CHAPTER 3: COMPARATIVE ANALYSIS OF A MULTIDIMENSIONAL ION EXCHANGE CHROMATOGRAPHY OF PROTEINS *VERSUS* PEPTIDE ION EXCHANGE CHROMATOGRAPHY OF DEPLETED HUMAN PLASMA

3.1 ABSTRACT

Blood and the identification of biomarkers for the early detection of disease is a prominent area of biomedical interest. The rate of biomarker discovery over the past decade has dropped dramatically with many looking towards the field of proteomics to fill the void for the 21st century. A unique form of protein fractionation termed PROOF (protein repetitive orthogonal off-line fractionation) of proteins in native states was tested on human plasma samples and compared against the standard fractionation and analysis on peptides known as the MudPIT (bottom-up) approach for proteome analysis. Both methods identified a similar number of proteins at high stringency of identification confidence, 146 and 147 proteins for PROOF and MudPIT respectively. Of these proteins identified, 74% were identified by both techniques.

The development and implementation of a unique graphical interpretative method to conduct top-down proteomic analysis of the PROOF data was conducted, highlighting the protein inference problem in bottom-up proteomic analysis and the potential loss of, or misinterpretation of data from such bottom-up analysis alone. I was able to interpolate possible structural information from the PROOF analysis that could not be found from standard bottom-up proteomics. Of the 146 proteins graphically analysed from the PROOF analysis, 5 proteins (IPI00423461.3, IPI00298828.3, IPI00218192.3, IPI00019591.2 and IPI00303963.1) are presented in detail as part of the initial proof of concept for the method developed.

Protein IPI00218192.3 showed the cleavage of the protein into two units as is stated in the literature, and is a good proof of concept for both the PROOF and graphical interpretation method developed and explained within. Standard bottom-up analysis identified only one protein, whereas I identified the two fragments as two distinct protein species.

Protein IPI00423461.3 showed the possible N-terminal cleavage of two tryptic peptides, creating two entities, rather than as one protein identification with a larger molecular weight, as was found by the bottom-up approach alone.

Protein IPI00298828.3 showed that is was possibly in two distinct conformational forms, either due to associations with other proteins or itself in a multimeric

arrangement. It is apparent that this protein has at least two tertiary/quaternary structures, which would imply two or more functional states for this protein.

Proteins IPI00019591.2 and IPI00303963.1 show a relationship that highlights the use of the unique and non-unique peptide classification established herein, leading to the interpretation that IPI00019591.2 is cleaved at the 23-24 peptide tryptic fragment and is possibly the smaller variant, while this N-terminal piece is not a separate entity cleaved from the parent protein, but that the peptides, being non-unique, are from the protein IPI00303963.1.

These five proteins highlighted the ideology behind the technique demonstrated, lending weight to the argument that top-down proteomics can enhance the information value of traditional bottom-up proteomics, and that the development of top-down proteomics should be a vital part of proteomics and biomedical research both now and in the future.

3.2 PREAMBLE

This project was initiated due to the limitations recognised with the SAM Biochip MALDI project described in Chapter 2; mainly that it was unlikely to be anything more than a ‘concentration chip’ in its current incarnation. To date, only one publication has been published on the technology in peer reviewed literature [60]. I concluded that it was beyond the scope, timescale and funding of my project to implement any of the envisaged changes I believed could enhance the ‘concentration’ chip, and associated ‘sample clean-up’ or ‘affinity capture’ chips. Hence, I decided to investigate chromatographic fractionation of human plasma to highlight the potential pitfalls of current bottom-up proteomics methods, namely the misidentification of some proteins in a sample set. A secondary aim was developing a procedure that could be used prior to SAM Biochip analysis aimed at biomarker discovery in human plasma. This plasma protein fractionation work was run in parallel with the later stages of the SAM Biochip MALDI project in Chapter 2.

My work presented in this chapter is neither a traditional top-down or bottom-up proteomic approach. It is more of a hybrid bottom-up, generating some information as it pertains to protein isoforms or species, with aspirations to be used in tandem with traditional top-down MS analysis for the future. This distinction is made based on the premise that I fractionated the proteome based on the proteins (top-down) with LC, and carried out mass spectrometric analysis on the tryptic peptides, which I classify as a bottom-up experiment. However, the unique graphical representation of the information I developed can relate some of the information generated back to the species or isoforms of the proteins present, hence making it a top-down approach. The technique developed could be adapted to run in a traditional top-down configuration provided the appropriate mass spectrometry platform was available for MS of intact mass followed by MS/MS of these species.

It has long been known that a fundamental issue within the field of proteomics is that the majority of the sample preparation and mass spectrometric analysis is conducted on the peptides and then the identification of the proteins are inferred from this peptide data, known as the protein inference problem with bottom-up proteomic analysis [185, 377]. This protein inference problem leads to questions being asked as

to the validity and usefulness of the majority of data generated within the field of proteomics for which numerous methods of interpretation have and are being developed [185, 192, 378-380]. The protein inference problem relates to the assignment of a weighted probability for the identified peptides and the distinction between the unique and non-unique peptides in a dataset, a problem that both gel electrophoresis and multi-dimensional liquid chromatography face to varying degrees when the identification of proteins is based solely on MS/MS of peptide information [185]. It can be argued that the protein inference problem has surreptitiously played a critical role in the inability of proteomics to generate any stand alone tangible medical diagnostics since its inception [184, 298, 381].

