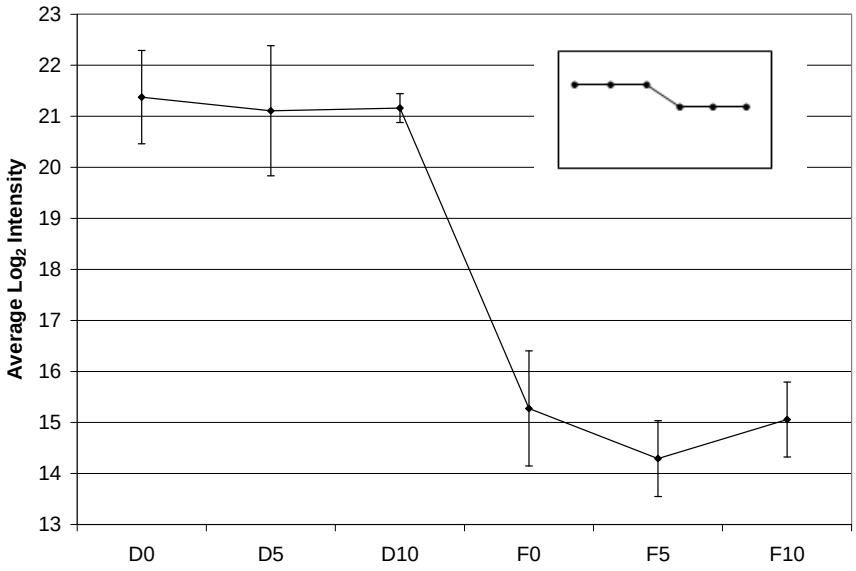
Expression profile from significant feature 49 from sub 10kDa intact samples



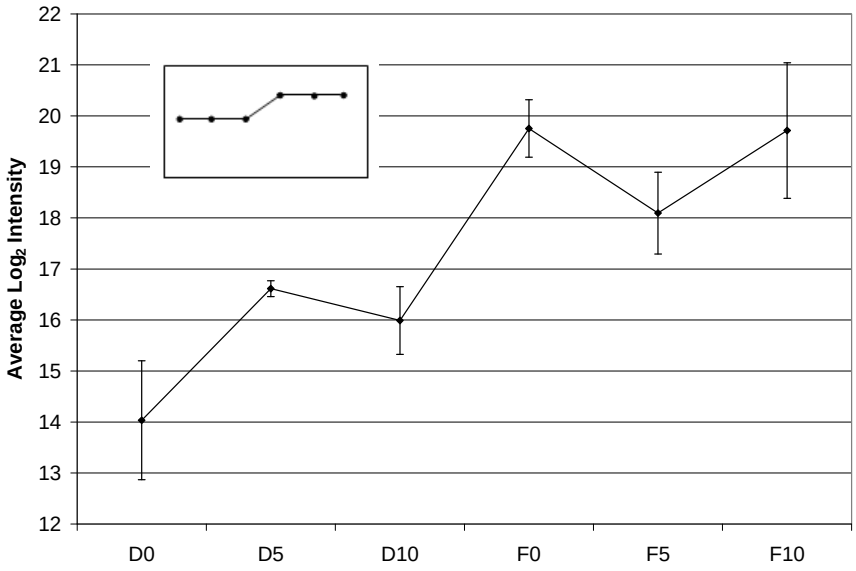
Expression profile from significant feature 50 from sub 10kDa intact samples

Appendix 4 – Assessment of Heat Treatment for Prevention of Proteomic Sample Degradation using Label Free Relative Quantitation

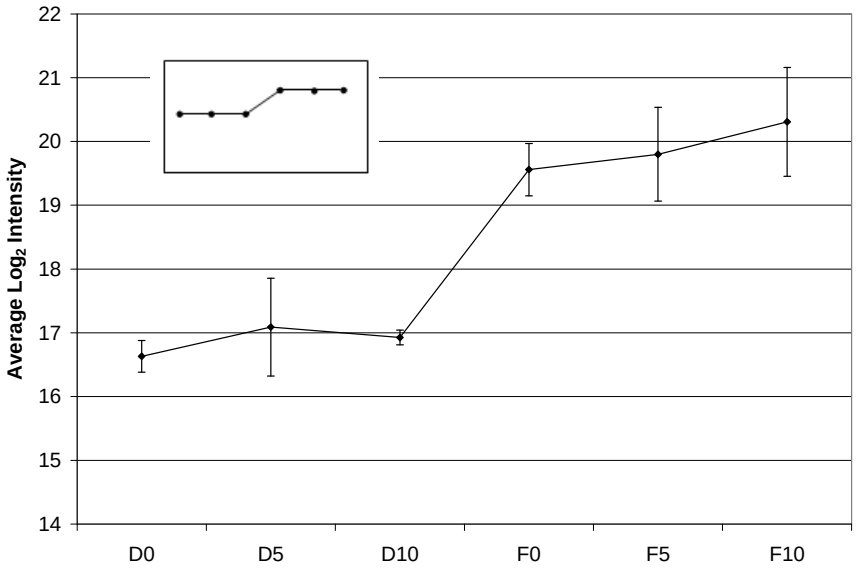
The figures shown below demonstrate the expression profiles exhibited by the 50 features with p-values ≤ 0.001 from the sub 10kDa digested samples, as discussed in section 3.4.2.2. From visual examination of the profiles it is demonstrated that 6 show expression profile as predicted by figure 3.2 C and 5 show expression profile as predicted by figure 3.2 D.



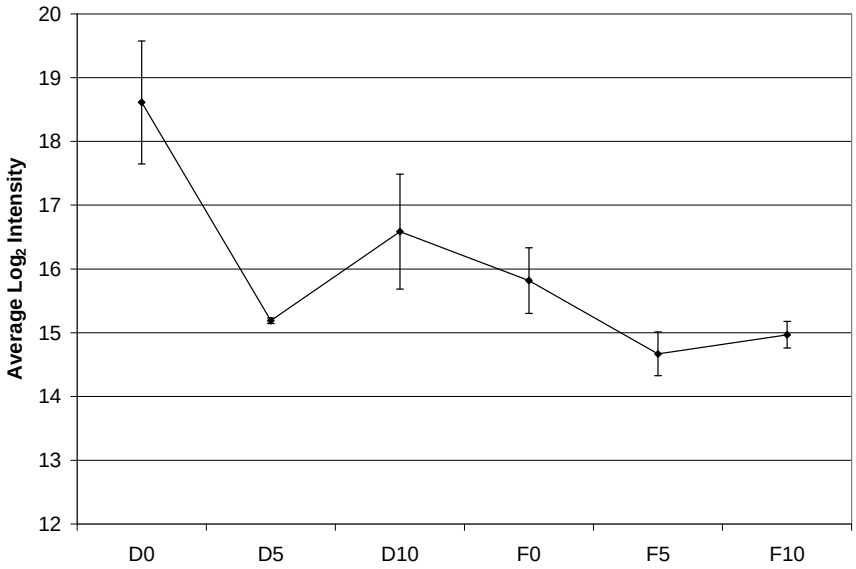
Expression profile from significant feature 1 from sub 10kDa digested samples



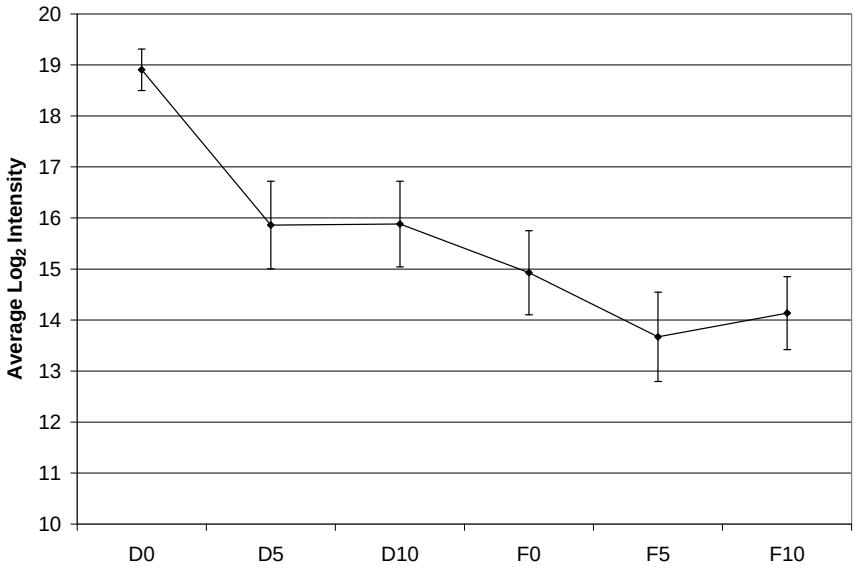
Expression profile from significant feature 2 from sub 10kDa digested samples



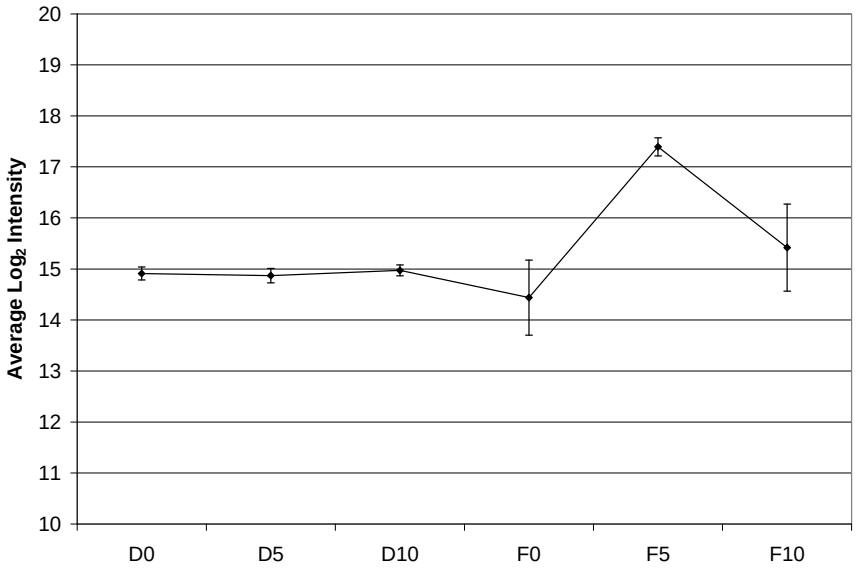
Expression profile from significant feature 3 from sub 10kDa digested samples



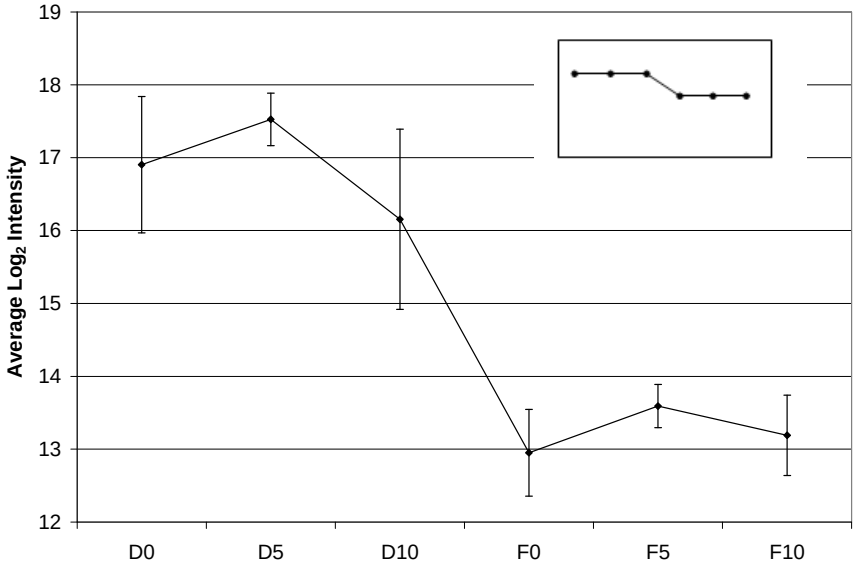
Expression profile from significant feature 4 from sub 10kDa digested samples



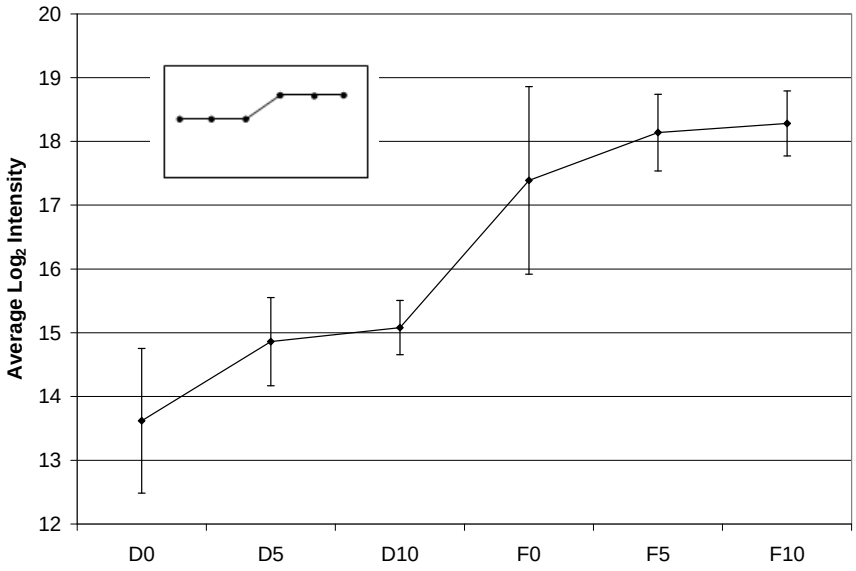
Expression profile from significant feature 5 from sub 10kDa digested samples



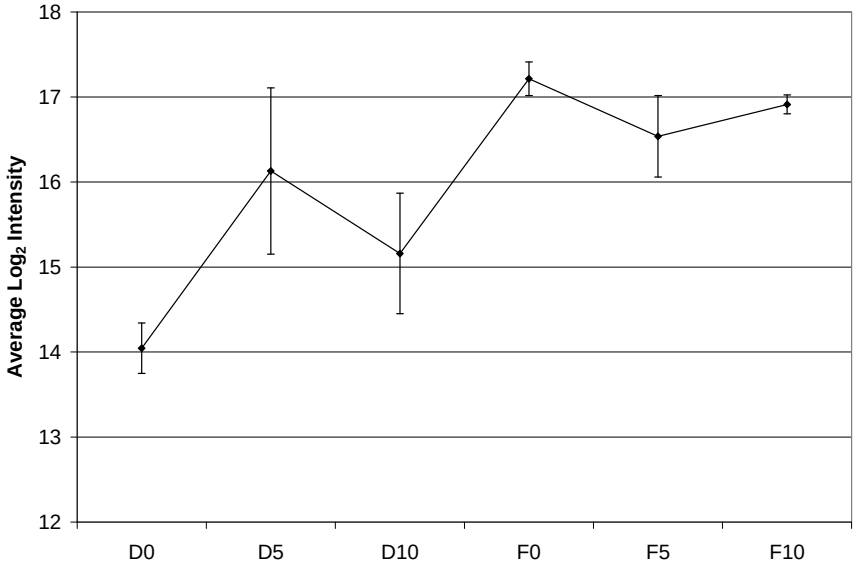
Expression profile from significant feature 6 from sub 10kDa digested samples



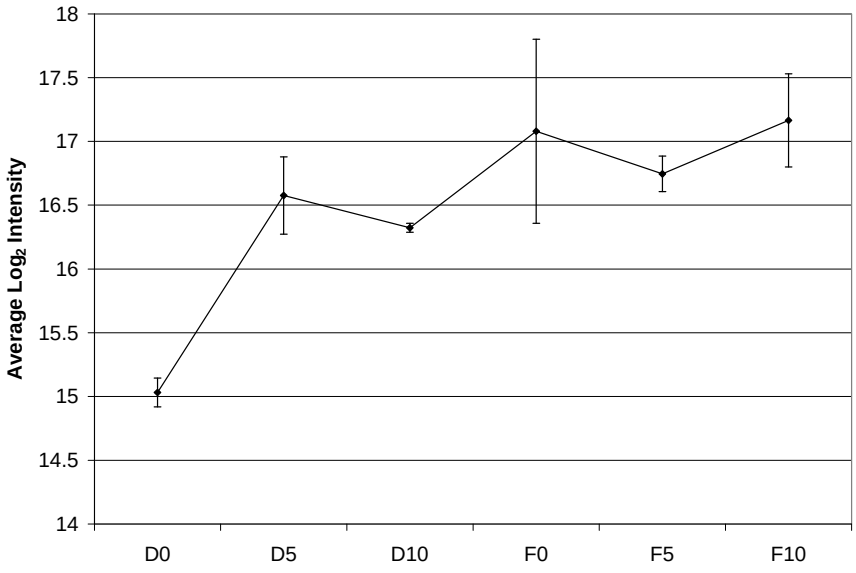
Expression profile from significant feature 7 from sub 10kDa digested samples



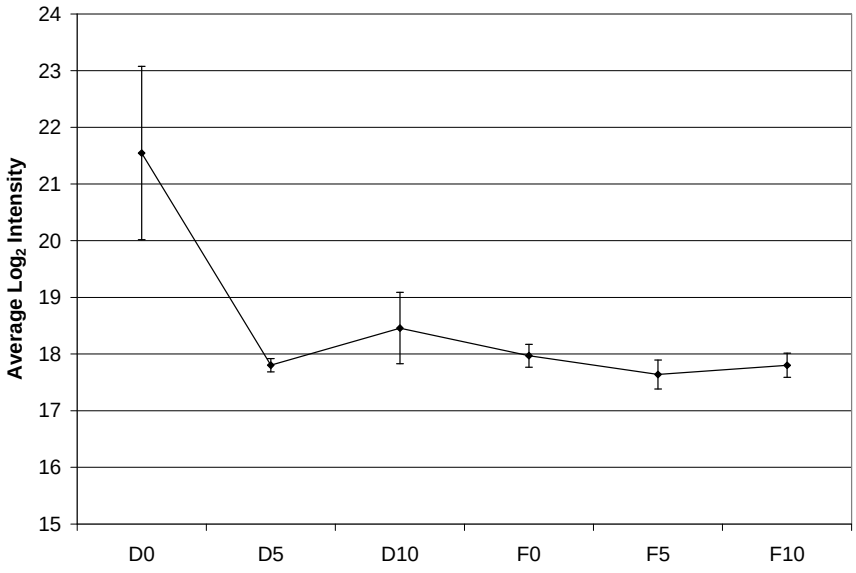
Expression profile from significant feature 8 from sub 10kDa digested samples