The chromatographic separation project I worked on was an offshoot of a larger project involving Dr Amit Kapur (Illawarra Health and Medical Research Institute), Dr Brian Hood (GE Healthcare) and Dr Hans-Rudolf Höpker (GE Healthcare). They were working on fractionating proteins from chicken egg yolk. The common theme of their work and mine was that we wanted to use liquid chromatography on the proteins as opposed to 2D-GE on the proteins or LC on the peptides, and we also wanted to try to maintain the proteins' biological activity throughout the fractionation process. This includes the endogenous conformation, and the identity of truncated and cleaved forms of the proteins that occur under normal *in vivo* conditions. By working with the proteins under mild salt conditions it was hypothesised that the structure and function of the proteins would be retained. Hence, we were trying to differentiate protein species with liquid chromatography and mass spectrometric analysis, analogous to top-down proteomics. This should mean that some of the endogenous conformational information normally lost in 2D-Gels and traditional liquid chromatography of peptides, for example any inter- or intra-molecular complexation, tertiary or quaternary structural elements, could be maintained during this protein liquid chromatography.

We refer to the basis of this fractionation concept as PROOF (protein repetitive orthogonal off-line fractionation) of proteins in native states, a precursor of the work later described as the Tandem IEX approach by Dr Hans-Rudolf Höpker *et al.* [311]. The PROOF technique is a compromise between the necessary conditions of the particular liquid chromatography employed, while maintaining similar conditions to the endogenous (*in vivo*) environment the proteins normally operate in. We employed

two ion exchange columns in tandem enabling binding and elution from the columns, and decomplexation (fractionation) of the sample, while maintaining the structural integrity of the proteins within.

For many of our experiments the PROOF fractions were collected, digested and further fractionated for mass spectrometric analysis to be conducted on the peptides. Alternatively, orthogonal off-line experiments could be conducted on the PROOF fractions to test for particular biological activity or protein-protein interaction studies [280]. Additionally, with the implementation of a different mass spectrometry platform, the use of top-down, intact-mass, mass spectrometry could also be employed and run on these types of protein fractions, either stand alone or in parallel with the LC-MS/MS of the peptides [215].

I chose to work with depleted human plasma due to its difficulty, dynamic range and biomedical significance [275, 294]. Hence, I implemented a stringent testing of the fractionation concept, automated the multidimensional protein fractionation with replicates, and up-scaled the analysis amount. I also compared it directly against a standard peptide-based MudPIT analysis, with SCX in the first dimension followed by RP LC-MS/MS in the second [146, 382-384]. Since I was not able to implement intact mass top-down MS and MS/MS proteomics analysis on the fractions due to instrumentation and time limitations, I tried to capitalise on the unique nature of the information that I believed to potentially exist within the data collected by creating an innovative way of viewing the peptide data and relating it back to the protein fractions so as to highlight the additional results achievable with the PROOF methodology.

For the protein fractionation of human plasma after abundant protein depletion, I used two High Performance columns, a HiLoad 26/10 SP- (sulphopropyl cation-exchanger) and Q-Sepharose (quaternary ammonium anion-exchanger). A physiological pH of 7.4 was maintained throughout the linear gradient applied by changing the ionic strength of the mobile phase. Due to the polyampholytic nature of proteins and the pIs of the known plasma proteins, we expected a greater binding and fractionation from the strong cation-exchanger column (SCX) than the strong anion-exchanger column (SAX) under the configuration we employed [311, 385].

It is important to note that since we were working with depleted plasma there is a possibility that the depletion process has interfered with the normal *in vivo* tertiary or quaternary structures of the proteins at some level, either due to the removal of the abundant proteins in the fluid or the chemical conditions under which the depletion chromatography was run.

With respect to the identification of truncated protein species into fragments or subunits, this process is facilitated by the novel graphical data representation developed and presented within this chapter. Thus, gene polymorphisms (DNA), alternative splicing (pre- and mRNA) and PTM cleavage events could be inferred. At the very least, I can identify from these truncated protein forms potential candidates for further investigation as bioactive proteins/peptides. These cleaved protein variants may possibly be an indicator of biomarkers from standard secretory mechanisms that change the shape or function of the protein, but are not currently distinguishable utilising the standard bottom-up identification methods of proteins based on peptides alone.

It is helpful to think of this study as a four-dimensional protein profiling of human plasma. I took human plasma and chromatographically depleted it to remove the high abundance species, followed by a novel two dimensional ion exchange chromatography, with a strong cation exchange and a strong anion exchanger (2D-SCX/SAX) on the proteins under mild salt conditions at physiological pH. These discrete protein fractions were then digested and the resulting peptides were fractionated by nanoflow RP-LC/MS/MS analysis. I then applied stringent selection criteria for identification based on the calculation of false discovery rates (FDR) and incorporating a triplicate exclusion criteria for any identifications not found in all three replicates. I then developed a schematic representation of the data to assist in relating the peptide information back to the protein fractions in order to identify truncated or cleaved elements that would be present endogenously within the sample and not identifiable by bottom-up proteomics.