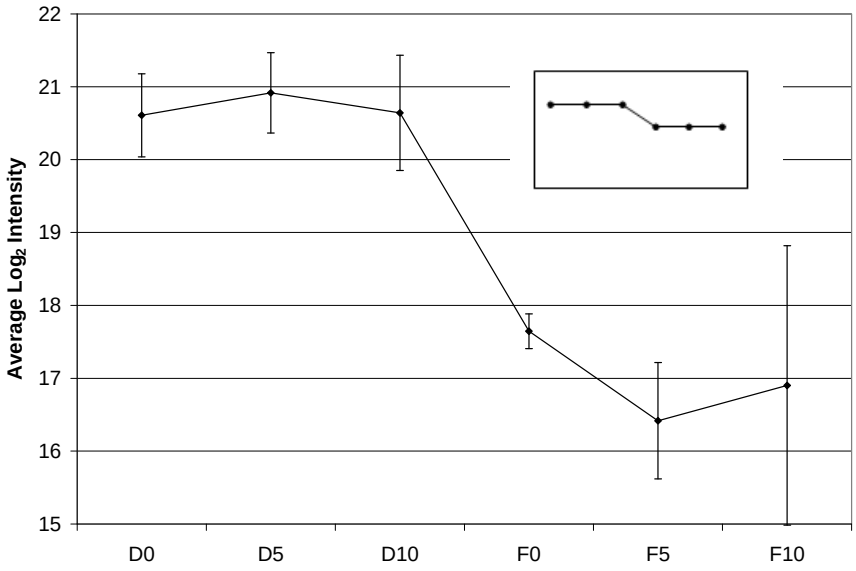
Expression profile from significant feature 9 from sub 10kDa digested samples



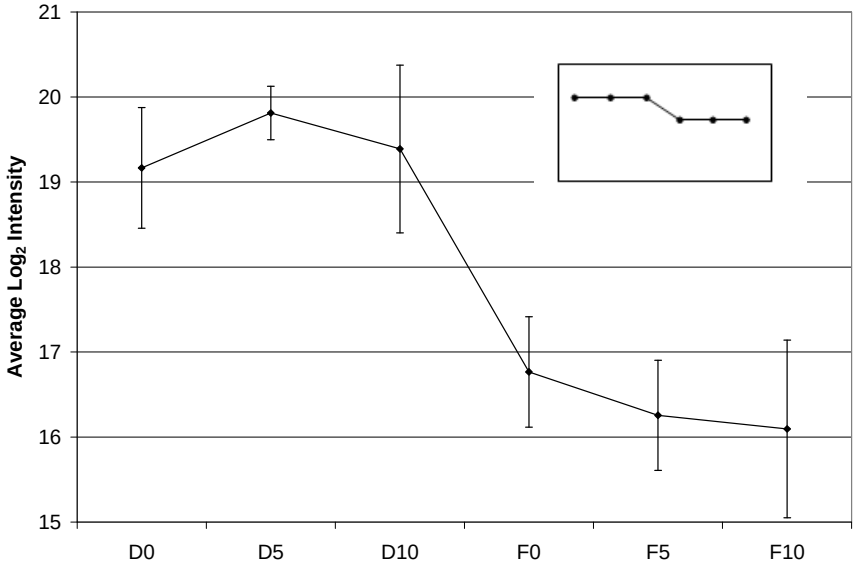
Expression profile from significant feature 10 from sub 10kDa digested samples



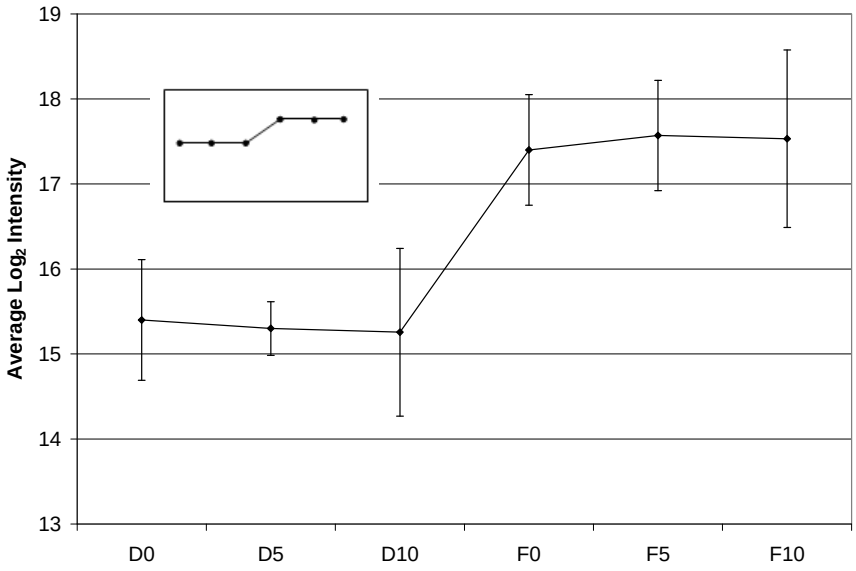
Expression profile from significant feature 11 from sub 10kDa digested samples



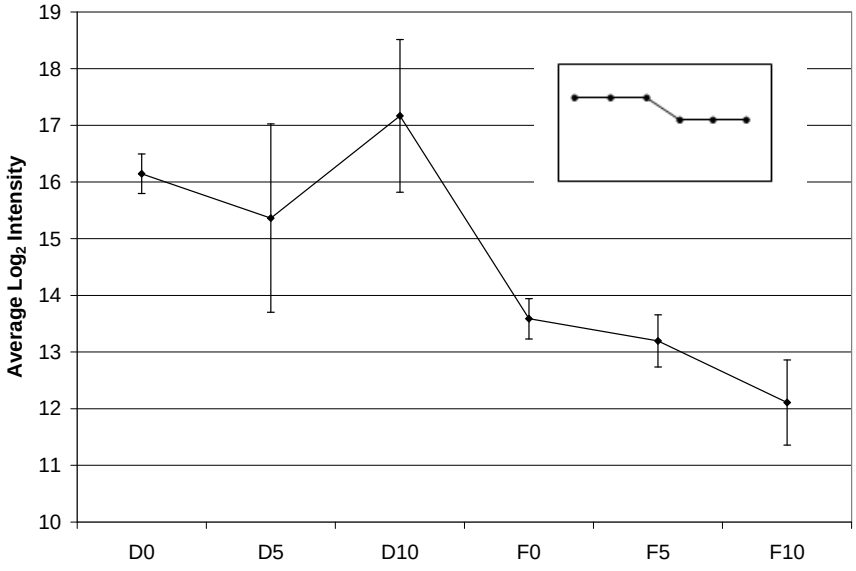
Expression profile from significant feature 12 from sub 10kDa digested samples



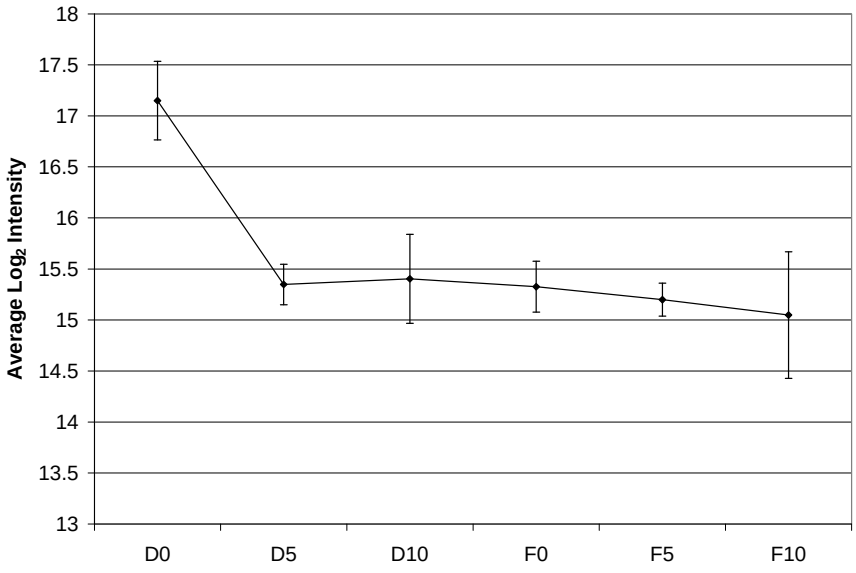
Expression profile from significant feature 13 from sub 10kDa digested samples



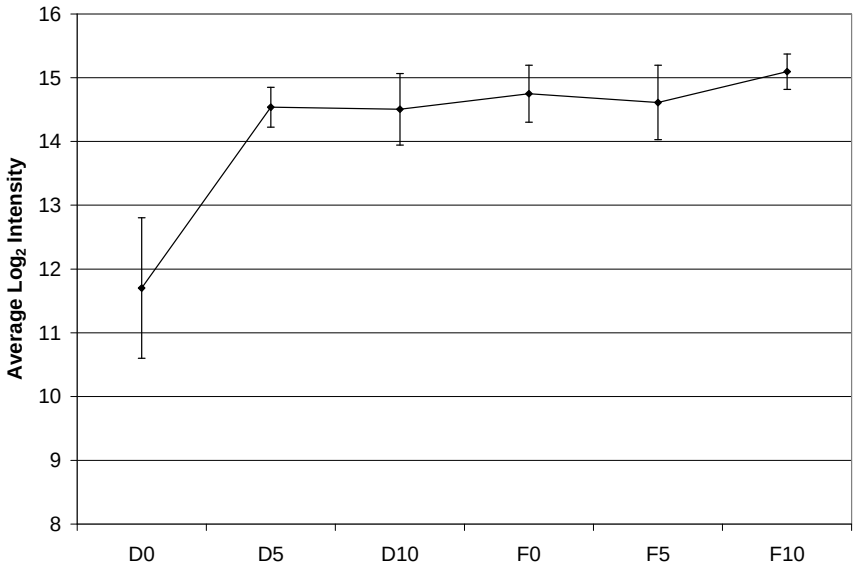
Expression profile from significant feature 14 from sub 10kDa digested samples



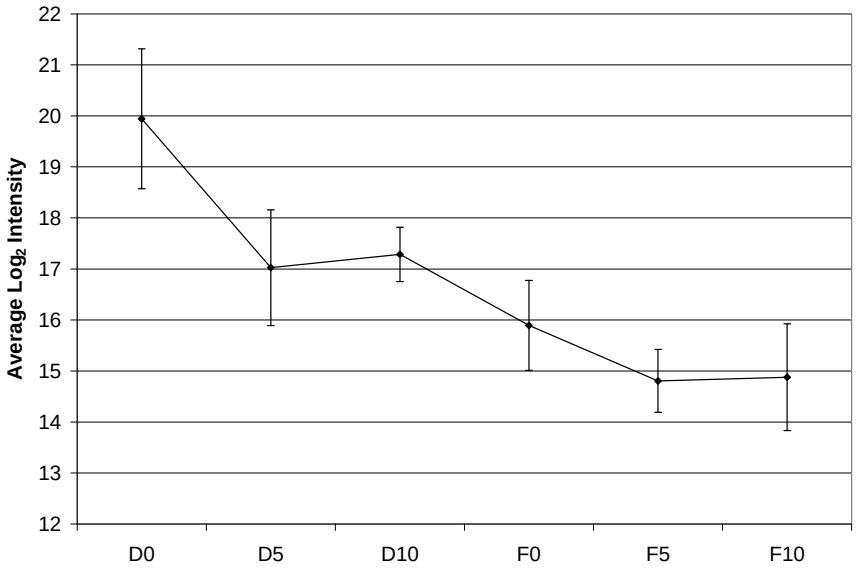
Expression profile from significant feature 15 from sub 10kDa digested samples



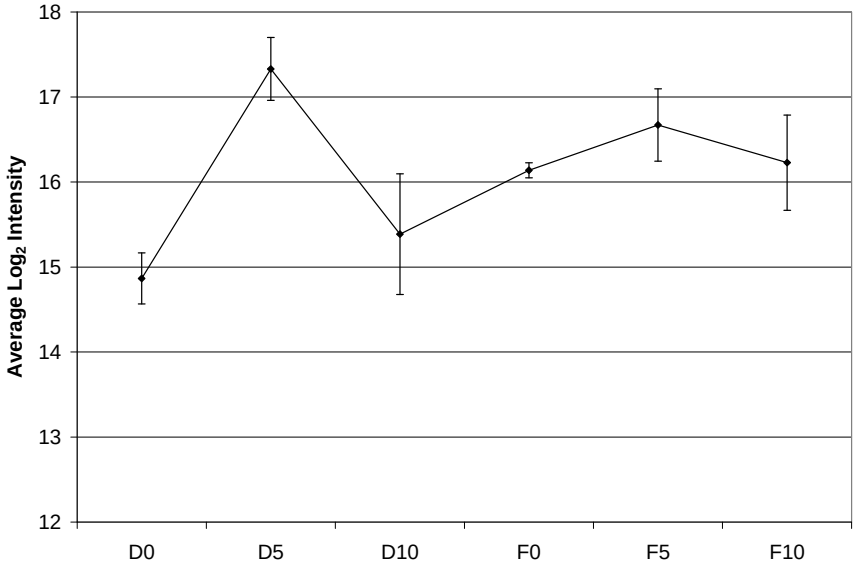
Expression profile from significant feature 16 from sub 10kDa digested samples



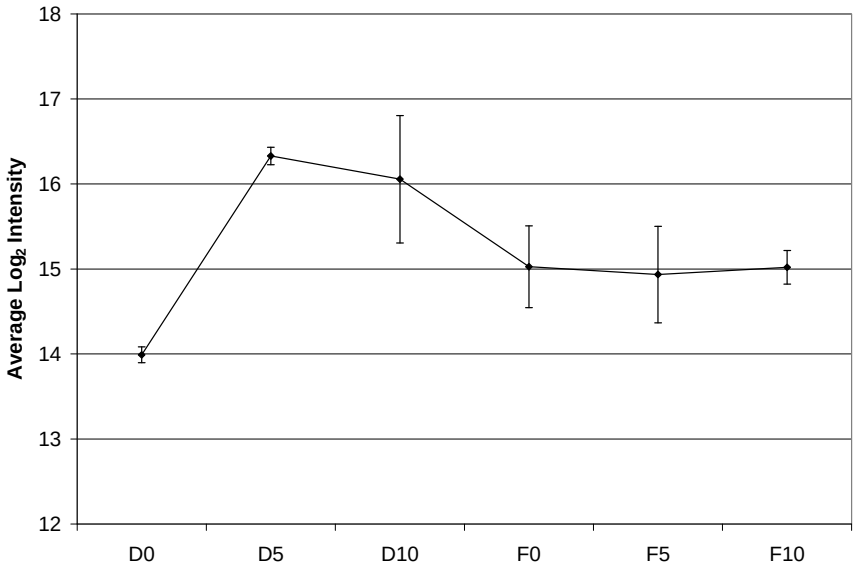
Expression profile from significant feature 17 from sub 10kDa digested samples



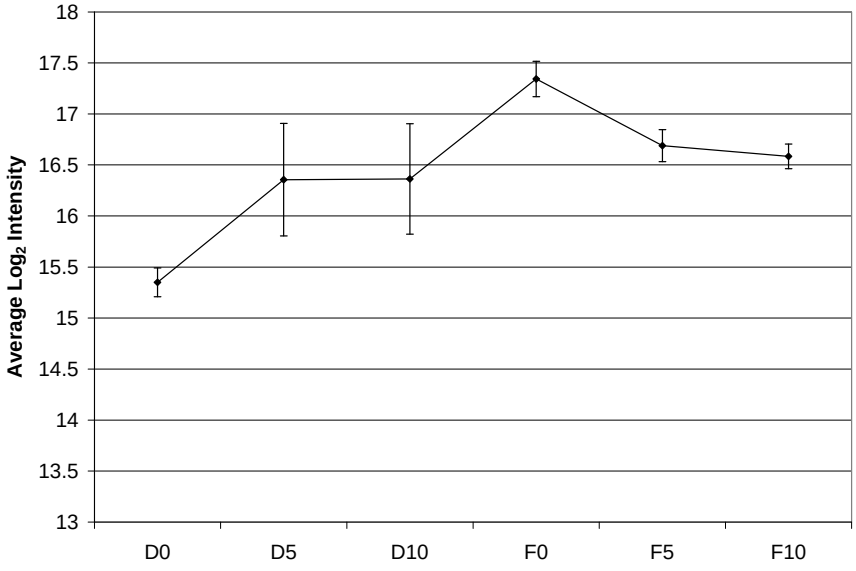
Expression profile from significant feature 18 from sub 10kDa digested samples



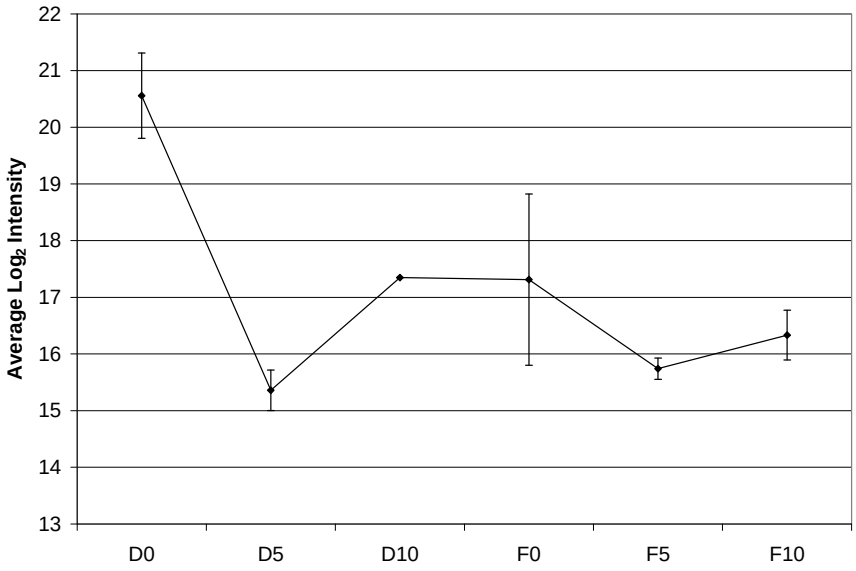
Expression profile from significant feature 19 from sub 10kDa digested samples



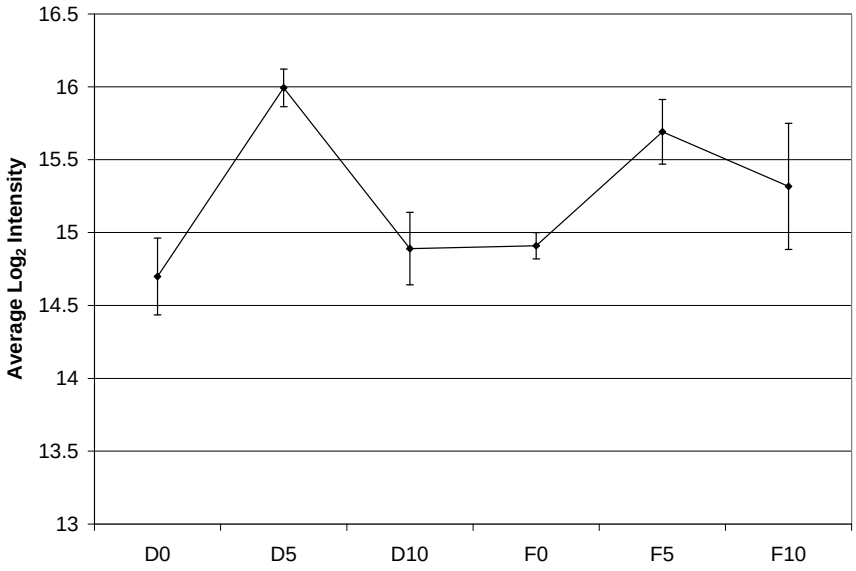
Expression profile from significant feature 20 from sub 10kDa digested samples



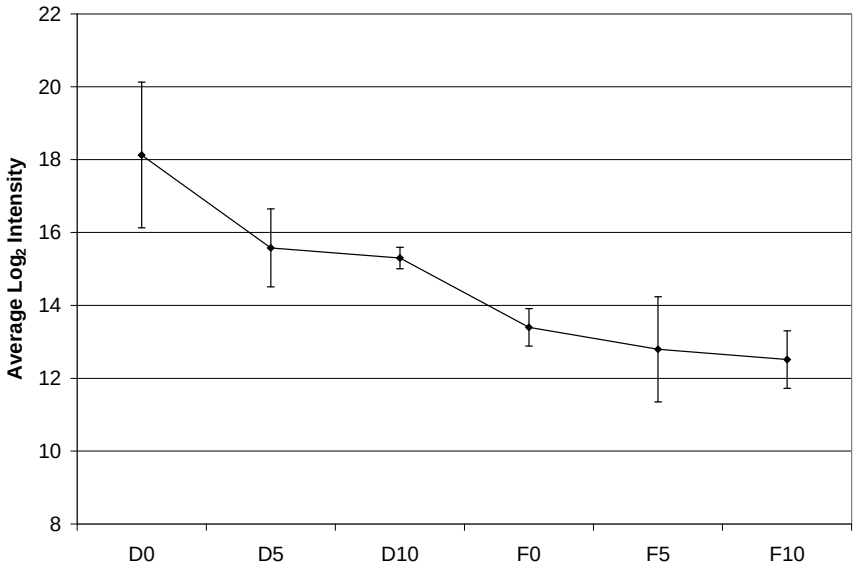
Expression profile from significant feature 21 from sub 10kDa digested samples



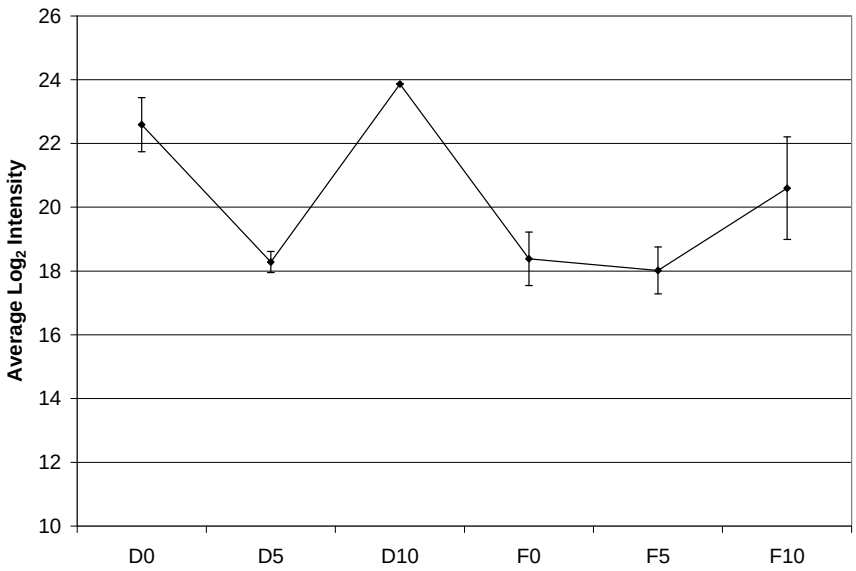
Expression profile from significant feature 22 from sub 10kDa digested samples



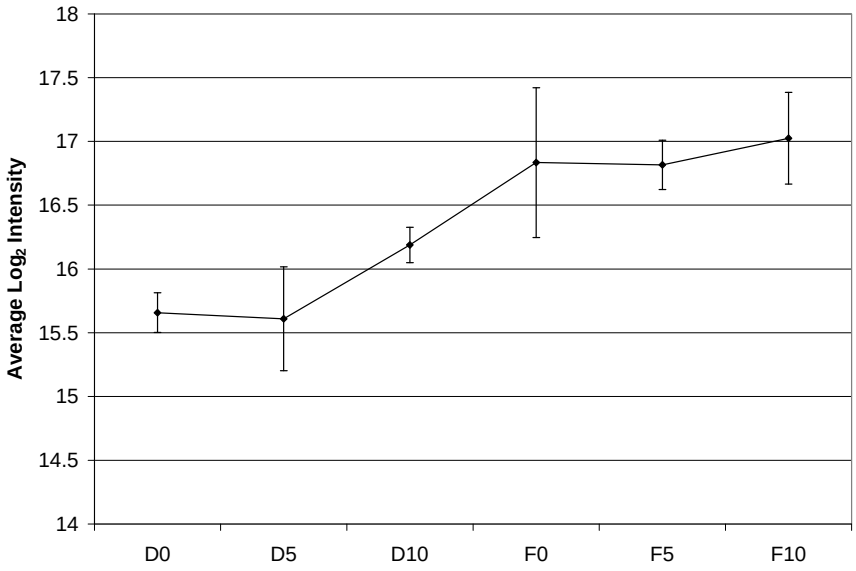
Expression profile from significant feature 23 from sub 10kDa digested samples



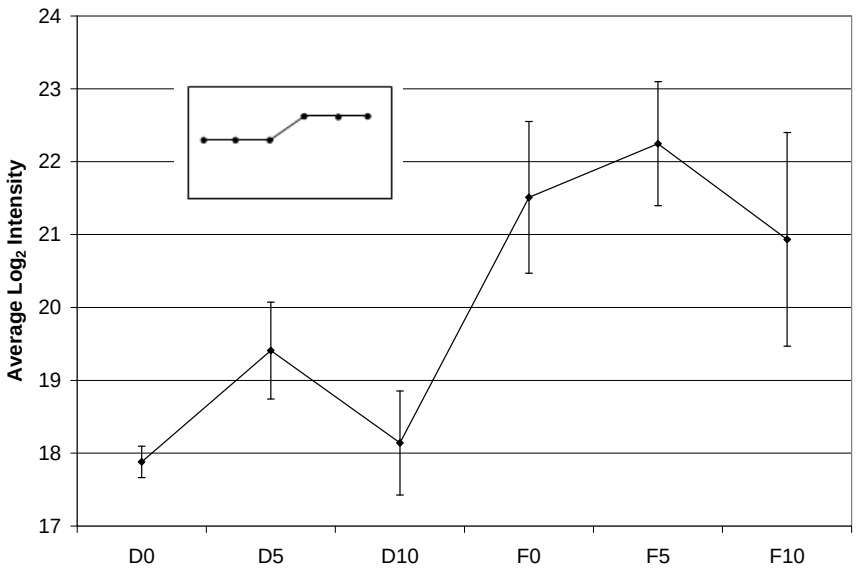
Expression profile from significant feature 24 from sub 10kDa digested samples



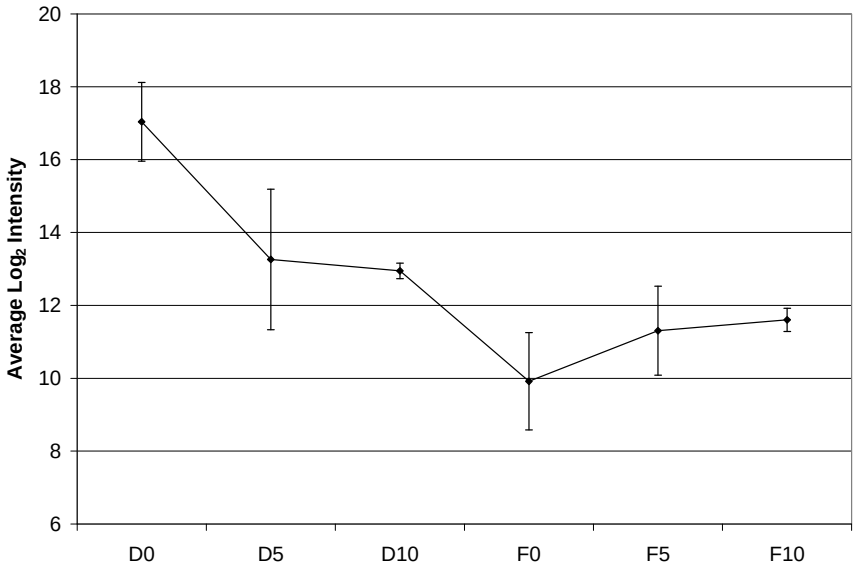
Expression profile from significant feature 25 from sub 10kDa digested samples



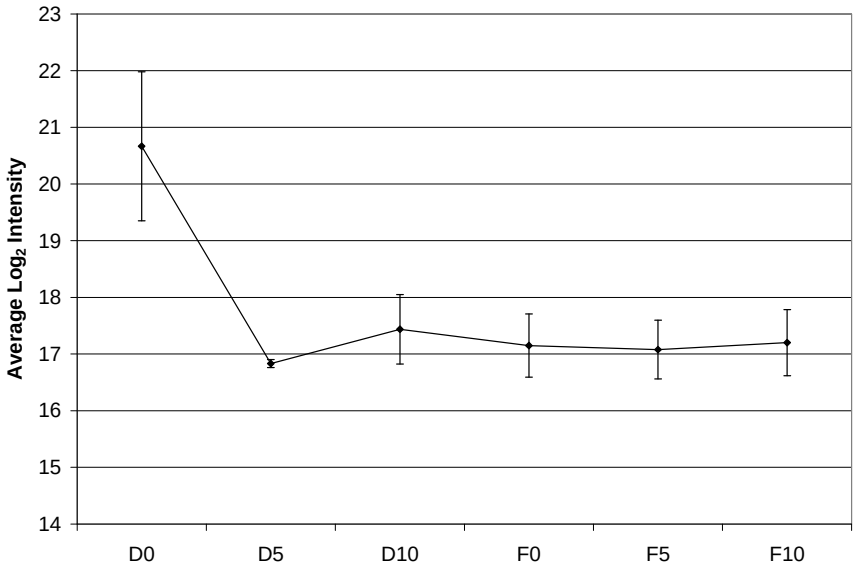
Expression profile from significant feature 26 from sub 10kDa digested samples



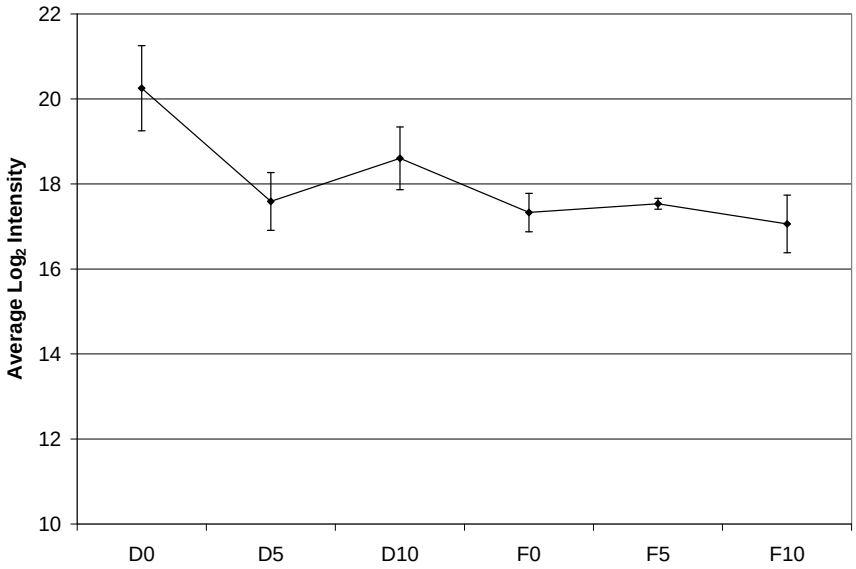
Expression profile from significant feature 27 from sub 10kDa digested samples



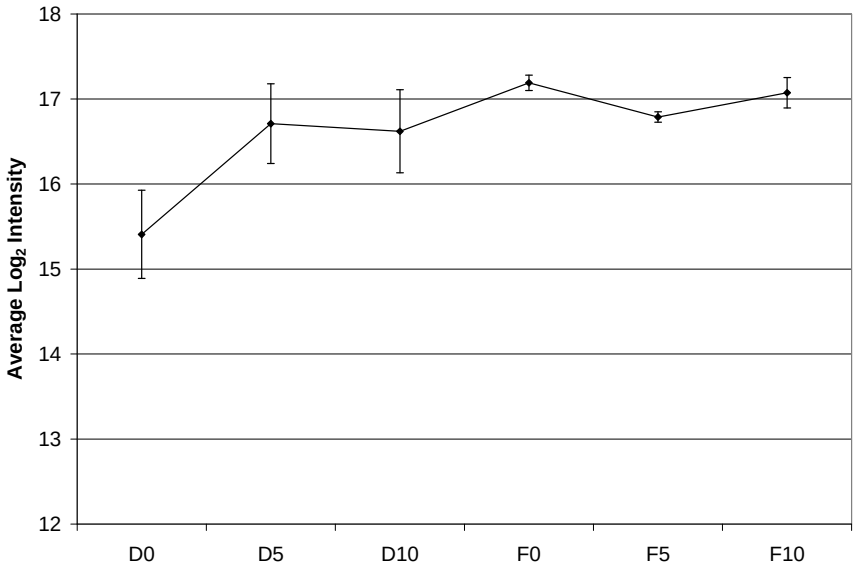
Expression profile from significant feature 28 from sub 10kDa digested samples



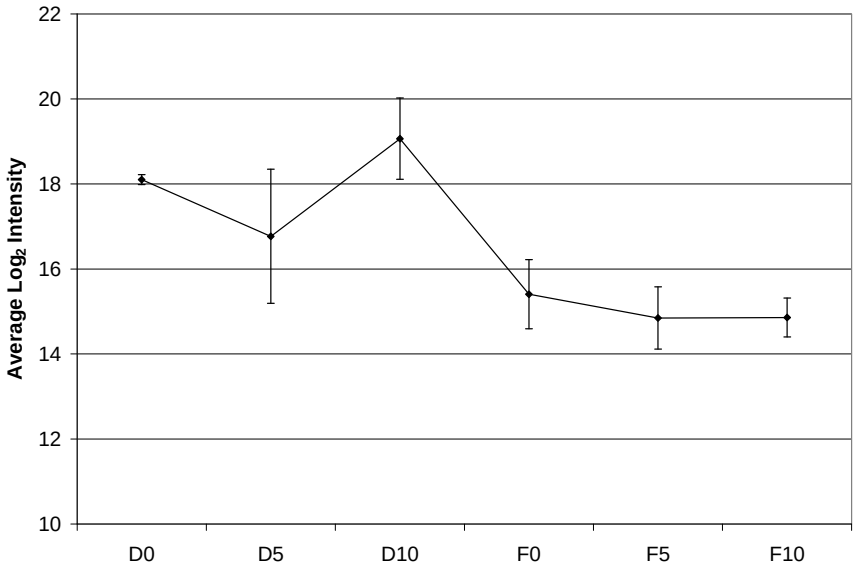
Expression profile from significant feature 29 from sub 10kDa digested samples



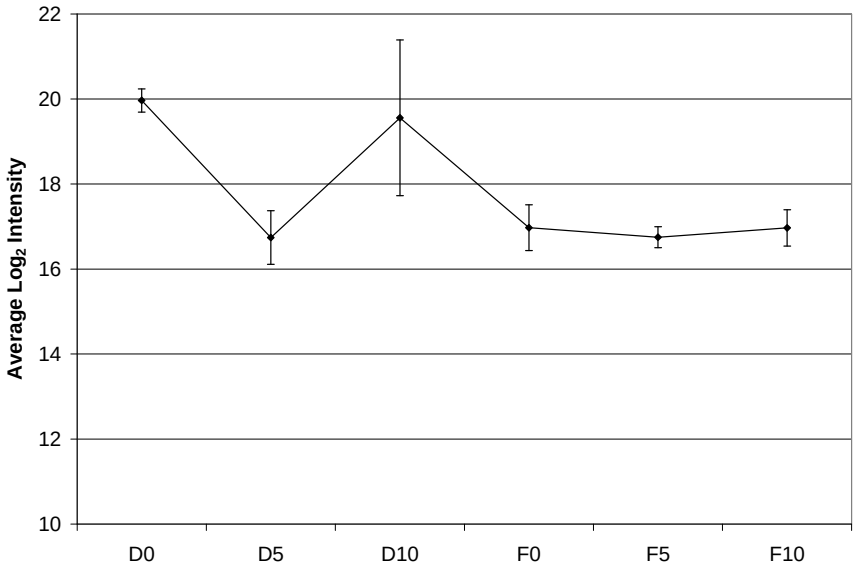
Expression profile from significant feature 30 from sub 10kDa digested samples



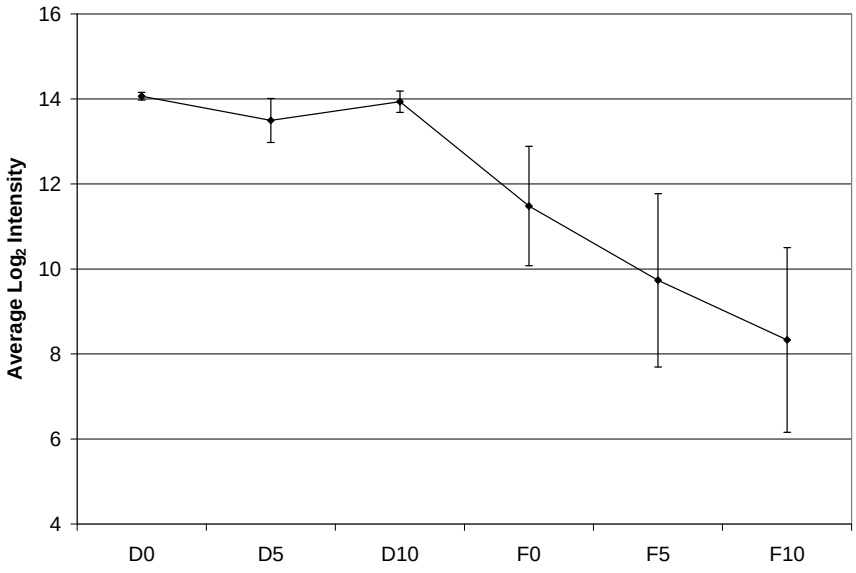
Expression profile from significant feature 31 from sub 10kDa digested samples