3.3 MATERIALS AND METHODS

3.3.1 Materials and Reagents

All chemicals and reagents were purchased from Sigma-Aldrich (St. Louis, Missouri, USA) unless otherwise stated.

3.3.2 Preparation of Depleted Human Plasma Samples

The Human Plasma used in this study was supplied by the Australian Red Cross Blood Service and was from healthy male and female donors of type B positive blood. The plasma was removed from the -80°C freezer and thawed overnight in the 4°C fridge and then left to rest at room temperature (RT) for 2 hours prior to manipulation. The plasma was centrifuged at 10,000 rpm for 20 min in 2 x 15 ml Falcon Tubes (FT). The supernatant was removed from the FT and diluted with Solubilising Buffer A (10 mM Tris HCl, 150 mM NaCl, 0.02% Azide at pH 7.4) in a ratio of 1:5 and filtered with a 0.02 µm filter (Millipore, USA). The flow through was collected and 5 x 10 ml aliquots collected for application to the depletion column, any unused flow through was placed in the freezer at -70°C. The MIXED12-LC20 depletion column (GenWay, San Diego, CA, USA) was used to deplete the 12 highest abundant proteins in the plasma according to the manufacture's specifications. Figure 3.1 depicts the elution gradient generated for the collection of the depleted plasma.

3.3.3 Concentration and Buffer Exchange of Depleted Plasma

The flow through from the GenWay depletion column was passed through 8 x 5 kDa cut-off Amicon spin tubes (Millipore, USA) and centrifuged at 5,000 rpm for 20 min. Once the volume in each tube reached approximately 5 ml, the concentrated mixture was buffer exchanged with 3 x 10 ml of 10 mM Tris, 10 mM di-sodium orthophosphate, pH 7.4 in each tube. The remaining supernatant was pooled and aliquoted into 1.5 ml tubes and placed in the freezer at -20°C.

3.3.4 One Dimensional Gel Electrophoresis of Depleted and non-Depleted Plasma

The depleted and non-depleted plasma were fractionated on an SDS-PAGE Gel (1D Gel) (Invitrogen). Bands were visualized with Coomassie G250 stain and scanned on a flat bed colour scanner. Figure 3.2 depicts the visualisation of the plasma samples.

3.3.5 Quantification of Depleted Human Plasma

Deep Purple (DP) protein quantification kit (Sigma-Aldrich, St. Louis, Missouri, USA) was used as per the manufacturers instructions to estimate the amount of protein present in the depleted plasma samples. A frozen depleted protein aliquot was diluted with ultra pure water (18.2 MΩ·cm) bringing the concentration of the Tris to approximately 30 µM. (Dilution of the sample was necessary because >500 mM of Tris will interfere with the quantitation). The final concentration of each 1.4 ml aliquot was calculated to be approximately 2,500 µg/ml.

3.3.6 Novel Two Dimensional Ion Exchange Chromatography of Proteins from Depleted Human Plasma Samples and Pooling Digestion

An ÄKTÄ Purifier system was configured to conduct the tandem ion exchange chromatography, employing two HiLoad 26/10 SP- and Q-Sepharose High Performance columns (Amersham Biosciences-GE Life Sciences, Sweden) for both the cationic and anionic ion exchange, respectively. A schematic of the chromatographic plumbing and workflow is illustrated in Figure 3.3, as are the chromatographs with the gradient profiles in Figures 3.4 and 3.5. The system was optimised for a continuous flow rate of 0.75 ml/min and was equilibrated before each run with eight column washes of Buffer A (10 mM Tris, 10 mM di-sodium orthophosphate, pH 7.4). An automated Super-Loop (Amersham Biosciences-GE Life Sciences, Sweden) was used to load 2 ml of diluted plasma (1:1 with Buffer A, protein concentration 1,250 µg/ml) for each of the replicates. The loaded sample passed through the SP-Sepharose column first to bind the cationic species, and any

unbound anionic species continued flowing to the secondary Q-Sepharose column. Proteins that did not bind to either column were flushed through with Buffer A and collected as flow through and analysed as Fraction 1. The first elution gradient was applied to the primary column of SP-Sepharose alone, while the secondary column of Q-Sepharose was kept loaded with sample and off-line. A linear gradient from flow point 14 – 42 ml (~37 min) was applied to the SP-Sepharose column from 0 to 50% Buffer B, (10 mM Tris, 10 mM di-sodium orthophosphate, 1 M sodium chloride, pH 7.4). The system switched to 100% Buffer B at 42 ml for 3 ml and then switched back to 100% Buffer A at 45 ml flow point. The valves were then switched so that a second elution gradient could be applied to the secondary column of Q-Sepharose, while the primary column of SP-Sepharose was kept wet and off-line. The gradient conditions for the secondary column were identical to the primary column. The totality of fractions collected were pooled into 12 fractions based on a manual inspection of the UV trace as shown in Figure 3.5, with fraction 1 the initial flow through, fraction 12 solely the Q-Sepharose eluent, and fractions 2 through to 11 were for the SP-Sepharose eluent. The 12 fractions were concentrated under vacuum and digested with Trypsin.

3.3.7 Digestion of both the Depleted Human Plasma Samples after the Novel Two Dimensional Ion Exchange Chromatography of Proteins and the Depleted Human Plasma Samples before standard Ion Exchange Chromatography

The method used was the same for both samples, but was applied at different points in the sample preparation. For the novel two-dimensional ion exchange chromatography of proteins, it was after the protein chromatography. For the Std-SCX of peptides, it was before the peptide chromatography. The digestion was undertaken by resuspending the proteins in 20 µl of Urea (6 M), followed by the addition of 2 µl of DTT (200 mM) and incubation at RT in the dark for 1 hour, followed by the addition of 4 µl of iodoacetamide (200 mM) and incubation at RT for 1 hour. Then, 900 µl of Ammonium Bicarbonate (50 mM, pH 8.0) was added followed by 35 µl of trypsin (0.1 µg/ml) and digestion proceeded overnight at 37°C. The digestion reaction was quenched the next day with 1% formic acid and stored in the -20°C freezer until required for nano-RP-LC-MS/MS.

3.3.8 Standard Ion Exchange Chromatography of Peptides from Depleted Human Plasma Samples

SCX chromatography was employed using an Agilent 1100 quaternary HPLC pump (Agilent Technologies, USA) with a polysulfoethyl aspartamide column (PolyLC) (200 mm x 2.1 mm, 5 µm, 200 Å). The column was equilibrated with Buffer A, 5 mM sodium phosphate with 25% (v/v) acetonitrile (ACN) at pH 2.7, which was also used for the sample re-suspension, sample injection and peptide adsorption to the column. A total of 21 injections were conducted, each with approximately 350 µg per injection, which when combined gives triplicate analysis of 2,500 µg per replicate. Peptide elution was achieved with a 70 min linear gradient to 100% Buffer B (5 mM sodium phosphate with 25% (v/v) ACN, 350 mM KCl, pH 2.7) at flow rate 300 µl/min. Peptides were collected into 35 fractions and pooled into 12 fractions of approximately the same amount based on a manual inspection of the elution gradient UV trace as shown in Figure 3.6. Each of the 12 fractions were dried in a vacuum centrifuge and stored at -20°C until required for nano-RP-LC-MS/MS.

3.3.9 In-line Reverse Phase Electrospray Ionisation Mass Spectrometry with a Thermo LTQ-XL

Each of the 12 fractions from each method of fractionation were analysed by nano-RP-LC-MS/MS using an LTQ-XL linear ion trap mass spectrometer with a fused silica capillary with an integrated electrospray tip (Thermo, USA). Each of the peptide extracts were desalted using C18 tips (Omix, Varian, Inc., CA) and the eluate was dried using a vacuum centrifuge followed by resuspension in 0.1% (v/v) formic acid. The methods of nanoflow LC-MS/MS, previously followed by Chick *et al.* [243], were adopted. The reversed-phase columns were packed in-house to approximately 7 cm (100 µm ID) in length using a 100 Å, 5 µm Zorbax C18 resin (Agilent Technologies, USA). An electrospray voltage of 1.8 kV was applied via a liquid junction upstream of the C18 column. Samples were injected onto the column using a Surveyor Autosampler (Thermo, USA), which was followed by an initial wash step with Buffer A, 5% (v/v) ACN, 0.1% (v/v) formic acid for 10 min at 1 µl/min. Peptides were eluted from the column with a 0-100% Buffer B, 95% (v/v) ACN, 0.1% (v/v) formic acid for

90 minutes at 500 nl/min. The column eluate was directed into a nanospray ionisation source of the mass spectrometer. Spectra were acquired in a positive ion mode with scanning over the range 400-1500 amu using the Xcalibur software (Version 2.06, Thermo, USA), and automated peak recognition, dynamic exclusion and MS/MS of the top six most intense precursor ions at 35% normalisation collision energy were performed.

3.3.10 Database Searching, Statistical Analysis and Data Processing for Protein Identification

The LTQ-XL raw output files were converted into mzXML format and searched against the Human IPI database (September, 2011), using the global proteome machine (GPM-XE) software (version 2.1.1) and the X!Tandem algorithm. The 12 fractions of each replicate were processed sequentially with output files generated for each fraction, as well as a merged non-redundant output file for protein identifications with log(e) values < -1. Peptide identification was determined using a 0.4 Da fragment mass error. Carbamidomethylation was considered as a complete modification and partial modifications considered included oxidation of methionine and tryptophan, and phosphorylation of serine, threonine and tyrosine.

Reverse database searching was used for estimating false discovery rates (FDRs). A protein log(e) value of -9 was used as an initial cut-off to create the first refined list of identifications with a protein FDR of ~1%, as shown in Table 3.1. Protein FDR was calculated by (number of reversed protein hits / total number of proteins in the list) x 100. Peptide FDR was calculated by 2 x (total number of peptides identified from reversed protein hits / total number of peptides identified for all proteins in the list) x 100 [386]. A secondary selection criteria was applied, with identifications that were not found in all three replicate analyses excluded, using an in-house software package. Venn diagrams as shown in Figure 3.6 were generated to depict the difference and similarities between the total numbers of proteins identified by each method [387]. Complete protein and peptide identification information is available in the supplemental data 3.1 provided with this thesis.