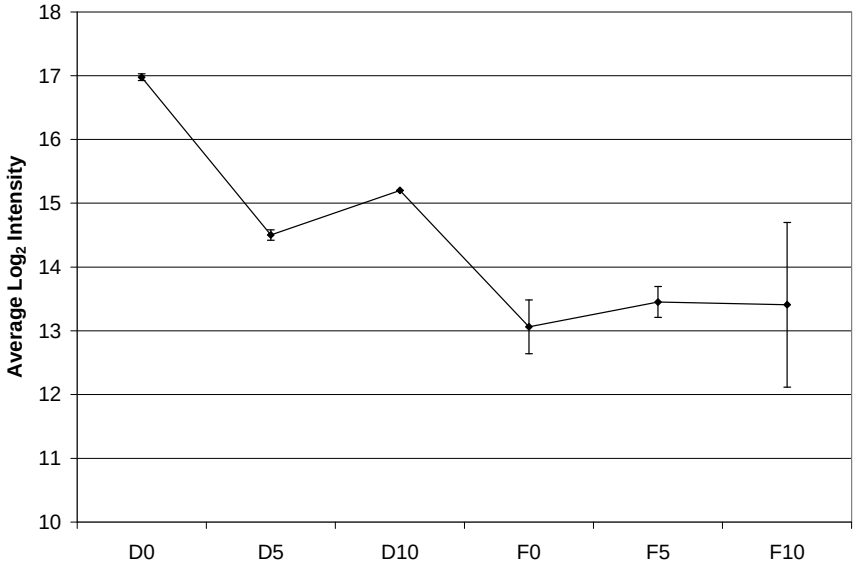
Expression profile from significant feature 32 from sub 10kDa digested samples



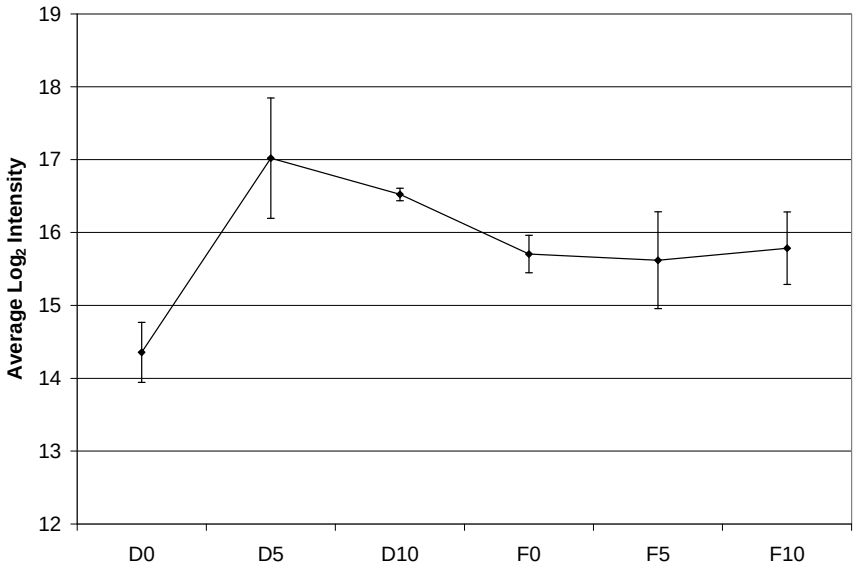
Expression profile from significant feature 33 from sub 10kDa digested samples



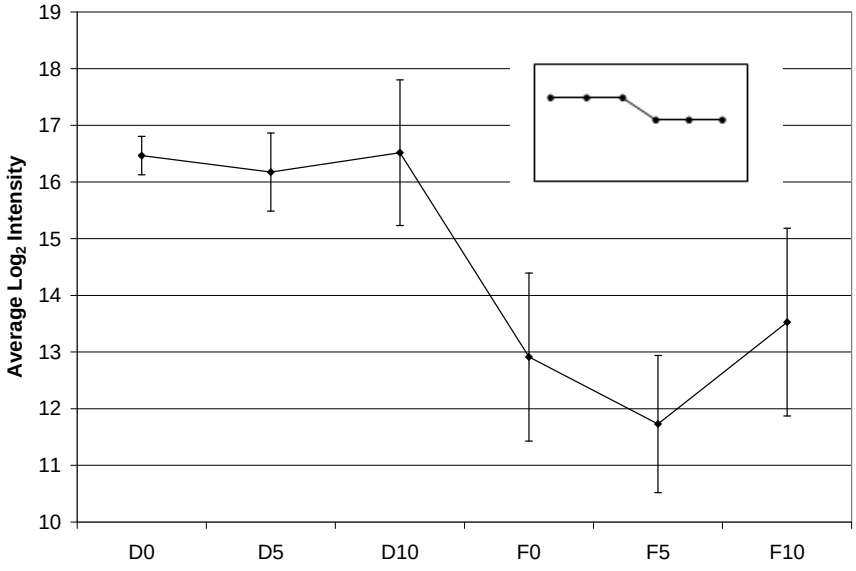
Expression profile from significant feature 34 from sub 10kDa digested samples



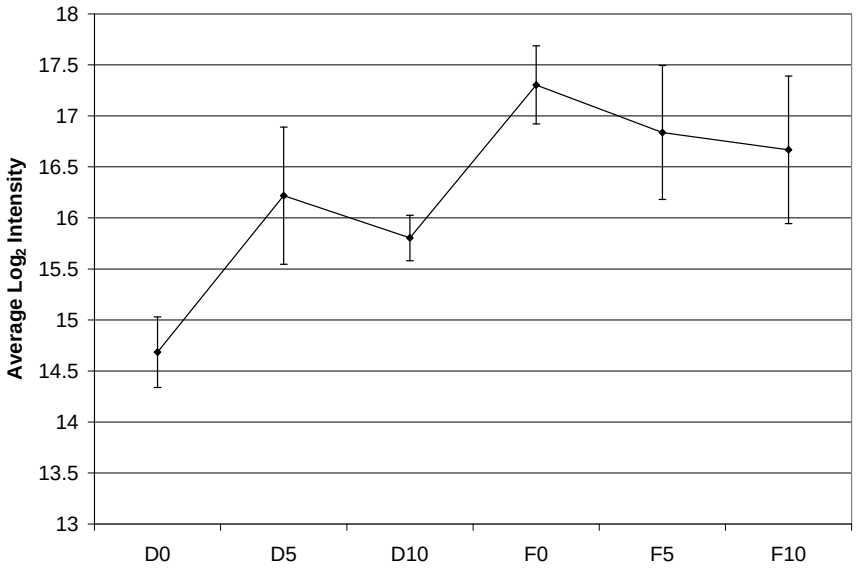
Expression profile from significant feature 35 from sub 10kDa digested samples



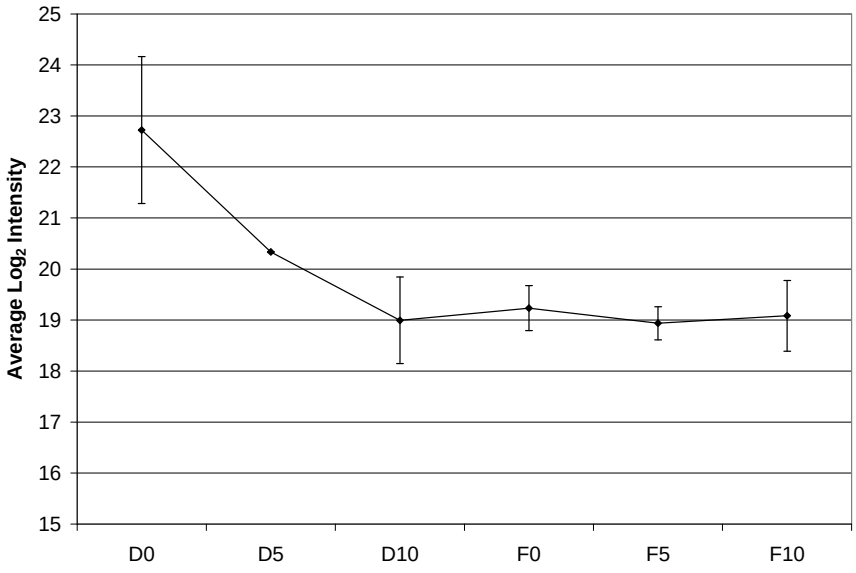
Expression profile from significant feature 36 from sub 10kDa digested samples



Expression profile from significant feature 37 from sub 10kDa digested samples



Expression profile from significant feature 38 from sub 10kDa digested samples



Expression profile from significant feature 39 from sub 10kDa digested samples

Reference List

(1) Persidis A. Proteomics - An ambitious drug development platform attempts to link gene sequence to expressed phenotype under various physiological states. Nature Biotechnology 1998; 16(4):393-394.

(2) Supratim Choudhuri. The Path from Nuclein to Human Genome: A Brief History of DNA with a Note on Human Genome. Bulletin of Science, Technology and Society 2003; 291(5507):1195.

(3) Collins FS, Lander ES, Rogers J, Waterston RH, Int Human Genome Sequencing Conso. Finishing the euchromatic sequence of the human genome. Nature 2004; 431(7011):931-945.

(4) James P. Protein identification in the post-genome era: the rapid rise of proteomics. Quarterly Reviews of Biophysics 1997; 30(4):279-331.

(5) Wasinger VC, Cordwell SJ, Cerpapoljak A, Yan JX, Gooley AA, Wilkins MR et al. Progress with Gene-Product Mapping of the Mollicutes - Mycoplasma-Genitalium. Electrophoresis 1995; 16(7):1090-1094.

(6) Jones MB, Krutzsch H, Shu HJ, Zhao YM, Liotta LA, Kohn EC et al. Proteomic analysis and identification of new biomarkers and therapeutic targets for invasive ovarian cancer. Proteomics 2002; 2(1):76-84.

(7) Wu W, Hu W, Kavanagh JJ. Proteomics in cancer research. International Journal of Gynecological Cancer 2002; 12(5):409-423.

(8) Anderson NL, Anderson NG. The human plasma proteome - History, character, and diagnostic prospects. Molecular & Cellular Proteomics 2002; 1(11):845-867.

(9) Veenstra TD, Conrads TP, Hood BL, Avellino AM, Ellenbogen RG, Morrison RS. Biomarkers: Mining the biofluid proteome. Molecular & Cellular Proteomics 2005; 4(4):409-418.

(10) Ma YF, Liu GS, Du M, Stayton I. Recent developments in the determination of urinary cancer biomarkers by capillary electrophoresis. Electrophoresis 2004; 25(10-11):1473-1484.

(11) Theodorescu D, Mischak H. Mass spectrometry based proteomics in urine biomarker discovery. World Journal of Urology 2007; 25:435-443.

(12) Han WK, Waikar SS, Johnson A, Betensky RA, Dent CL, Devarajan P et al. Urinary biomarkers in the early diagnosis of acute kidney injury (vol 73, pg 863, 2008). Kidney International 2009; 76(3):348-349.

(13) Vestergaard P, Leverett R. Constancy of Urinary Creatinine Excretion. Journal of Laboratory and Clinical Medicine 1958; 51(2):211-218.

(14) Schiffer E, Mischak H, Novak J. High resolution proteome/peptidome analysis of body fluids by capillary electrophoresis coupled with MS. Proteomics 2006; 6(20):5615-5627.

(15) Kussmann M, Raymond F, Affolter M. OMICS-driven biomarker discovery in nutrition and health. Journal of Biotechnology 2006; 124(4):758-787.

(16) Schaub S, Wilkins J, Weiler T, Sangster K, Rush D, Nickerson P. Urine protein profiling with surface-enhanced laser-desorption/ionization time-of-flight mass spectrometry. Kidney International 2004; 65(1):323-332.

(17) Mischak H, Apweiler R, Banks RE, Conaway M, Coon J, Dominiczak A et al. Clinical proteomics: A need to define the field and to begin to set adequate standards. Proteomics Clinical Applications 2007; 1(2):148-156.

(18) Dayarathna MKDR, Hancock WS, Hincapie M. A two step fractionation approach for plasma proteomics using immunodepletion of abundant proteins and multi-lectin affinity chromatography: Application to the analysis of obesity, diabetes, and hypertension diseases. Journal of Separation Science 2008; 31(6-7):1156-1166.

(19) Abdul-Salam VB, Paul GA, Ali JO, Gibbs SR, Rahman D, Taylor GW et al. Identification of plasma protein biomarkers associated with idiopathic pulmonary arterial hypertension. Proteomics 2006; 6(7):2286-2294.

(20) Berhane BT, Zong CG, Liem DA, Huang A, Le S, Edmondson RD et al. Cardiovascular-related proteins identified in human plasma by the HUPO Plasma Proteome Project pilot Phase. Proteomics 2005; 5(13):3520-3530.

(21) Anderson L. Candidate-based proteomics in the search for biomarkers of cardiovascular disease. Journal of Physiology-London 2005; 563(1):23-60.

(22) Donahue MP, Rose K, Vonderscher J, Bougueleret L, Nelson CL, Cohen D et al. Proteomic discovery of coronary artery disease proteins using industrial scale analysis of pooled plasma: Results from the pooled plasma proteomics project. Circulation 2004; 110(17):1011.

(23) Florian-Kujawski M, Hussain W, Chyna B, Kahn S, Hoppensteadt D, Leya E et al. Biomarker profiling of plasma from acute coronary syndrome patients - Application of ProteinChip Array analysis. International Angiology 2004; 23(3):246-254.

(24) Hoffman SA, Joo WA, Echan LA, Speicher DW. Higher dimensional (Hi-D) separation strategies dramatically improve the potential for cancer biomarker detection in serum and plasma. Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences 2007; 849(1-2):43-52.

(25) Pietrowska M. Tumor markers studied with proteomic methods in blood and serum plasma. Biotechnologia (Poznan) 2009;(2):39-53.

(26) Schaub NP, Jones KJ, Nyalwidhe JO, Cazares LH, Karbassi ID, Semmes OJ et al. Serum Proteomic Biomarker Discovery Reflective of Stage and Obesity in Breast Cancer Patients. Journal of the American College of Surgeons 2009; 208(5):970-978.

(27) Vasudev NS, Ferguson RE, Cairns DA, Stanley AJ, Selby PJ, Banks RE. Serum biomarker discovery in renal cancer using 2-DE and prefractionation by immunodepletion and isoelectric focusing; increasing coverage or more of the same? Proteomics 2008; 8(23-24):5074-5085.

(28) Grote T, Siwak DR, Fritsche HA, Joy C, Mills GB, Simeone D et al. Validation of reverse phase protein array for practical screening of potential biomarkers in serum and plasma: Accurate detection of CA19-9 levels in pancreatic cancer. Proteomics 2008; 8(15):3051-3060.

(29) Omenn GS, States DJ, Adamski M, Blackwell TW, Menon R, Hermjakob H et al. Overview of the HUPO Plasma Proteome Project: Results from the pilot phase with 35 collaborating laboratories and multiple analytical groups, generating a core dataset of 3020 proteins and a publicly-available database. Proteomics 2005; 5(13):3226-3245.

(30) Shi M, Caudle WM, Zhang J. Biomarker discovery in neurodegenerative diseases: A proteomic approach. Neurobiology of Disease 2009; 35(2):157-164.

(31) de Roos B. Proteomic analysis of human plasma and blood cells in nutritional studies: development of biomarkers to aid disease prevention. Expert Review of Proteomics 2008; 5(6):819-826.

(32) Ahmed FE. Laser microdissection: Application to carcinogenesis. Cancer Genomics & Proteomics 2006; 3(3-4):217-225.

(33) Chaurand P, Sanders ME, Jensen RA, Caprioli RM. Proteomics in diagnostic pathology - Profiling and imaging proteins directly in tissue sections. American Journal of Pathology 2004; 165(4):1057-1068.

(34) Vincek V, Nassiri M, Nadji M, Morales AR. A tissue fixative that protects macromolecules (DNA, RNA, and protein) and histomorphology in clinical samples. Laboratory Investigation 2003; 83(10):1427-1435.

(35) Ahram M, Flaig MJ, Gillespie JW, Duray PH, Linehan WM, Ornstein DK et al. Evaluation of ethanol-fixed, paraffin-embedded tissues for proteomic applications. Proteomics 2003; 3(4):413-421.

(36) Griffin TJ, Lock CM, Li XJ, Patel A, Chervetsova L, Lee H et al. Abundance ratio-dependent proteomic analysis by mass spectrometry. Analytical Chemistry 2003; 75(4):867-874.

(37) Wright ME, Han DK, Aebersold R. Mass spectrometry-based expression profiling of clinical prostate cancer. Molecular & Cellular Proteomics 2005; 4(4):545-554.

(38) Etzioni R, Urban N, Ramsey S, McIntosh M, Schwartz S, Reid B et al. The case for early detection. Nature Reviews Cancer 2003; 3(4):243-252.

(39) Ludwig JA, Weinstein JN. Biomarkers in cancer staging, prognosis and treatment selection. Nature Reviews Cancer 2005; 5(11):845-856.

(40) Kiga C, Sakurai H, Goto H, Hayashi K, Shimada Y, Saiki I. Proteomic identification of haptoglobin as a stroke plasma biomarker in spontaneously hypertensive stroke-prone rats. Life Sciences 2008; 83(17-18):625-631.

(41) Silbiger V, Luchessi A, Hirata R, Neto L, Pastorelli C, Ueda E et al. Time course proteomic profiling of human myocardial infarction plasma samples: An approach to new biomarker discovery. Clinica Chimica Acta 412[11-12], 1086-1093. 12-5-2011.

(42) Hsieh SY, Chen RK, Pan YH, Lee HL. Systematical evaluation of the effects of sample collection procedures on low-molecular-weight serum/plasma proteome profiling. Proteomics 2006; 6(10):3189-3198.

(43) Rai AJ, Gelfand CA, Haywood BC, Warunek DJ, Yi JZ, Schuchard MD et al. HUPO Plasma Proteome Project specimen collection and handling: Towards the standardization of parameters for plasma proteome samples. Proteomics 2005; 5(13):3262-3277.

(44) Sampson M, Ruddel M, Albright S, Elin RJ. Positive interference in lithium determinations from clot activator in collection container. Clinical Chemistry 1997; 43(4):675-679.

(45) Drake SK, Bowen RAR, Remaley AT, Hortin GL. Potential interferences from blood collection tubes in mass spectrometric analyses of serum polypeptides. Clinical Chemistry 2004; 50(12):2398-2401.

(46) Villanueva J, Philip J, Chaparro CA, Li YB, Toledo-Crow R, DeNoyer L et al. Correcting common errors in identifying cancer-specific serum peptide signatures. Journal of Proteome Research 2005; 4(4):1060-1072.

(47) Baumann S, Ceglarek U, Fiedler GM, Lembcke J, Leichtle A, Thiery J. Standardized approach to proteome profiling of human serum based on magnetic bead separation and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Clinical Chemistry 2005; 51(6):973-980.

(48) Capila I, Linhardt RJ. Heparin - Protein interactions. Angewandte Chemie-International Edition 2002; 41(3):391-412.

(49) White JG. EDTA-induced changes in platelet structure and function: clot retraction. Platelets 2000; 11(1):49-55.

(50) Marshall J, Kupchak P, Zhu WM, Yantha J, Vrees T, Furesz S et al. Processing of serum proteins underlies the mass spectral fingerprinting of myocardial infarction. Journal of Proteome Research 2003; 2(4):361-372.

(51) West-Hielsen M, Hogdall EV, Marchiori E, Hogdall CK, Schou C, Heegaard NHH. Sample handling for mass spectrometric proteomic investigations of human sera. Analytical Chemistry 2005; 77(16):5114-5123.

(52) Kolch W, Mischak H, Pitt AR. The molecular make-up of a tumour: proteomics in cancer research. Clinical Science 2005; 108(5):369-383.

(53) Koomen J, Hawke D, Kobayashi R. Developing an understanding an introduction to biological of proteomics: Mass spectrometry. Cancer Investigation 2005; 23(1):47-59.

(54) Stephens WE. A Pulsed Mass Spectrometer with Time Dispersion. Physical Review 1946; 69(11-1):691.

(55) Hays EE, Richards PI, Goudsmit SA. Mass Measurements with A Magnetic Time-Of-Flight Mass Spectrometer. Physical Review 1951; 84(4):824-829.

(56) Wiley WC, Mclaren IH. Time-Of-Flight Mass Spectrometer with Improved Resolution. Review of Scientific Instruments 1955; 26(12):1150-1157.

(57) Cotter RJ. Time-Of-Flight Mass-Spectrometry for the Structural-Analysis of Biological Molecules. Analytical Chemistry 1992; 64(21):A1027-A1039.

(58) Paul W, Steinwedel H. *Ein Neues Massenspektrometer Ohne Magnetfeld. Zeitschrift fur Naturforschung Section A-A Journal of Physical Sciences 1953; 8(7):448-450.

(59) Stafford GC, Kelley PE, Syka JEP, Reynolds WE, Todd JFJ. Recent Improvements in and Analytical Applications of Advanced Ion Trap Technology. International Journal of Mass Spectrometry and Ion Processes 1984; 60(SEP):85-98.

(60) Louris JN, Cooks RG, Syka JEP, Kelley PE, Stafford GC, Todd JFJ. Instrumentation, Applications, and Energy Deposition in Quadrupole Ion-Trap Tandem Mass-Spectrometry. Analytical Chemistry 1987; 59(13):1677-1685.

(61) Jonscher K, Currie G, Mccormack AL, Yates JR. Matrix-Assisted Laser Desorption of Peptides and Proteins on A Quadrupole Ion Trap Mass-Spectrometer. Rapid Communications in Mass Spectrometry 1993; 7(1):20-26.

(62) Pappin DJC, Hojrup P, Bleasby AJ. Rapid Identification of Proteins by Peptide-Mass Finger Printing (Vol 3, Pg 327, 1993). Current Biology 1993; 3(7):487.

(63) Henry KD, Williams ER, Wang BH, Mclafferty FW, Shabanowitz J, Hunt DF. Fourier-Transform Mass-Spectrometry of Large Molecules by Electrospray Ionization. Proceedings of the National Academy of Sciences of the United States of America 1989; 86(23):9075-9078.

(64) Beu SC, Senko MW, Quinn JP, Mclafferty FW. Improved Fourier-Transform Ion-Cyclotron-Resonance Mass-Spectrometry of Large Biomolecules. Journal of the American Society for Mass Spectrometry 1993; 4(2):190-192.

(65) Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM. Electrospray Ionization for Mass-Spectrometry of Large Biomolecules. Science 1989; 246(4926):64-71.

(66) Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM. Electrospray Ionization-Principles and Practice. Mass Spectrometry Reviews 1990; 9(1):37-70.

(67) Wilm M, Mann M. Analytical properties of the nanoelectrospray ion source. Analytical Chemistry 1996; 68(1):1-8.

(68) Korner R, Wilm M, Morand K, Schubert M, Mann M. Nano electrospray combined with a quadrupole ion trap for the analysis of peptides and protein digests. Journal of the American Society for Mass Spectrometry 1996; 7(2):150-156.

(69) Valaskovic GA, Kelleher NL, Mclafferty FW. Attomole protein characterization by capillary electrophoresis mass spectrometry. Science 1996; 273(5279):1199-1202.

(70) Wilm M, Shevchenko A, Houthaeve T, Breit S, Schweigerer L, Fotsis T et al. Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry. Nature 1996; 379(6564):466-469.

(71) Lebrilla CB, Wang DTS, Hunter RL, Mciver RT. Detection of Mass 31830 Ions with An External Ion-Source Fourier-Transform Mass-Spectrometer. Analytical Chemistry 1990; 62(8):878-880.

(72) Huang XH, Shear JB, Zare RN. Quantitation of Ribonucleotides from Base-Hydrolyzed Rna Using Capillary Zone Electrophoresis. Analytical Chemistry 1990; 62(18):2049-2051.

(73) Loo JA, Williams ER, Amster IJ, Furlong JJP, Wang BH, Mclafferty FW et al. Cf-252 Plasma Desorption with Fourier-Transform Mass-Spectrometry. Analytical Chemistry 1987; 59(14):1880-1882.

(74) Bowers WD, Delbert SS, Hunter RL, Mciver RT. Fragmentation of Oligopeptide Ions Using Ultraviolet-Laser Radiation and Fourier-Transform Mass-Spectrometry. Journal of the American Chemical Society 1984; 106(23):7288-7289.

(75) Mccrery DA, Gross ML. Laser Desorption Fourier-Transform Mass-Spectrometry for the Study of Nucleosides, Oligosaccharides, and Glycosides. Analytica Chimica Acta 1985; 178(1):91-103.

(76) Karas M, Hillenkamp F. Laser Desorption Ionization of Proteins with Molecular Masses Exceeding 10000 Daltons. Analytical Chemistry 1988; 60(20):2299-2301.

(77) Chapman S. Carrier mobility spectra of spray electrified liquids. Physical Review 1937; 52(3):0184-0190.

(78) Dole M, Mack LL, Hines RL. Molecular Beams of Macroions. Journal of Chemical Physics 1968; 49(5):2240-&.

(79) Yamashita M, Fenn JB. Electrospray Ion-Source - Another Variation on the Free-Jet Theme. Journal of Physical Chemistry 1984; 88(20):4451-4459.

(80) Yamashita M, Fenn JB. Negative-Ion Production with the Electrospray Ion-Source. Journal of Physical Chemistry 1984; 88(20):4671-4675.

(81) Ikonomou MG, Blades AT, Kebarle P. Investigations of the Electrospray Interface for Liquid-Chromatography Mass-Spectrometry. Analytical Chemistry 1990; 62(9):957-967.

(82) Cahn JW. Stability of Electrically Charged Conducting Droplets. Physics of Fluids 1962; 5(12):1662-1663.

(83) Gomez A, Tang KQ. Charge and Fission of Droplets in Electrostatic Sprays. Physics of Fluids 1994; 6(1):404-414.

(84) Cole RB. Some tenets pertaining to electrospray ionization mass spectrometry. Journal of Mass Spectrometry 2000; 35(7):763-772.

(85) Smith RD, Lightwahl KJ. The Observation of Noncovalent Interactions in Solution by Electrospray-Ionization Mass-Spectrometry - Promise, Pitfalls and Prognosis. Biological Mass Spectrometry 1993; 22(9):493-501.

(86) Dulcks T, Juraschek R. Electrospray as an ionisation method for mass spectrometry. Journal of Aerosol Science 1999; 30(7):927-943.

(87) Juraschek R, Dulcks T, Karas M. Nanoelectrospray - More than just a minimized-flow electrospray ionization source. Journal of the American Society for Mass Spectrometry 1999; 10(4):300-308.

(88) Iribarne JV, Thomson BA. Evaporation of Small Ions from Charged Droplets. Journal of Chemical Physics 1976; 64(6):2287-2294.

(89) Kebarle P, Tang L. From Ions in Solution to Ions in the Gas-Phase - the Mechanism of Electrospray Mass-Spectrometry. Analytical Chemistry 1993; 65(22):A972-A986.

(90) Gamero-Castano M, de la Mora JF. Mechanisms of electrospray ionization of singly and multiply charged salt clusters. Analytica Chimica Acta 2000; 406(1):67-91.

(91) Wilm MS, Mann M. Electrospray and Taylor-Cone Theory, Doles Beam of Macromolecules at Last. International Journal of Mass Spectrometry 1994; 136(2-3):167-180.

(92) Emmett MR, Caprioli RM. Micro-Electrospray Mass-Spectrometry - Ultra-High-Sensitivity Analysis of Peptides and Proteins. Journal of the American Society for Mass Spectrometry 1994; 5(7):605-613.

(93) Andren PE, Emmett MR, Caprioli RM. Micro-Electrospray - Zeptomole-Attomole Per Microliter Sensitivity for Peptides. Journal of the American Society for Mass Spectrometry 1994; 5(9):867-869.

(94) Davis MT, Stahl DC, Hefta SA, Lee TD. A Microscale Electrospray Interface for Online, Capillary Liquid-Chromatography Tandem Mass-Spectrometry of Complex Peptide Mixtures. Analytical Chemistry 1995; 67(24):4549-4556.

(95) Gatlin CL, Kleemann GR, Hays LG, Link AJ, Yates JR. Protein identification at the low femtomole level from silver-stained gels using a new fritless electrospray interface for liquid chromatography microspray and nanospray mass spectrometry. Analytical Biochemistry 1998; 263(1):93-101.

(96) de la Mora JF. Electrospray ionization of large multiply charged species proceeds via Dole's charged residue mechanism. Analytica Chimica Acta 2000; 406(1):93-104.

(97) Scott RPW. Modern Liquid-Chromatography. Chemical Society Reviews 1992; 21(2):137-145.

(98) van Deemter J.J., Zuiderweg F.J., Klinkenberg A. Longitudinal diffusion and resistance to mass transfer as causes of nonideality in chromatography. Chemical Engineering Science 5, 271-289. 1956.

(99) Schneiderheinze J, Walden Z, Dufleld R, Demarest C. Rapid online proteolytic mapping of PEGylated rhGH for identity confirmation, quantitation of methionine oxidation and quantitation of UnPEGylated N-terminus using HPLC with UV detection. Journal of Chromatography B 2009; 877(31):4065-4070.

(100) Eggink M, Romero W, Vreuls RJ, Lingeman H, Niessen WMA, Irth H. Development and optimization of a system for comprehensive two-dimensional liquid chromatography with UV and mass spectrometric detection for the separation of complex samples by multi-step gradient elution. Journal of Chromatography A 2008; 1188(2):216-226.

(101) Qiao XQ, Wang L, Ma JF, Deng QL, Liang Z, Zhang LH et al. High sensitivity analysis of water-soluble, cyanine dye labeled proteins by high-performance liquid chromatography with fluorescence detection. Analytica Chimica Acta 2009; 640(1-2):114-120.

(102) Hirter P, Walther HJ, Datwyler P. Micro Column Liquid-Chromatography Mass-Spectrometry Using A Capillary Interface. Journal of Chromatography 1985; 323(1):89-98.

(103) Apffel JA, Brinkman UAT, Frei RW, Evers EAIM. Gas-Nebulized Direct Liquid Introduction Interface for Liquid-Chromatography Mass-Spectrometry. Analytical Chemistry 1983; 55(14):2280-2284.

(104) Erni F. Liquid Chromatography-Mass Spectrometry in the Pharmaceutical-Industry - Objectives and Needs. Journal of Chromatography 1982; 251(2):141-151.

(105) Noga M, Sucharski F, Suder P, Silberring J. A practical guide to nano-LC troubleshooting. Journal of Separation Science 2007; 30:2179-2189.

(106) Wolters DA, Washburn MP, Yates JR. An automated multidimensional protein identification technology for shotgun proteomics. Analytical Chemistry 2001; 73(23):5683-5690.

(107) Huang YF, Huang CC, Hu CC, Chang HT. Capillary electrophoresis-based separation techniques for the analysis of proteins. Electrophoresis 2006; 27(18):3503-3522.

(108) Jabeen R, Payne D, Wiktorowcz J, Mohammad A, Petersen J. Capillary electrophoresis and the clinical laboratory. Electrophoresis 2006; 27(12):2413-2438.

(109) Hutterer K, Dolnik V. Capillary electrophoresis of proteins 2001-2003. Electrophoresis 2003; 24(22-23):3998-4012.

(110) Hapuarachchi S, Janaway GA, Aspinwall CA. Capillary electrophoresis with a UV light-emitting diode source for chemical monitoring of native and derivatized fluorescent compounds. Electrophoresis 2006; 27(20):4052-4059.

(111) Kolch W, Neususs C, Peizing M, Mischak H. Capillary electrophoresis - Mass spectrometry as a powerful tool in clinical diagnosis and biomarker discovery. Mass Spectrometry Reviews 2005; 24(6):959-977.

(112) Mischak H, Kaiser T, Walden M, Hillmann M, Wittke S, Herrmann A et al. Proteomic analysis for the assessment of diabetic renal damage in humans. Clinical Science 2004; 107(5):485-495.

(113) Wittke S, Haubitz M, Walden M, Rohde F, Schwarz A, Mengel M et al. Detection of acute tubulointerstitial rejection by proteomic analysis of urinary samples in renal transplant recipients. American Journal of Transplantation 2005; 5(10):2479-2488.

(114) Kolch W, Neususs C, Peizing M, Mischak H. Capillary electrophoresis - Mass spectrometry as a powerful tool in clinical diagnosis and biomarker discovery. Mass Spectrometry Reviews 2005; 24(6):959-977.

(115) Meier M, Kaiser T, Herrmann A, Knueppel S, Hillmann M, Koester P et al. Identification of urinary protein pattern in Type 1 diabetic adolescents with early diabetic nephropathy by a novel combined proteome analysis. Journal of Diabetes and Its Complications 2005; 19(4):223-232.

(116) Ramstrom M, Bergquist J. Miniatuized proteomics and pepitdomics using capillary liquid separation and high mass spectrometry. Febs Letters 2004; 567(1):92-95.

(117) Marouga R, David S, Hawkins E. The development of the DIGE system: 2D fluorescence difference gel analysis technology. Analytical and Bioanalytical Chemistry 2005; 382(3):669-678.

(118) Shaw J, Rowlinson R, Nickson J, Stone T, Sweet A, Williams K et al. Evaluation of saturation labelling two-dimensional difference gel electrophoresis fluorescent dyes. Proteomics 2003; 3(7):1181-1195.

(119) Steinberg TH, Jones LJ, Haugland RP, Singer VL. SYPRO Orange and SYPRO Red protein gel stains: One-step fluorescent staining of denaturing gels for detection of nanogram levels of protein. Analytical Biochemistry 1996; 239(2):223-237.

(120) Tonge R, Shaw J, Middleton B, Rowlinson R, Rayner S, Young J et al. Validation and development of fluorescence two-dimensional differential gel electrophoresis proteomics technology. Proteomics 2001; 1(3):377-396.

(121) Wheelock AM, Morin D, Bartosiewicz M, Buckpitt AR. Use of a fluorescent internal protein standard to achieve quantitative two-dimensional gel electrophoresis. Proteomics 2006; 6(5):1385-1398.

(122) Chen DY, Dovichi NJ. Yoctomole Detection Limit by Laser-Induced Fluorescence in Capillary Electrophoresis. Journal of Chromatography B-Biomedical Applications 1994; 657(2):265-269.

(123) Zhao JY, Chen DY, Dovichi NJ. Low-Cost Laser-Induced Fluorescence Detector for Micellar Capillary Zone Electrophoresis - Detection at the Zeptomol Level of Tetramethylrhodamine Thiocarbamyl Amino-Acid Derivatives. Journal of Chromatography 1992; 608(1-2):117-120.

(124) Wu SO, Dovichi NJ. Capillary Zone Electrophoresis Separation and Laser-Induced Fluorescence Detection of Zeptomole Quantities of Fluorescein Thiohydantoin Derivatives of Amino-Acids. Talanta 1992; 39(2):173-178.

(125) Weiss S. Fluorescence spectroscopy of single biomolecules. Science 1999; 283(5408):1676-1683.

(126) Kinsey JL. Laser-Induced Fluorescence. Annual Review of Physical Chemistry 1977; 28:349-372.

(127) Vrabel P, Taborsky P, Ryvolova M, Havel J, Preisler J. Sensitive detection and separation of fluorescent derivatives using capillary electrophoresis with laser-induced fluorescence detection with 532 nm Nd : YAG laser. Journal of Luminescence 2006; 118(2):283-292.

(128) Mujumdar SR, Mujumdar RB, Grant CM, Waggoner AS. Cyanine-labeling reagents: Sulfobenzindocyanine succinimidyl esters. Bioconjugate Chemistry 1996; 7(3):356-362.

(129) Holler MG, Campo LF, Brandelli A, Stefani V. Synthesis and spectroscopic characterisation of 2-(2 '-hydroxyphenyl)benzazole isothiocyanates as new fluorescent probes for proteins. Journal of Photochemistry and Photobiology A-Chemistry 2002; 149(1-3):217-225.

(130) Smith SN, Steer RP. The photophysics of Lissamine rhodamine-B sulphonyl chloride in aqueous solution: implications for fluorescent protein-dye conjugates. Journal of Photochemistry and Photobiology A-Chemistry 2001; 139(2-3):151-156.

(131) Zhao SL, Yuan HY, Xiao D. Optical fiber light-emitting diode-induced fluorescence detection for capillary electrophoresis. Electrophoresis 2006; 27(2):461-467.

(132) McKeon J, Khaledi MG. Quantitative nuclear and cytoplasmic localization of antisense oligonucleotides by capillary electrophoresis with laser-induced fluorescence detection. Electrophoresis 2001; 22(17):3765-3770.

(133) Khan HA. Detection and semi-quantitative determination of low abundance GFAP mRNA in mouse brain by capillary electrophoresis coupled with laser-induced fluorescence. Brain Research Protocols 2004; 14(1):13-17.

(134) Roe MR, Griffin TJ. Gel-free mass spectrometry-based high throughput proteomics: Tools for studying biological response of proteins and proteomes. Proteomics 2006; 6(17):4678-4687.

(135) Kolkman A, Dirksen EHC, Slijper M, Heck AJR. Double standards in quantitative proteomics - Direct comparative assessment of difference in gel electrophoresis and metabolic stable isotope labeling. Molecular & Cellular Proteomics 2005; 4(3):255-266.

(136) Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nature Biotechnology 1999; 17(10):994-999.

(137) Gygi SP, Rist B, Griffin TJ, Eng J, Aebersold R. Proteome analysis of low-abundance proteins using multidimensional chromatography and isotope-coded affinity tags. Journal of Proteome Research 2002; 1(1):47-54.

(138) Smolka MB, Zhou HL, Purkayastha S, Aebersold R. Optimization of the isotope-coded affinity tag-labeling procedure for quantitative proteome analysis. Analytical Biochemistry 2001; 297(1):25-31.

(139) Thompson A, Schafer J, Kuhn K, Kienle S, Schwarz J, Schmidt G et al. Tandem mass tags: A novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. Analytical Chemistry 2003; 75(8):1895-1904.

(140) Bottari P, Aebersold R, Turecek F, Gelb MH. Design and synthesis of visible isotope-coded affinity tags for the absolute quantification of specific proteins in complex mixtures. Bioconjugate Chemistry 2004; 15(2):380-388.

(141) Jenkins RE, Kitteringham NR, Hunter CL, Webb S, Hunt TJ, Elsby R et al. Relative and absolute quantitative expression profiling of cytochromes P450 using isotope-coded affinity tags. Proteomics 2006; 6(6):1934-1947.

(142) Qu J, Straubinger RM. Improved sensitivity for quantification of proteins using triply charged cleavable isotope-coded affinity tag peptides. Rapid Communications in Mass Spectrometry 2005; 19(19):2857-2864.