3.3.11 Novel Graphical Representation of the PROOF Data

Once the final list of the identified proteins with a high degree of confidence were generated, computer scripts were written to semi-automatically interrogate the GPM result files so as to extract the “sequence” files, the “all” files and the “modified” files for each of the proteins identified, from each of the replicates inside the GPM-XE operating system. The graphs were then generated to display each of the fractions collected along the x-axis, from 1 – 12 from left to right respectively. Fraction 1 is flow through, fraction 12 is the Q-Sepharose eluate, and fractions 2-11 are the SP-Sepharose eluate. The y-axis was orientated to show the peptide strings used to identify the particular protein in order of N-terminus to C-terminus from top to bottom, respectively. The y-axis was composed by compiling the information supplied in the “sequence” data (amino acid sequence) and the “all” data (peptide strings) for each protein, which was then positioned in relation to the three replicates and the corresponding “all” data files. Thus, for each peptide identified a spherical dot appears next to the peptide string and in-line with the particular fraction that it was identified in. The size of each dot is directly related to the number of identifications/hits for that peptide string in that fraction, as was shown in the GPM data files.

I created a distinction within the peptides identified and shown on the y-axis between unique peptides in red and non-unique peptides in black. This distinction is based on searching each identified peptide string against the compiled set of 10740 identified peptides from the 146 high stringency protein identifications. It is important to note that what I refer to as unique may not in fact be a unique peptide within the IPI human protein database, but it is a unique peptide within my identified dataset. I chose to use this approach and terminology to avoid confusion when I was referring to, for example, non-unique (within the IPI human database) which were only present once, and hence unique, within our dataset. Using this criterion, I identified 3848 unique and 6892 non-unique peptides within the 10740 peptides identified from the 146 proteins.

3.4 RESULTS AND DISCUSSION

3.4.1 Fractionation – Decomplexation of Human Plasma

The GenWay depletion column performed reproducibly as shown by the UV trace in Figure 3.1. The 1D Gel analysis of the human plasma samples, shown in Figure 3.2 was conducted to verify that the depletion column was working satisfactorily. Upon visual inspection of the 1D Gel it is clear that there is a difference between the depleted and non-depleted samples. I cannot conclude definitively whether the sample is depleted with the complete removal of the top 12 most abundant proteins, or not. I can say that the depleted sample is different from the non-depleted sample and showing signs of depletion that are reproducible across the columns in the 1D Gel despite the concentration or amount loaded changing [388]. Particularly, with respect to the reduced presence of a large band (protein species) around 60 kDa, which is most likely Albumin [290]. There was also increased visualisation of more bands in the depleted sample at the higher and lower molecular weight ranges, which would denote a reduction in the dynamic range of the sample due to the removal of some high abundant species, which enabled visualization of bands that we can not see in the non-depleted sample [314]. I took these results as an indication that the sample was depleted to some degree of abundant proteins, which justified the running of the next stages of the experiments.

PROOF is the term we have used to describe the particular system of multi-dimensional ion exchange chromatography we employed for fractionating depleted plasma proteins. PROOF, to our knowledge at the time of these experiments, is unique in using this configuration as shown in Figures 3.3 (A – D) on a ÄKTÄ HPLC system. The plumbing with two PV908 and one PV907 valves on one ÄKTÄ HPLC system enabled us to achieve automated multidimensional liquid chromatography. An additional feature of these experiments was the idea of performing ion exchange chromatography under mild salt conditions and at physiological pH, so as not to disrupt any of the tertiary or quaternary structural elements of the protein mixture.

The concept of fractionating multiple proteins simultaneously using chromatography has been around for a long time, but it has long been thought of as problematic,

mainly due to issues with differential protein solubility and precipitation [148, 389]. We believe we demonstrate here that in this particular configuration, this is no longer a justification for not undertaking these types of experiments when analysing plasma. The entire process of equilibration, sample loading and binding, with separated elution from each ion exchange chromatographic column, and including analysis of three replicates, was entirely automated. This approach can, of course, be used for multiple injections and can be adapted for various protein mixtures and types of columns.