(143) Wu WW, Wang GH, Baek SJ, Shen RF. Comparative study of three proteomic quantitative methods, DIGE, cICAT, and iTRAQ, using 2D gel- or LC-MALDI TOF/TOF. Journal of Proteome Research 2006; 5(3):651-658.

(144) Ross PL, Huang YLN, Marchese JN, Williamson B, Parker K, Hattan S et al. Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. Molecular & Cellular Proteomics 2004; 3(12):1154-1169.

(145) Zieske LR. A perspective on the use of iTRAQ (TM) reagent technology for protein complex and profiling studies. Journal of Experimental Botany 2006; 57(7):1501-1508.

(146) Mann M. Functional and quantitative proteomics using SILAC. Nature Reviews Molecular Cell Biology 2006; 7(12):952-958.

(147) Blagoev B, Kratchmarova I, Ong SE, Nielsen M, Foster LJ, Mann M. A proteomics strategy to elucidate functional protein-protein interactions applied to EGF signaling. Nature Biotechnology 2003; 21(3):315-318.

(148) Ong SE, Kratchmarova I, Mann M. Properties of C-13-substituted arginine in stable isotope labeling by amino acids in cell culture (SILAC). Journal of Proteome Research 2003; 2(2):173-181.

(149) Ong SE, Foster LJ, Mann M. Mass spectrometric-based approaches in quantitative proteomics. Methods 2003; 29(2):124-130.

(150) Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A et al. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Molecular & Cellular Proteomics 2002; 1(5):376-386.

(151) Molina H, Parmigiani G, Pandey A. Assessing reproducibility of a protein dynamics study using in vivo labeling and liquid chromatography tandem mass spectrometry. Analytical Chemistry 2005; 77(9):2739-2744.

(152) Bantscheff M, Dumpelfeld B, Kuster B. Femtomol sensitivity post-digest O-18 labeling for relative quantification of differential protein complex composition. Rapid Communications in Mass Spectrometry 2004; 18(8):869-876.

(153) Sakai J, Kojima S, Yanagi K, Kanaoka M. O-18-labeling quantitative proteomics using an ion trap mass spectrometer. Proteomics 2005; 5(1):16-23.

(154) Lee YH, Han H, Chang SB, Lee SW. Isotope-coded N-terminal sulfonation of peptides allows quantitative proteomic analysis with increased de novo peptide sequencing capability. Rapid Communications in Mass Spectrometry 2004; 18(24):3019-3027.

(155) Wienkoop S, Larrainzar E, Niemann M, Gonzalez EM, Lehmann U, Weckwerth W. Stable isotope-free quantitative shotgun proteomics combined with sample pattern recognition for rapid diagnostics. Journal of Separation Science 2006; 29(18):2793-2801.

(156) Roy SM, Anderle M, Lin H, Becker CH. Differential expression profiling of serum proteins and metabolites for biomarker discovery. International Journal of Mass Spectrometry 2004; 238(2):163-171.

(157) Silva JC, Denny R, Dorschel CA, Gorenstein M, Kass IJ, Li GZ et al. Quantitative proteomic analysis by accurate mass retention time pairs. Analytical Chemistry 2005; 77(7):2187-2200.

(158) Schmidt A, Karas M, Dulcks T. Effect of different solution flow rates on analyte ion signals in nano-ESI MS, or: When does ESI turn into nano-ESI? Journal of the American Society for Mass Spectrometry 2003; 14(5):492-500.

(159) Shen YF, Zhao R, Berger SJ, Anderson GA, Rodriguez N, Smith RD. High-efficiency nanoscale liquid chromatography coupled on-line with mass spectrometry using nanoelectrospray ionization for proteomics. Analytical Chemistry 2002; 74(16):4235-4249.

(160) Old WM, Meyer-Arendt K, Aveline-Wolf L, Pierce KG, Mendoza A, Sevinsky JR et al. Comparison of label-free methods for quantifying human proteins by shotgun proteomics. Molecular & Cellular Proteomics 2005; 4(10):1487-1502.

(161) America AHP, Cordewener JHG, van Geffen MHA, Lommen A, Vissers JPC, Bino RJ et al. Alignment and statistical difference analysis of complex peptide data sets generated by multidimensional LC-MS. Proteomics 2006; 6(2):641-653.

(162) Liu T, Qian WJ, Strittmatter EF, Camp DG, Anderson GA, Thrall BD et al. High-throughput comparative proteome analysis using a quantitative cysteinyl-peptide enrichment technology. Analytical Chemistry 2004; 76(18):5345-5353.

(163) Wang GH, Wu WW, Zeng WH, Chou CL, Shen RF. Label-free protein quantification using LC-coupled ion trap or FT mass spectrometry: Reproducibility, linearity, and application with complex proteomes. Journal of Proteome Research 2006; 5(5):1214-1223.

(164) America AHP, Cordewener JHG. Comparative LC-MS: A landscape of peaks and valleys. Proteomics 2008; 8(4):731-749.

(165) Simpson KL, Whetton AD, Dive C. Quantitative mass spectrometry-based techniques for clinical use: Biomarker identification and quantification. Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences 2009; 877(13):1240-1249.

(166) Ru QHC, Zhu LWA, Silberman J, Shriver CD. Label-free semiquantitative peptide feature profiling of human breast cancer and breast disease sera via two-dimensional liquid chromatography-mass spectrometry. Molecular & Cellular Proteomics 2006; 5(6):1095-1104.

(167) Patwardhan AJ, Strittmatter EF, David GC, Smith RD, Pallavicini MG. Quantitative proteome analysis of breast cancer cell lines using O-18-labeling and an accurate mass and time tag strategy. Proteomics 2006; 6(9):2903-2915.

(168) Vissers JPC, Langridge JI, Aerts JMFG. Analysis and quantification of diagnostic serum markers and protein signatures for Gaucher disease. Molecular & Cellular Proteomics 2007; 6(5):755-766.

(169) Kemperman RFJ, Horvatovich PL, Hoekman B, Reijmers TH, Muskiet FAJ, Bischoff R. Comparative urine analysis by liquid chromatography-mass spectrometry and multivariate statistics: Method development, evaluation, and application to proteinuria. Journal of Proteome Research 2007; 6(1):194-206.

(170) Huang JTJ, McKenna T, Hughes C, Leweke FM, Schwarz E, Bahn S. CSF biomarker discovery using label-free nano-LC-MS based proteomic profiling: Technical aspects. Journal of Separation Science 2007; 30(2):214-225.

(171) Metz TO, Qian WJ, Jacobs JM, Gritsenko MA, Moore RJ, Polpitiya AD et al. Application of proteomics in the discovery of candidate protein biomarkers in a diabetes autoantibody standardization program sample subset. Journal of Proteome Research 2008; 7(2):698-707.

(172) Canas B, Pineiro C, Calvo E, Lopez-Ferrer D, Gallardo JM. Trends in sample preparation for classical and second generation proteomics. Journal of Chromatography A 2007; 1153(1-2):235-258.

(173) Tantipaiboonwong P, Sinchaikul S, Sriyam S, Phutrakul S, Chen ST. Different techniques for urinary protein analysis of normal and lung cancer patients. Proteomics 2005; 5(4):1140-1149.

(174) Zellner M, Winkler W, Hayden H, Diestinger M, Eliasen M, Gesslbauer B et al. Quantitative validation of different protein precipitation methods in proteome analysis of blood platelets. Electrophoresis 2005; 26(12):2481-2489.

(175) Huber LA, Pfaller K, Vietor I. Organelle proteomics - Implications for subcellular fractionation in proteomics. Circulation Research 2003; 92(9):962-968.

(176) Klose J, Link A. Fractionated extraction of total tissue proteins from mouse and human for 2-D electrophoresis. Methods in Molecular Biology; 2-D proteome analysis protocols 1999;67-85.

(177) Pasquali C, Fialka I, Huber LA. Subcellular fractionation, electromigration analysis and mapping of organelles. Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences 1999; 722(1-2):89-102.

(178) Stasyk T, Huber LA. Zooming in: Fractionation strategies in proteomics. Proteomics 2004; 4(12):3704-3716.

(179) Wu CC, MacCoss MJ, Mardones G, Finnigan C, Mogelsvang S, Yates JR et al. Organellar proteomics reveals Golgi arginine dimethylation. Molecular Biology of the Cell 2004; 15(6):2907-2919.

(180) Rice GE, Georgiou HM, Ahmed N, Shi G, Kruppa G. Translational proteomics: Developing a predictive capacity - A review. Placenta 2006; 27:S76-S86.

(181) Zhang X, Leung SM, Morris CR, Shigenaga MK. Evaluation of a Novel, Integrated Approach Using Functionalized Magnetic Beads, Bench-Top MALDI-TOF-MS with Prestructured Sample Supports, and Pattern Recognition Software for Profiling Potential Biomarkers in Human Plasma. J Biomol Tech 2004; 15(3):167-175.

(182) Merchant M, Weinberger SR. Recent advancements in surface-enhanced laser desorption/ionization-time of flight-mass spectrometry. Electrophoresis 2000; 21(6):1164-1177.

(183) Luque-Garcia JL, Neubert TA. Sample preparation for serum/plasma profiling and biomarker identification by mass spectrometry. Journal of Chromatography A 2007; 1153(1-2):259-276.

(184) Tammen H, Schulte L, Hess R, Menzel C, Kellmann M, Mohring T et al. Peptidomic analysis of human blood specimens: Comparison between plasma specimens and serum by differential peptide display. Proteomics 2005; 5(13):3414-3422.

(185) Tirumalai RS, Chan KC, Prieto DA, Issaq HJ, Conrads TP, Veenstra TD. Characterization of the low molecular weight human serum proteome. Molecular & Cellular Proteomics 2003; 2(10):1096-1103.

(186) Whiteaker JR, Zhang HD, Eng JK, Fang RH, Piening BD, Feng LC et al. Head-to-head comparison of serum fractionation techniques. Journal of Proteome Research 2007; 6(2):828-836.

(187) Granger J, Siddiqui J, Copeland S, Remick D. Albumin depletion of human plasma also removes low abundance proteins including the cytokines. Proteomics 2005; 5(18):4713-4718.

(188) Zhang H, Yi EC, Li XJ, Mallick P, Kelly-Spratt KS, Masselon CD et al. High throughput quantitative analysis of serum proteins using glycopeptide capture and liquid chromatography mass spectrometry. Molecular & Cellular Proteomics 2005; 4(2):144-155.

(189) Zhang H, Li XJ, Martin DB, Aebersold R. Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. Nature Biotechnology 2003; 21(6):660-666.

(190) Roth J. Protein N-glycosylation along the secretory pathway: Relationship to organelle topography and function, protein quality control, and cell interactions. Chemical Reviews 2002; 102(2):285-303.

(191) Murray CJL, Lopez AD. Mortality by cause for eight regions of the world: Global Burden of Disease Study. Lancet 1997; 349(9061):1269-1276.

(192) Donnan GA, Fisher M, Macleod M, Davis SM. Stroke. Lancet 2008; 371(9624):1612-1623.

(193) Bonita R. Epidemiology of Stroke. Lancet 1992; 339(8789):342-344.

(194) Lopez AD, Mathers CD, Ezzati M, Jamison DT, Murray CJL. Global and regional burden of disease and risk factors, 2001: systematic analysis of population health data. Lancet 2006; 367(9524):1747-1757.

(195) Heart and Stroke Facts Statistics: Dallas: American Heart Association. American Heart Association . 1997.

(196) Rothwell PM. The high cost of not funding stroke research: a comparison with heart disease and cancer. Lancet 2001; 357(9268):1612-1616.

(197) Whisnant JP. Modeling of risk factors for ischemic stroke - The Willis Lecture. Stroke 1997; 28(9):1840-1844.

(198) Yusuf S, Hawken S, Ounpuu S. Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study (vol 364, pg 937, 2004). Lancet 2004; 364(9449):1938.

(199) Fiebach JB, Schellinger PD, Jansen O, Meyer M, Wilde P, Bender J et al. CT and diffusion-weighted MR imaging in randomized order - Diffusion-weighted imaging results in higher accuracy and lower interrater variability in the diagnosis of hyperacute ischemic stroke. Stroke 2002; 33(9):2206-2210.

(200) Auer RN, Sutherland GR. Primary intracerebral hemorrhage: Pathophysiology. Canadian Journal of Neurological Sciences 2005; 32:S3-S12.

(201) Mead GE, Wardlaw JM, Dennis MS, Lewis SC. Extensive haemorrhagic transformation of infarct: might it be an important cause of primary intracerebral haemorrhage? Age and Ageing 2002; 31(6):429-433.

(202) Thrift AG, Donnan GA, Mcneil JJ. Epidemiology of intracerebral hemorrhage. Epidemiologic Reviews 1995; 17(2):361-381.

(203) Thrift AG, Dewey HM, Macdonell RAL, Mcneil JJ, Donnan GA. Incidence of the major stroke subtypes - Initial findings from the North East Melbourne Stroke Incidence Study (NEMESIS). Stroke 2001; 32(8):1732-1738.

(204) Fisher M, Garcia JH. Evolving stroke and the ischemic penumbra. Neurology 1996; 47(4):884-888.

(205) Dirnagl U, Iadecola C, Moskowitz MA. Pathobiology of ischaemic stroke: an integrated view. Trends in Neurosciences 1999; 22(9):391-397.

(206) Fisher M. The ischemic penumbra: Identification, evolution and treatment concepts. Cerebrovascular Diseases 2004; 17:1-6.

(207) Hankey GJ, Jamrozik K, Broadhurst RJ, Forbes S, Burvill PW, Anderson CS et al. Long-term risk of first recurrent stroke in the Perth Community Stroke Study. Stroke 1998; 29(12):2491-2500.

(208) Weimar C, Ziegler A, Konig IR, Diener HC, German Stroke SC. Predicting functional outcome and survival after acute ischemic stroke. Journal of Neurology 2002; 249(7):888-895.

(209) Donnan GA, Omalley HM, Quang L, Hurley S, Bladin PF. The Capsular Warning Syndrome - Pathogenesis and Clinical-Features. Neurology 1993; 43(5):957-962.

(210) Johnston SC, Gress DR, Browner WS, Sidney S. Short-term prognosis after emergency department diagnosis of TIA. Jama-Journal of the American Medical Association 2000; 284(22):2901-2906.

(211) Coull AJ, Lovett JK, Rothwell PM, Oxford VS. Population based study of early risk of stroke after transient ischaemic attack or minor stroke: implications for public education and organisation of services. British Medical Journal 2004; 328(7435):326-328.

(212) Rothwell PM, Buchan A, Johnston SC. Recent advances in management of transient ischaemic attacks and minor ischaemic strokes. Lancet Neurology 2006; 5(4):323-331.

(213) Coutts SB, Simon JE, Eliasziw M, Sohn CH, Hill MD, Barber PA et al. Triaging transient ischemic attack and minor stroke patients using acute magnetic resonance imaging. Annals of Neurology 2005; 57(6):848-854.

(214) Bonita R, Beaglehole R. Does treatment of hypertension explain the decline in mortality from stroke? British Medical Journal 292, 191-192. 1986.

(215) Bornstein N, Silvestrelli G, Caso V, Parnetti L. Arterial hypertension and stroke prevention: An update. Clinical and Experimental Hypertension 2006; 28(3-4):317-326.

(216) Bonita R, Beaglehole R. The Decline in Stroke Mortality - the Limited Role of Antihypertensive Therapy. New Zealand Medical Journal 1987; 100(828):454-456.

(217) Neal B, MacMahon S, Chapman N, Cutler J, Fagard R, Neal B et al. Effects of ACE inhibitors, calcium antagonists, and other blood-pressure-lowering drugs: results of prospectively designed overviews of randomised trials. Lancet 2000; 356(9246):1955-1964.

(218) Lawes CMM, Bennett DA, Feigin VL, Rodgers A. Blood pressure and stroke - An overview of published reviews (vol 35, pg 776, 2004). Stroke 2004; 35(4):1024-1033.

(219) Taylor FC, Cohen H, Ebrahim S. Systematic review of long term anticoagulation or antiplatelet treatment in patients with non-rheumatic atrial fibrillation. British Medical Journal 2001; 322(7282):321-326.

(220) Amarenco P, Labreuche J, Lavallee P, Touboul PJ. Statins in stroke prevention and carotid atherosclerosis - Systematic review and up-to-date meta-analysis. Stroke 2004; 35(12):2902-2909.

(221) Ridker PM, Cook NR, Lee IM, Gordon D, Gaziano JM, Manson JE et al. A randomized trial of low-dose aspirin in the primary prevention of cardiovascular disease in women. New England Journal of Medicine 2005; 352(13):1293-1304.

(222) A randomized trial of aspirin and sulfinpyrazone in threatened stroke. The Canadian Cooperative Study Group. New England Journal of Medicine 299, 53-59. 1978.

(223) Sandercock P, Collins R, Counsell C, Farrell B, Peto R, Slattery J et al. The International Stroke Trial (IST): A randomised trial of aspirin, subcutaneous heparin, both, or neither among 19 435 patients with acute ischaemic stroke. Lancet 1997; 349(9065):1569-1581.

(224) CAST: randomised placebo-controlled trial of early aspirin use in 20,000 patients with acute ischaemic stroke. CAST (Chinese Acute Stroke Trial) Collaborative Group. Lancet 1997; 349(9066):1641-1649.

(225) Gilligan AK, Thrift AG, Sturm JW, Dewey HM, Macdonell RAL, Donnan GA. Stroke units, tissue plasminogen activator, aspirin and neuroprotection: Which stroke intervention could provide the greatest community benefit? Cerebrovascular Diseases 2005; 20(4):239-244.

(226) De Cristobal J, Moro MA, Davalos A, Castillo J, Leza JC, Camarero J et al. Neuroprotective effect of aspirin by inhibition of glutamate release after permanent focal cerebral ischaemia in rats. Journal of Neurochemistry 2001; 79(2):456-459.

(227) Eur Stroke Prev Study Group. The European Stroke Prevention Study Esps Principal End-Points. Lancet 1987; 2(8572):1351-1354.

(228) Koudstaal PJ, Dehaene I, Dhooghe M, Marchau M, Vanzandijcke M, Delwaide C et al. Secondary Prevention in Nonrheumatic Atrial-Fibrillation After Transient Ischemic Attack Or Minor Stroke. Lancet 1993; 342(8882):1255-1262.

(229) Beneficial effect of carotid endarterectomy in symptomatic patients with high-grade carotid stenosis. North American Symptomatic Carotid Endarterectomy Trial Collaborators. N Engl J Med 1991; 325(7):445-453.

(230) Eur Carotid Surg Trialists' Collab Group. Mrc European Carotid Surgery Trial Interim Results for Symptomatic Patients with Severe 70-99 Percent Or with Mild 0-29 Percent Carotid Stenosis. Lancet 1991; 337(8752):1235-1243.

(231) Gent M, Beaumont D, Blanchard J, Bousser MG, Coffman J, Easton JD et al. A randomised, blinded, trial of clopidogrel versus aspirin in patients at risk of ischaemic events (CAPRIE). Lancet 1996; 348(9038):1329-1339.

(232) Randomised trial of a perindopril-based blood-pressure-lowering regimen among 6,105 individuals with previous stroke or transient ischaemic attack. Lancet 2001; 358(9287):1033-1041.