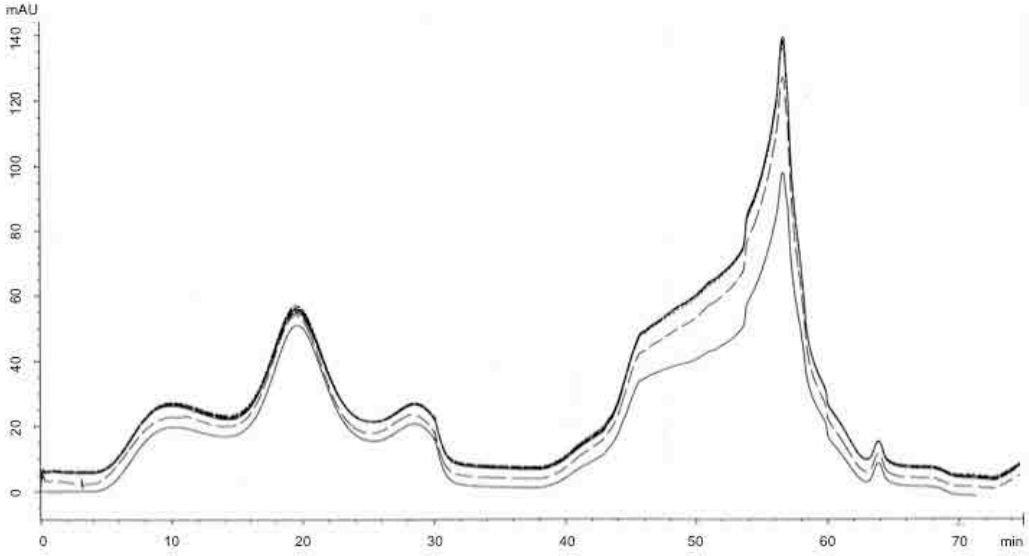


Figure 3.1 UV trace superimposition of multiple runs of the GenWay depletion column for the removal of the top 12 most abundant proteins in human plasma. The peak between 35 min and 70 min is the depleted plasma fragment of the eluent that was collected and used for the later fractionation experiments. The UV trace lines show high reproducibility across runs.

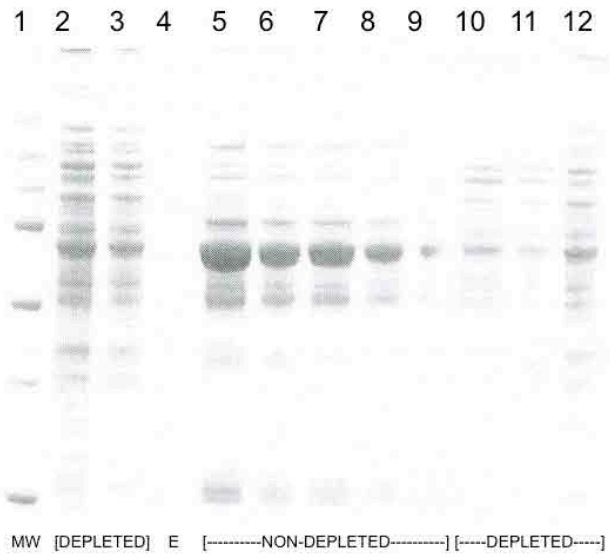


Figure 3.2 1D Gel of the GenWay depleted and non-depleted plasma. Row 1 is the molecular weight markers. The non-depleted plasma are found across rows 5 – 9 of differing concentration and loading amounts. The depleted plasma are found across rows 2 – 3 and 10 - 12 of differing concentrations and loading amounts. Note the improved visualisation of bands across the molecular weight range for the depleted compared to non-depleted. Additionally, the reduced appearance of the band around 60 kDa for depleted compared to non-depleted. Lane 1 = MWM, 2 = Depleted Plasma 10 µl of ½ dilution, 3 = Depleted Plasma 5 µl of ½ dilution, 4 = Empty, 5 = Non-depleted Plasma 10 µl of 1/10 dilution, 6 = Non-depleted Plasma 5 µl of 1/10 dilution, 7 = Non-depleted Plasma 10 µl of 1/50 dilution, 8 = Non-depleted Plasma 5 µl of 1/50 dilution, 9 = Non-depleted Plasma 1 µl of 1/50 dilution,10 = Depleted Plasma 10 µl of 1/10 dilution, 11 = Depleted Plasma 5 µl of 1/10 dilution, Depleted Plasma 5 µl of ½ dilution respectively.

Once the system was optimised, we injected test samples of non-depleted and depleted plasma to assess the enhanced chromatographic resolution achieved when using a depleted sample as opposed to a non-depleted sample in these types of systems. The GenWay depletion technique was shown to change the resultant elution profile of both columns compared to non-depleted plasma. These changes in chromatography can easily be attributed to the removal of the 12 highest abundant species. The UV trace in Figure 3.4 highlights how the depleted sample has a reduced signal for the SAX column at 62 ml, as well the removal of the peak at 15 ml (SCX) and reduced signal for the large peak at 28 ml for the SCX column. We interpret this to be further evidence that the GenWay sample (pink) we are using is depleted of the most abundant proteins when compared to the non-depleted (blue).

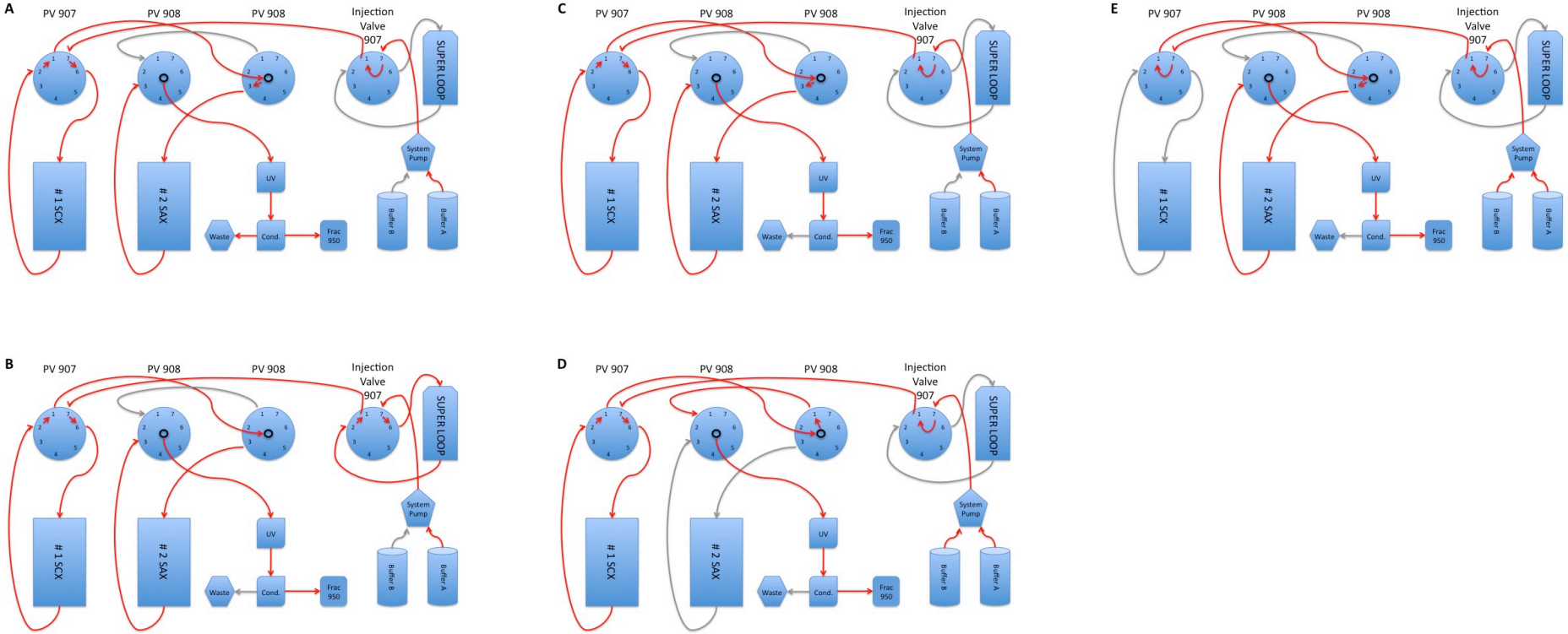


Figure 3.3 Schematic representation of the novel plumbing architecture for the PROOF system developed with a ÄKTÄ HPLC for these experiments. (A) Column equilibration stage with 100% Buffer A, the super-loop is off-line while the rest of the system in-line. (B) Super-Loop is brought in-line and the sample is injected and allowed to pass over both cationic (#1 SCX) and anionic (#2 SAX) columns to bind. (C) Super-Loop is taken off-line and flow rates are maintained, allowing extra flow time for the sample to completely pass through the system for complete binding to both columns. (D) Column #2 (SAX) is bypassed and elution gradient for column #1 (SCX) with introduction of Buffer B is undertaken. (E) The system switches to bypass column #1 (SCX) and bring column #2 (SAX) in-line and the secondary elution gradient with Buffer B is undertaken. Once completed the system switches back to position (A) until the next sample is injected. Grey lines indicated off-line, while red lines indicated on-line.

Upon further examination of the UV trace for depleted plasma, the SAX column does not seem to conduct any discernable type of chromatographic separation, which led us to question whether or not we should include this secondary column in the experiments. However, since we were aiming to compare peptide and protein based multidimensional chromatographic approaches, we decided to keep the system as it was for these proof of concept experiments. The enhanced chromatographic resolution in the UV trace in Figure 3.4 provided justification for using depleted plasma rather than non-depleted, even though some of the tertiary and quaternary structural information may be reduced or lost when using a depleted sample with a GenWay column. Also, protein-protein interaction elements in the chromatographic peaks were removed from the non-depleted plasma samples by the GenWay column, though this information was deemed not pivotal for this development of a proof of the concept. However, this possible removal or change in the protein-protein interactions results in the exclusion of one of the potential benefits of conducting these types of fractionation experiments in mild conditions on proteins in their native state. Thus, no conclusive remarks concerning this area will be made here.

The major reasons for these experiments were to automate, up-scale and validate PROOF against a standard system, based on the highest possible number of stringent protein identifications for a triplicate comparison. Hence, depleted plasma was used to achieve a higher number of protein identifications allowing a better distinction if any between the two techniques at the expense of the protein-protein interaction information at this stage of the technique development. It is important to note that the GenWay depletion step should not have interfered with any endogenous truncation or cleavage fragments of the proteins found in the native states, and hence this aspect of the technique can still be used for enhanced interrogation of the data compared to the Std-SCX of peptides method used within.