(233) Heart Outcomes PE. Effects of ramipril on cardiovascular and microvascular outcomes in people with diabetes mellitus: results of the HOPE study and MICROHOPE substudy (vol 355, pg 253, 2000). Lancet 2000; 356(9232):860.

(234) Amarenco P, Bogousslavsky J, Callahan A, Goldstein LB, Hennerici M, Rudolph AE et al. High-dose atorvastatin after stroke or transient ischemic attack. New England Journal of Medicine 2006; 355(6):549-559.

(235) Ringleb PA, Allenberg J, Berger J, Bruckmann H, Eckstein HH, Fraedrich G et al. 30 day results from the SPACE trial of stent-protected angioplasty versus carotid endarterectomy in symptomatic patients: a randomised non-inferiority trial. Lancet 2006; 368(9543):1239-1247.

(236) Brown MM, Rogers J, Bland JM, CAVATAS I. Endovascular versus surgical treatment in patients with carotid stenosis in the Carotid and Vertebral Artery Transluminal Angioplasty Study (CAVATAS): a randomised trial. Lancet 2001; 357(9270):1729-1737.

(237) Mas J, Chatellier G, Beyssen B, Branchereau A, Moulin T, Becquemin JP et al. Endarterectomy versus stenting in patients with symptomatic severe carotid stenosis. New England Journal of Medicine 2006; 355(16):1660-1671.

(238) Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). UK Prospective Diabetes Study (UKPDS) Group. Lancet 1998; 352(9131):837-853.

(239) Donnan GA, Adena MA, Omalley HM, Mcneil JJ, Doyle AE, Neill GC. Smoking As A Risk Factor for Cerebral-Ischemia. Lancet 1989; 2(8664):643-647.

(240) Reynolds K, Lewis LB, Nolen JDL, Kinney GL, Sathya B, He J. Alcohol consumption and risk of stroke - A meta-analysis. Jama-Journal of the American Medical Association 2003; 289(5):579-588.

(241) He FJ, Nowson CA, MacGregor GA. Fruit and vegetable consumption and stroke: meta-analysis of cohort studies. Lancet 2006; 367(9507):320-326.

(242) Wendel-Vos GCW, Schuit AJ, Feskens EJM, Boshuizen HC, Verschuren WMM, Saris WHM et al. Physical activity and stroke. A meta-analysis of observational data. International Journal of Epidemiology 2004; 33(4):787-798.

(243) Langhorne P, Williams BO, Gilchrist W, Howie K. Do Stroke Units Save Lives. Lancet 1993; 342(8868):395-398.

(244) Langhorne P, Dey P, Woodman M, Kalra L, Wood-Dauphinee S, Patel N et al. Is stroke unit care portable? A systematic review of the clinical trials. Age and Ageing 2005; 34(4):324-330.

(245) Indredavik B, Bakke F, Slordahl SA, Rokseth R, Haheim LL. Treatment in a combined acute and rehabilitation stroke unit - Which aspects are most important? Stroke 1999; 30(5):917-923.

(246) Cadilhac DA, Ibrahim J, Pearce DC, Ogden KJ, McNeill J, Davis SM et al. Multicenter comparison of processes of care between Stroke Units and conventional care wards in Australia. Stroke 2004; 35(5):1035-1040.

(247) Hacke W, Donnan G, Fieschi C, Kaste M, von Kummer R, Broderick JP et al. Association of outcome with early stroke treatment: pooled analysis of ATLANTIS, ECASS, and NINDS rt-PA stroke trials. Lancet 2004; 363(9411):768-774.

(248) Frank JI. Large Hemispheric Infarction, Deterioration, and Intracranial-Pressure. Neurology 1995; 45(7):1286-1290.

(249) Davis SM, Broderick J, Hennerici M, Brun NC, Diringer MN, Mayer SA et al. Hematoma growth is a determinant of mortality and poor outcome after intracerebral hemorrhage. Neurology 2006; 66(8):1175-1181.

(250) Mayer SA, Brun NC, Begtrup K, Broderick J, Davis S, Diringer MN et al. Recombinant activated factor VII for acute intracerebral hemorrhage. New England Journal of Medicine 2005; 352(8):777-785.

(251) Smith WS, Sung G, Starkman S, Saver JL, Kidwell CS, Gobin YP et al. Safety and efficacy of mechanical embolectomy in acute ischemic stroke - Results of the MERCI trial. Stroke 2005; 36(7):1432-1438.

(252) Furlan AJ, Fisher M. Devices, drugs, and the Food and Drug Administration - Increasing implications for ischemic stroke. Stroke 2005; 36(2):398-399.

(253) Fisher M, Stroke Therapy Acad Ind Roundtable. Enhancing the development and approval of acute stroke therapies - Stroke Therapy Academic Industry Roundtable. Stroke 2005; 36(8):1808-1813.

(254) Chaurand P, Stoeckli M, Caprioli RM. Direct profiling of proteins in biological tissue sections by MALDI mass spectrometry. Analytical Chemistry 1999; 71(23):5263-5270.

(255) Stoeckli M, Chaurand P, Hallahan DE, Caprioli RM. Imaging mass spectrometry: A new technology for the analysis of protein expression in mammalian tissues. Nature Medicine 2001; 7(4):493-496.

(256) Chaurand P, Schwartz SA, Caprioli RM. Imaging mass spectrometry: a new tool to investigate the spatial organization of peptides and proteins in mammalian tissue sections. Current Opinion in Chemical Biology 2002; 6(5):676-681.

(257) Chaurand P, Caprioli RM. Direct profiling and imaging of peptides and proteins from mammalian cells and tissue sections by mass spectrometry. Electrophoresis 2002; 23(18):3125-3135.

(258) Murphy RC, Hankin JA, Barkley RM. Imaging of lipid species by MALDI mass spectrometry. Journal of Lipid Research 2009; 50:S317-S322.

(259) Andren P. Frontier and perspectives - Specific and sensitive identification of peptides using peptidomics and MALDI imaging mass spectrometry. Journal of Peptide Science 2008; 14(8):40.

(260) Goodwin RJA, Pennington SR, Pitt AR. Protein and peptides in pictures: Imaging with MALDI mass spectrometry. Proteomics 2008; 8(18):3785-3800.

(261) Cornett DS, Mobley JA, Dias EC, Andersson M, Arteaga CL, Sanders ME et al. A novel histology-directed strategy for MALDI-MS tissue profiling that improves throughput and

cellular specificity in human breast cancer. Molecular & Cellular Proteomics 2006; 5(10):1975-1983.

(262) Meistermann H, Norris JL, Aerni HR, Cornett DS, Friedlein A, Erskine AR et al. Biomarker discovery by imaging mass spectrometry - Transthyretin is a biomarker for gentamicin-induced nephrotoxicity in rat. Molecular & Cellular Proteomics 2006; 5(10):1876-1886.

(263) Skold K, Svensson M, Nilsson A, Zhang XQ, Nydahl K, Caprioli RM et al. Decreased striatal levels of PEP-19 following MPTP lesion in the mouse. Journal of Proteome Research 2006; 5(2):262-269.

(264) Chaurand P, Rahman MA, Hunt T, Mobley JA, Gu G, Latham JC et al. Monitoring mouse prostate development by profiling and imaging mass spectrometry. Molecular & Cellular Proteomics 2008; 7(2):411-423.

(265) Shimma S, Sugiura Y, Hayasaka T, Hoshikawa Y, Noda T, Setou M. MALDI-based imaging mass spectrometry revealed abnormal distribution of phospholipids in colon cancer liver metastasis. Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences 2007; 855(1):98-103.

(266) Reyzer ML, Caprioli RM. MALDI mass spectrometry for direct tissue analysis: A new tool for biomarker discovery. Journal of Proteome Research 2005; 4(4):1138-1142.

(267) Meistermann H, Norris JL, Aerni HR, Cornett DS, Friedlein A, Erskine AR et al. Biomarker discovery by imaging mass spectrometry - Transthyretin is a biomarker for gentamicin-induced nephrotoxicity in rat. Molecular & Cellular Proteomics 2006; 5(10):1876-1886.

(268) Schwartz SA, Weil RJ, Thompson RC, Shyr Y, Moore JH, Toms SA et al. Proteomic-based prognosis of brain tumor patients using direct-tissue matrix-assisted laser desorption ionization mass spectrometry. Cancer Research 2005; 65(17):7674-7681.

(269) Roepstorff P, Fohlman J. Proposal for A Common Nomenclature for Sequence Ions in Mass-Spectra of Peptides. Biomedical Mass Spectrometry 1984; 11(11):601.

(270) Barton SJ, Whittaker JC. Review of Factors That Influence the Abundance of Ions Produced in A Tandem Mass Spectrometer and Statistical Methods for Discovering These Factors. Mass Spectrometry Reviews 2009; 28(1):177-187.

(271) Chen R, Brentnall TA, Pan S, Cooke K, Moyes KW, Lane Z et al. Quantitative proteomics analysis reveals that proteins differentially expressed in chronic pancreatitis are also frequently involved in pancreatic cancer. Molecular & Cellular Proteomics 2007; 6(8):1331-1342.

(272) Kikuchi T, Carbone DP. Proteomics analysis in lung cancer: Challenges and opportunities. Respirology 2007; 12(1):22-28.

(273) Lemhire R, Menguellet SA, Stauber J, Marchaudon V, Lucot JP, Collinet P et al. Specific MALDI imaging and profiling for biomarker hunting and validation: Fragment of the 11S proteasome activator complex, reg alpha fragment, is a new potential ovary cancer biomarker. Journal of Proteome Research 2007; 6(11):4127-4134.

(274) Pan S, Shi M, Jin JH, Albin RL, Lieberman A, Gearing M et al. Proteomics identification of proteins in human cortex using multidimensional separations and MALDI tandem mass spectrometer. Molecular & Cellular Proteomics 2007; 6(10):1818-1823.

(275) Goodwin RJA, Dungworth JC, Cobb SR, Pitt AR. Time-dependent evolution of tissue markers by MALDI-MS imaging. Proteomics 2008; 8(18):3801-3808.

(276) Skold K, Svensson M, Norrman M, Sjogren B, Svenningsson P, Andren PE. The significance of biochemical and molecular sample integrity in brain proteomics and peptidomics: Stathmin 2-20 and peptides as sample quality indicators. Proteomics 2007; 7(24):4445-4456.

(277) Galli C, Racagni G. Use of Microwave Techniques to Inactivate Brain-Enzymes Rapidly. Methods in Enzymology 1982; 86:635-642.

(278) O'Callaghan JP, Sriram K. Focused microwave irradiation of the brain preserves in vivo protein phosphorylation: comparison with other methods of sacrifice and analysis of multiple phosphoproteins. Journal of Neuroscience Methods 2004; 135(1-2):159-168.

(279) Theodorsson E, Stenfors C, Mathe AA. Microwave Irradiation Increases Recovery of Neuropeptides from Brain-Tissues. Peptides 1990; 11(6):1191-1197.

(280) Mathe AA, Stenfors C, Brodin E, Theodorsson E. Neuropeptides in Brain - Effects of Microwave Irradiation and Decapitation. Life Sciences 1990; 46(4):287-293.

(281) Nylander I, Stenfors C, Tanno K, Mathe AA, Terenius L. A comparison between microwave irradiation and decapitation: basal levels of dynorphin and enkephalin and the effect of chronic morphine treatment on dynorphin peptides. Neuropeptides 1997; 31(4):357-365.

(282) Svensson M, Skold K, Svenningsson P, Andren PE. Peptidomics-based discovery of novel neuropeptides. Journal of Proteome Research 2003; 2(2):213-219.

(283) Che FY, Lim J, Pan H, Biswas R, Fricker LD. Quantitative neuropeptidomics of microwave-irradiated mouse brain and pituitary. Molecular & Cellular Proteomics 2005; 4(9):1391-1405.

(284) Mikesh LM, Ueberheide B, Chi A, Coon JJ, Syka JEP, Shabanowitz J et al. The utility of ETD mass spectrometry in proteomic analysis. Biochimica et Biophysica Acta-Proteins and Proteomics 2006; 1764(12):1811-1822.

(285) Bakhtiar R, Guan ZQ. Electron capture dissociation mass spectrometry in characterization of peptides and proteins. Biotechnology Letters 2006; 28(14):1047-1059.

(286) Goodwin RJA, Lang AM, Allingham H, Boren M, Pitt AR. Stopping the clock on proteomic degradation by heat treatment at the point of tissue excision. Proteomics 2010; 10(9):1751-1761.

(287) Olsen JV, Ong SE, Mann M. Trypsin cleaves exclusively C-terminal to arginine and lysine residues. Molecular & Cellular Proteomics 2004; 3(6):608-614.

(288) Bensadek D, Monigatti F, Steen JAJ, Steen H. Why b, y's? Sodiation-induced tryptic peptide-like fragmentation of non-tryptic peptides. International Journal of Mass Spectrometry 2007; 268(2-3):181-189.

(289) Mikesh LM, Ueberheide B, Chi A, Coon JJ, Syka JEP, Shabanowitz J et al. The utility of ETD mass spectrometry in proteomic analysis. Biochimica et Biophysica Acta-Proteins and Proteomics 2006; 1764(12):1811-1822.

(290) Sun RX, Dong MQ, Chi H, Yang B, Xiu LY, Wang LH et al. ECD/ETD-based Tandem Mass Spectrometry in Proteomics. Progress in Biochemistry and Biophysics 2010; 37(1):94-102.

(291) Jansson M, Moller M, Hederstedt L, James P. Charge-directed fragmentation of non-tryptic peptides. Molecular & Cellular Proteomics 2005; 4(8):S322.

(292) Elliott WI. Systemic hypertension. Current Problems in Cardiology 2007; 32(4):201-259.

(293) Kearney PM, Whelton M, Reynolds K, Muntner P, Whelton PK, He J. Global burden of hypertension: analysis of worldwide data. Lancet 2005; 365:217-223.

(294) Kannel WB. Blood pressure as a cardiovascular risk factor - Prevention and treatment. Jama-Journal of the American Medical Association 1996; 275(20):1571-1576.

(295) Lee MLT, Rosner BA, Weiss ST. Relationship of blood pressure to cardiovascular death: The effects of pulse pressure in the elderly. Annals of Epidemiology 1999; 9(2):101-107.

(296) Staessen JA, Wang YG, Thijs L. Cardiovascular protection and blood pressure reduction: a meta-analysis. Lancet 2001; 358(9290):1305-1315.

(297) Kannel WB. Risk stratification in hypertension: New insights from the Framingham Study. American Journal of Hypertension 2000; 13(1):3S-10S.

(298) Cheung BMY, Ong KL, Man YB, Lam KSL, Lau CP. Prevalence, awareness, treatment, and control of hypertension: United States National Health and Nutrition Examination Survey 2001-2002. J Clin Hypertens (Greenwich) 2006; 8(2):93-98.

(299) Aram V.Chobanian M, George L.Bakris M, Henry R.Black M, William C.Cushman M, Lee A.Green MM, Joseph L.Izzo JM et al. The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure. Journal of Americal Medical Association. In press.

(300) Oparil S, Zaman A, Calhoun DA. Pathogenesis of hypertension. Annals of Internal Medicine 2003; 139(9):761-776.

(301) De Wardener HE, Macgregor GA. Sodium and blood pressure. Current Opinion in Cardiology 2002; 17(4):360-367.

(302) Mayer G. An update on the relationship between the kidney, salt and hypertension. Wien Med Wochenschr 2008; 158(13-14):365-369.

(303) Meneton P, Jeunemaitre X, De Wardener HE, Macgregor GA. Links between dietary salt intake, renal salt handling, blood pressure, and cardiovascular diseases. Physiological Reviews 2005; 85(2):679-715.

(304) Koga Y, Hirooka Y, Araki S, Nozoe M, Kishi T, Sunagawa K. High Salt Intake Enhances Blood Pressure Increase during Development of Hypertension via Oxidative Stress in Rostral Ventrolateral Medulla of Spontaneously Hypertensive Rats. Hypertension Research 2008; 31(11):2075-2083.

(305) Sanders PW. Dietary Salt Intake, Salt Sensitivity, and Cardiovascular Health. Hypertension 2009; 53(3):442-445.

(306) Lifton RP, Gharavi AG, Geller DS. Molecular mechanisms of human hypertension. Cell 2001; 104(4):545-556.

(307) Blaustein MP, Zhang J, Chen L, Hamilton BP. How does salt retention raise blood pressure? American Journal of Physiology-Regulatory Integrative and Comparative Physiology 2006; 290(3):R514-R523.

(308) Lohmeier TE, Hildebrandt DA, Warren S, May PJ, Cunningham JT. Recent insights into the interactions between the baroreflex and the kidneys in hypertension. American Journal of Physiology-Regulatory Integrative and Comparative Physiology 2005; 288(4):R828-R836.

(309) Guyton AC, Coleman TG. Quantitative Analysis of Pathophysiology of Hypertension. Circulation Research 1969; 24(5S1):I1-&.

(310) Hamlyn JM, Blaustein MP. Sodium-Chloride, Extracellular Fluid Volume, and Blood-Pressure Regulation. American Journal of Physiology 1986; 251(4):F563-F575.

(311) Freis ED. Mechanism of the Antihypertensive Effects of Diuretics Possible Role of Salt in Hypertension. Clinical Pharmacology & Therapeutics 1960; 1(3):337-344.

(312) Curtis JJ, Luke RG, Dustan HP, Kashgarian M, Whelchel JD, Jones P et al. Remission of Essential-Hypertension After Renal-Transplantation. New England Journal of Medicine 1983; 309(17):1009-1015.

(313) Guidi E, Menghetti D, Milani S, Montagnino G, Palazzi P, Bianchi G. Hypertension may be transplanted with the kidney in humans: A long-term historical prospective follow-up of recipients grafted with kidneys coming from donors with or without hypertension in their families. Journal of the American Society of Nephrology 1996; 7(8):1131-1138.

(314) Rettig R, Grisk O. The kidney as a determinant of genetic hypertension - Evidence from renal transplantation studies. Hypertension 2005; 46(3):463-468.

(315) Nakasaki Y, Iwamoto T, Hanada H, Imagawa T, Shigekawa M. Cloning of the Rat Aortic Smooth-Muscle Na+/Ca-2+ Exchanger and Tissue-Specific Expression of Isoforms. Journal of Biochemistry 1993; 114(4):528-534.

(316) O'Shaughnessy KM, Karet FE. Salt handling and hypertension. Journal of Clinical Investigation 2004; 113(8):1075-1081.

(317) Takahashi N, Smithies O. Gene targeting approaches to analyzing hypertension. Journal of the American Society of Nephrology 1999; 10(7):1598-1605.

(318) Feinleib M, Garrison RJ, Fabsitz R, Christian JC, Hrubec Z, Borhani NO et al. Nhlbi Twin Study of Cardiovascular-Disease Risk-Factors - Methodology and Summary of Results. American Journal of Epidemiology 1977; 106(4):284-295.

(319) Longini IM, Higgins MW, Hinton PC, Moll PP, Keller JB. Environmental and Genetic Sources of Familial Aggregation of Blood-Pressure in Tecumseh, Michigan. American Journal of Epidemiology 1984; 120(1):131-144.

(320) Johnson RJ, Rodriguez-Iturbe B, Nakagawa T, Kang DH, Feig DI, Herrera-Acosta J. Subtle renal injury is likely a common mechanism for salt-sensitive essential hypertension. Hypertension 2005; 45(3):326-330.

(321) Sutherla DJ, Ruse JL, Laidlaw JC. Hypertension Increased Aldosterone Secretion and Low Plasma Renin Activity Relieved by Dexamethasone. Canadian Medical Association Journal 1966; 95(22):1109-&.

(322) Pascoe L, Curnow KM, Slutsker L, Rosler A, White PC. Mutations in the Human Cyp11B2 (Aldosterone Synthase) Gene Causing Corticosterone Methyloxidase-Ii Deficiency. Proceedings of the National Academy of Sciences of the United States of America 1992; 89(11):4996-5000.

(323) New MI, Levine LS, Biglieri EG, Pareira J, Ulick S. Evidence for An Unidentified Steroid in A Child with Apparent Mineralocorticoid Hypertension. Journal of Clinical Endocrinology and Metabolism 1977; 44(5):924-933.

(324) Mune T, Rogerson FM, Nikkila H, Agarwal AK, White PC. Human Hypertension Caused by Mutations in the Kidney Isozyme of 11-Beta-Hydroxysteroid Dehydrogenase. Nature Genetics 1995; 10(4):394-399.

(325) Geller DS, Farhi A, Pinkerton N, Fradley M, Moritz M, Spitzer A et al. Activating mineralocorticoid receptor mutation in hypertension exacerbated by pregnancy. Science 2000; 289(5476):119-123.

(326) Sesso HD, Buring JE, Rifai N, Blake GJ, Gaziano JM, Ridker PM. C-reactive protein and the risk of developing hypertension. Jama-Journal of the American Medical Association 2003; 290(22):2945-2951.

(327) Vasan RS, Evans JC, Larson MG, Wilson PWF, Meigs JB, Rifai N et al. Serum aldosterone and the incidence of hypertension in nonhypertensive persons. New England Journal of Medicine 2004; 351(1):33-41.

(328) Wang TJ, Evans JC, Meigs JB, Rifai N, Fox CS, D'Agostino RB et al. Low-grade albuminuria and the risks of hypertension and blood pressure progression - Response. Circulation 2005; 112(9):E121.

(329) Freitag MH, Larson MG, Levy D, Benjamin EJ, Wang TJ, Leip EP et al. Plasma brain natriuretic peptide levels and blood pressure tracking in the Framingham Heart Study. Hypertension 2003; 41(4):978-983.

(330) Wang TJ, Gona P, Larson MG, Levy D, Benjamin EJ, Tofler GH et al. Multiple biomarkers and the risk of incident hypertension. Hypertension 2007; 49(3):432-438.

(331) Carty DM, Siwy J, Brennand JE, Zurbig P, Mullen W, Franke J et al. Urinary Proteomics for Prediction of Preeclampsia. Hypertension 2011; 57(3):561-U387.

(332) Dayarathna MKDR, Hancock WS, Hincapie M. A two step fractionation approach for plasma proteomics using immunodepletion of abundant proteins and multi-lectin affinity chromatography: Application to the analysis of obesity, diabetes, and hypertension diseases. Journal of Separation Science 2008; 31(6-7):1156-1166.

(333) Graham D, McBride MW, Gaasenbeek M, Gilday K, Beattie E, Miller WH et al. Candidate genes that determine response to salt in the stroke-prone spontaneously hypertensive rat - Congenic analysis. Hypertension 2007; 50:1134-1141.

(334) Clark JS, Jeffs B, Davidson AO, Lee WK, Anderson NH, Bihoreau MT et al. Quantitative trait loci in genetically hypertensive rats - Possible sex specificity. Hypertension 1996; 28(5):898-906.

(335) Kerr S, Brosnan MJ, McIntyre M, Reid JL, Dominiczak AF, Hamilton CA. Superoxide anion production is increased in a model of genetic hypertension - Role of the endothelium. Hypertension 1999; 33(6):1353-1358.

(336) Jeffs B, Negrin CD, Graham D, Clark JS, Anderson NH, Gauguier D et al. Applicability of a "speed" congenic strategy to dissect blood pressure quantitative trait loci on rat chromosome 2. Hypertension 2000; 35(1):179-187.

(337) McBride MW, Carr FJ, Graham D, Anderson NH, Clark JS, Lee WK et al. Microarray analysis of rat chromosome 2 congenic strains. Hypertension 2003; 41(3):847-853.

(338) McBride MW, Brosnan MJ, Mathers J, McLellan LI, Miller WH, Graham D et al. Reduction of Gstm1 expression in the stroke-prone spontaneously hypertension rat contributes to increased oxidative stress. Hypertension 2005; 45(4):786-792.

(339) Lander ES, Schork NJ. Genetic Dissection of Complex Traits. Science 1994; 265(5181):2037-2048.

(340) Ransohoff DF. Opinion - Bias as a threat to the validity of cancer molecular-marker research. Nature Reviews Cancer 2005; 5(2):142-149.

(341) Ransohoff DF. Opinion - Rules of evidence for cancer molecular-marker discovery and validation. Nature Reviews Cancer 2004; 4(4):309-314.

(342) Pepe MS, Etzioni R, Feng ZD, Potter JD, Thompson ML, Thornquist M et al. Phases of biomarker development for early detection of cancer. Journal of the National Cancer Institute 2001; 93(14):1054-1061.

(343) Anderson NL, Anderson NG. The human plasma proteome - History, character, and diagnostic prospects. Molecular & Cellular Proteomics 2002; 1(11):845-867.

(344) Tam SW, Pirro J, Hinerfeld D. Depletion and fractionation technologies in plasma proteomic analysis. Expert Review of Proteomics 2004; 1(4):411-420.

(345) Veenstra TD, Conrads TP, Hood BL, Avellino AM, Ellenbogen RG, Morrison RS. Biomarkers: Mining the biofluid proteome. Molecular & Cellular Proteomics 2005; 4(4):409-418.

(346) Echan LA, Tang HY, Ali-Khan N, Lee K, Speicher DW. Depletion of multiple high-abundance proteins improves protein profiling capacities of human serum and plasma. Proteomics 2005; 5(13):3292-3303.

(347) Ahmed N, Barker G, Oliva K, Garfin D, Talmadge K, Georgiou H et al. An approach to remove albumin for the proteomic analysis of low abundance biomarkers in human serum. Proteomics 2003; 3(10):1980-1987.

(348) Steel LF, Trotter MG, Nakajima PB, Mattu TS, Gonye G, Block T. Efficient and specific removal of albumin from human serum samples. Molecular & Cellular Proteomics 2003; 2(4):262-270.

(349) Adkins JN, Varnum SM, Auberry KJ, Moore RJ, Angell NH, Smith RD et al. Toward a human blood serum proteome - Analysis by multidimensional separation coupled with mass spectrometry. Molecular & Cellular Proteomics 2002; 1(12):947-955.

(350) Bjorhall K, Miliotis T, Davidsson P. Comparison of different depletion strategies for improved resolution in proteomic analysis of human serum samples. Proteomics 2005; 5(1):307-317.

(351) Whiteaker JR, Zhang HD, Eng JK, Fang RH, Piening BD, Feng LC et al. Head-to-head comparison of serum fractionation techniques. Journal of Proteome Research 2007; 6:828-836.

(352) Castronovo V, Kischel P, Guillonneau F, de Leval L, Defechereux T, De Pauw E et al. Identification of specific reachable molecular targets in human breast cancer using a versatile ex vivo proteomic method. Proteomics 2007; 7(8):1188-1196.

(353) Gronwall C, Sjoberg A, Ramstrom M, Hoidn-Guthenberg I, Hober S, Jonasson P et al. Affibody-mediated transferrin depletion for proteomics applications. Biotechnology Journal 2007; 2(11):1389-1398.

(354) Zolotarjova N, Martosella J, Nicol G, Bailey J, Boyes BE, Barrett WC. Differences among techniques for high-abundant protein depletion. Proteomics 2005; 5(13):3304-3313.

(355) Brand J, Haslberger T, Zolg W, Pestlin G, Palme S. Depletion efficiency and recovery of trace markers from a multiparameter immunodepletion column. Proteomics 2006; 6(11):3236-3242.

(356) Gong Y, Li X, Yang B, Ying WT, Li D, Zhang YJ et al. Different immunoaffinity fractionation strategies to characterize the human plasma proteome. Journal of Proteome Research 2006; 5(6):1379-1387.

(357) Pernemalm M, Orre LM, Lengqvist J, Wikstrom P, Lewensohn R, Lehtio J. Evaluation of three principally different intact protein prefractionation methods for plasma biomarker discovery. Journal of Proteome Research 2008; 7(7):2712-2722.

(358) Liu T, Qian WJ, Mottaz HM, Gritsenko MA, Norbeck AD, Moore RJ et al. Evaluation of multiprotein immunoaffinity subtraction for plasma proteomics and candidate biomarker discovery using mass spectrometry. Molecular & Cellular Proteomics 2006; 5(11):2167-2174.

(359) Huang L, Fang X. Immunoaffinity fractionation of plasma proteins by chicken IgY antibodies. Methods in Molecular Biology 2008;41-51.

(360) Zolotarjova N, Mrozinski P, Chen H, Martosella J. Combination of affinity depletion of abundant proteins and reversed-phase fractionation in proteomic analysis of human plasma/serum. Journal of Chromatography A 2008; 1189(1-2):332-338.

(361) Levin Y, Schwarz E, Wang L, Leweke FM, Bahn S. Label-free LC-MS/MS quantitative proteomics for large-scale biomarker discovery in complex samples. Journal of Separation Science 2007; 30:2198-2203.

(362) Gao MX, Deng CH, Yu WJ, Zhang Y, Yang PY, Zhang XM. Large scale depletion of the high-abundance proteins and analysis of middle- and low-abundance proteins in human liver proteome by multidimensional liquid chromatography. Proteomics 2008; 8(5):939-947.

(363) Sennels L, Salek M, Lomas L, Boschetti E, Righetti PG, Rappsilber J. Proteomic analysis of human blood serum using peptide library beads. Journal of Proteome Research 2007; 6:4055-4062.

(364) Jonasson T, Ohlin AK, Gottsater A, Hultberg B, Ohlin H. Plasma homocysteine and markers for oxidative stress and inflammation in patients with coronary artery disease - a prospective randomized study of vitamin supplementation. Clinical Chemistry and Laboratory Medicine 2005; 43(6):628-634.

(365) Hussain SP, Hofseth LJ, Harris CC. Radical causes of cancer. Nature Reviews Cancer 2003; 3(4):276-285.

(366) Shi QL, Gibson GE. Oxidative stress and transcriptional regulation in Alzheimer disease. Alzheimer Disease & Associated Disorders 2007; 21(4):276-291.

(367) Dalle-Donne I, Giustarini D, Colombo R, Rossi R, Milzani A. Protein carbonylation in human diseases. Trends in Molecular Medicine 2003; 9(4):169-176.

(368) Mirzaei H, Baena B, Barbas C, Regnier F. Identification of oxidized proteins in rat plasma using avidin chromatography and tandem mass spectrometry. Proteomics 2008; 8(7):1516-1527.

(369) Liu T, Qian WJ, Strittmatter EF, Camp DG, Anderson GA, Thrall BD et al. High-throughput comparative proteome analysis using a quantitative cysteinyl-peptide enrichment technology. Analytical Chemistry 2004; 76(18):5345-5353.

(370) Qiu RQ, Regnier FE. Comparative glycoproteomics of N-linked complex-type glycoforms containing sialic acid in human serum. Analytical Chemistry 2005; 77(22):7225-7231.

(371) Kobata A, Amano J. Altered glycosylation of proteins produced by malignant cells, and application for the diagnosis and immunotherapyof tumours. Immunology and Cell Biology 2005; 83(4):429-439.

(372) Zhao J, Simeone DM, Heidt D, Anderson MA, Lubman DM. Comparative serum glycoproteomics using lectin selected sialic acid glycoproteins with mass spectrometric analysis: Application to pancreatic cancer serum. Journal of Proteome Research 2006; 5(7):1792-1802.

(373) Zhang H, Li XJ, Martin DB, Aebersold R. Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. Nature Biotechnology 2003; 21(6):660-666.

(374) Larsen MR, Jensen SS, Jakobsen LA, Heegaard NHH. Exploring the sialiome using titanium dioxide chromatography and mass spectrometry. Molecular & Cellular Proteomics 2007; 6:1778-1787.

(375) Zhang QB, Tang N, Schepmoes AA, Phillips LS, Smith RD, Metz TO. Proteomic profiling of nonenzymatically glycated proteins in human plasma and erythrocyte membranes. Journal of Proteome Research 2008; 7(5):2025-2032.

(376) Sun XS, Chiu JF, He QY. Application of immobilized metal affinity chromatography in proteomics. Expert Review of Proteomics 2005; 2(5):649-657.

(377) Cheng AJ, Chen LC, Chien KY, Chen YJ, Chang JTC, Wang HM et al. Oral cancer plasma tumor marker identified with bead-based affinity-fractionated proteomic technology. Clinical Chemistry 2005; 51(12):2236-2244.

(378) Navare A, Zhou M, McDonald J, Noriega FG, Sullards MC, Fernandez FM. Serum biomarker profiling by solid-phase extraction with particle-embedded micro tips and matrix-assisted laser desorption/ionization mass spectrometry. Rapid Communications in Mass Spectrometry 2008; 22(7):997-1008.

(379) Gatlin-Bunai CL, Cazares LH, Cooke WE, Semmes OJ, Malyarenko DI. Optimization of MALDI-TOF MS detection for enhanced sensitivity of affinity-captured proteins spanning a 100 kDa mass range. Journal of Proteome Research 2007; 6:4517-4524.

(380) Granger J, Siddiqui J, Copeland S, Remick D. Albumin depletion of human plasma also removes low abundance proteins including the cytokines. Proteomics 2005; 5(18):4713-4718.

(381) Tukey JW. Some Thoughts on Clinical-Trials, Especially Problems of Multiplicity. Science 1977; 198(4318):679-684.

(382) Greenhalgh T. How to read a paper - Statistics for the non-statistician .2. "significant" relations and their pitfalls. British Medical Journal 1997; 315(7105):422-425.

(383) Bland JM, Altman DG. Multiple Significance Tests - the Bonferroni Method .10. British Medical Journal 1995; 310(6973):170.

(384) Perneger TV. What's wrong with Bonferroni adjustments. BMJ 1998; 316(7139):1236-1238.

(385) Rothman KJ. No adjustments are needed for multiple comparisons. Epidemiology 1990; 1(1):43-46.

(386) Pawitan Y, Michiels S, Koscielny S, Gusnanto A, Ploner A. False discovery rate, sensitivity and sample size for microarray studies. Bioinformatics 2005; 21(13):3017-3024.

(387) Benjamini Y, Hochberg Y. Controlling the False Discovery Rate - A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B-Methodological 1995; 57(1):289-300.

(388) Benjamini Y, Hochberg Y. On the adaptive control of the false discovery fate in multiple testing with independent statistics. Journal of Educational and Behavioral Statistics 2000; 25(1):60-83.

(389) Keselman HJ, Cribbie R, Holland B. Controlling the rate of Type I error over a large set of statistical tests. British Journal of Mathematical & Statistical Psychology 2002; 55:27-39.

(390) Grant GR, Liu JM, Stoeckert CJ. A practical false discovery rate approach to identifying patterns of differential expression in microarray data. Bioinformatics 2005; 21(11):2684-2690.

(391) Hu JH, Zou F, Wright FA. Practical FDR-based sample size calculations in microarray experiments. Bioinformatics 2005; 21(15):3264-3272.

(392) Jung SH. Sample size for FDR-control in microarray data analysis. Bioinformatics 2005; 21(14):3097-3104.

(393) Jung SH, Jang W. How accurately can we control the FDR in analyzing microarray data? Bioinformatics 2006; 22(14):1730-1736.

(394) Li QB, Roxas BAP. An assessment of false discovery rates and statistical significance in label-free quantitative proteomics with combined filters. Bmc Bioinformatics 2009; 10.

(395) Maurer MH, Feldmann RE, Bromme JO, Kalenka A. Comparison of statistical approaches for the analysis of proteome expression data of differentiating neural stem cells. Journal of Proteome Research 2005; 4(1):96-100.

(396) Roxas BAP, Li QB. Significance analysis of microarray for relative quantitation of LC/MS data in proteomics. Bmc Bioinformatics 2008; 9.

(397) Choi H, Fermin D, Nesvizhskii AI. Significance Analysis of Spectral Count Data in Label-free Shotgun Proteomics. Molecular & Cellular Proteomics 2008; 7(12):2373-2385.

(398) Breitling R, Herzyk P. Rank-based methods as a non-parametric alternative of the T-statistic for the analysis of biological microarray data. Journal of Bioinformatics and Computational Biology 2005; 3(5):1171-1189.

(399) Breitling R, Armengauda P, Amtmann A, partial-least square discriminant analysis Herzyk P. Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments. FEBS letters 573, 83-92. 11-8-2004.

(400) Nikkila H, Gitlin JD, Mullereberhard U. Rat Hemopexin - Molecular-Cloning, Primary Structural Characterization, and Analysis of Gene-Expression. Biochemistry 1991; 30(3):823-829.

(401) Mullereberhard U. Hemopexin. Methods in Enzymology 1988; 163:536-565.

(402) Lecam A, Pages G, Auberger P, Lecam G, Leopold P, Benarous R et al. Study of A Growth Hormone-Regulated Protein Secreted by Rat Hepatocytes - Cdna Cloning, Anti-Protease Activity and Regulation of Its Synthesis by Various Hormones. Embo Journal 1987; 6(5):1225-1232.

(403) Ohkubo K, Ogata S, Misumi Y, Takami N, Ikehara Y. Molecular-Cloning and Characterization of Rat Contrapsin-Like Protease Inhibitor and Related Proteins. Journal of Biochemistry 1991; 109(2):243-250.

(404) Chao J, Chai KX, Chen LM, Xiong W, Chao S, Woodleymiller C et al. Tissue Kallikrein-Binding Protein Is A Serpin .1. Purification, Characterization, and Distribution in Normotensive and Spontaneously Hypertensive Rats. Journal of Biological Chemistry 1990; 265(27):16394-16401.

(405) Yoon JB, Towle HC, Seelig S. Growth-Hormone Induces 2 Messenger-Rna Species of the Serine Protease Inhibitor Gene Family in Rat-Liver. Journal of Biological Chemistry 1987; 262(9):4284-4289.

(406) Wang JY, Cheng KI, Yu FJ, Tsai HL, Huang TJ, Hsieh JS. Analysis of the correlation of plasma NO and ET-1 levels in rats with acute mesenteric ischemia. Journal of Investigative Surgery 2006; 19(3):155-161.

(407) Huang JTJ, McKenna T, Hughes C, Leweke FM, Schwarz E, Bahn S. CSF biomarker discovery using label-free nano-LC-MS based proteomic profiling: Technical aspects. Journal of Separation Science 2007; 30(2):214-225.