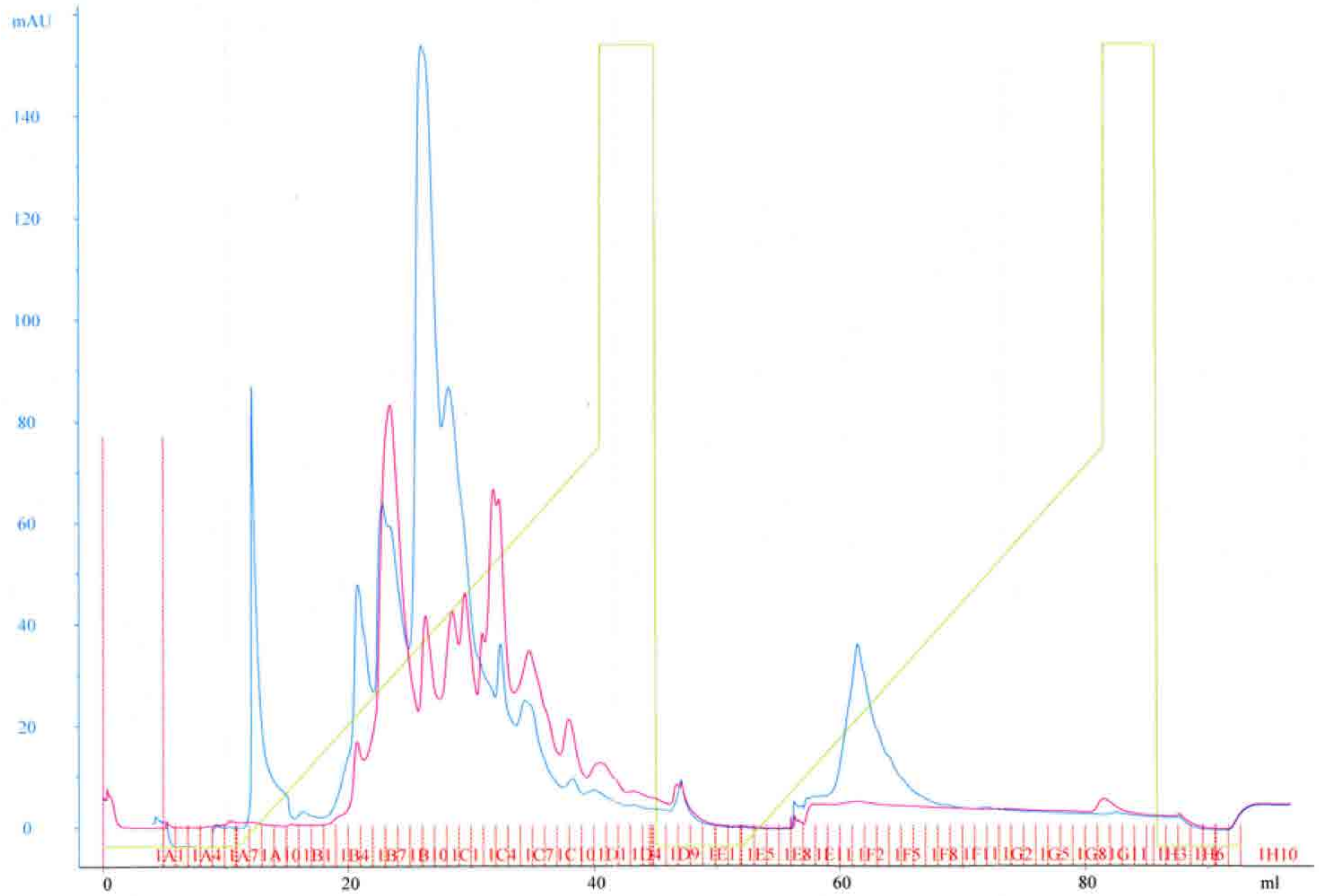


Figure 3.4 PROOF UV-trace of the GenWay depleted (pink) and non-depleted (blue) plasma proteins under mild conditions. The y-axis is UV absorbance in mAU. The x-axis depicts the amount of mobile phase run and collected, shown in ml's in black and the collection tubes in red. Both depleted and non-depleted plasma were run using the multi-dimensional ion exchange chromatography of PROOF on the ÄKTÄ HPLC system to highlight the ability of the system to do both sample types and the varied/enhanced chromatographic resolution with the depleted plasma compared to non-depleted. Note the enhanced resolution of the depleted plasma sample compared to the non-depleted for the SCX column. Additionally, note the removal of the single peak in the SAX column from the depleted sample compared to the non-depleted.

In order to maximise the number of protein identifications, equalisation of the amount of protein in each fraction was achieved through manual inspection of the UV trace and manual interpretation of the area under the curve. Since there was no discernable fractionation on the anionic column based on the chromatogram, it was pooled into one fraction, # 12, and the flowthrough from any unbound species was pooled in fraction # 1. The rest of the 10 fractions were broken up as equitably as possible, shown below in Figure 3.5. There were three main modifications to the elution gradient that we optimised and believe should be mentioned. The first was altering the seven-step gradient that was used in the beginning of the earliest proof of concepts for PROOF to a linear gradient. This at first did not seem to give any visible advantage, and it was not until we increased the length of the gradient to 120 min for each column that we observed enhanced chromatographic resolution. Having two 120 min gradients was deemed not practical at this stage of development, so we incorporated a truncated gradient that had a similar gradient angle as the 120 min gradient. We truncated it to 0 – 50 % buffer B, which is where the majority of the chromatographic resolution for ion exchange chromatography occurred, then stepped the gradient to 100% Buffer B, so as to make sure that any remaining bound species would be removed from the column and detected in our system. In conjunction with these gradient changes we altered the flow rate from 1 ml/min to 0.75 ml/min, which kept the pressure of the system to acceptable levels while showing enhanced chromatographic resolution as shown in Figure 3.5. The reproducibility of the system was sound enough that our own visual inspection of the UV trace chromatograms showed no discernable differences that would require the pooling of the samples from one run to the next to be tailored individually. Further evidence of the reproducibility of the system is shown in the mass spectrometric identification of the same peptides in similar fractions across replicates.

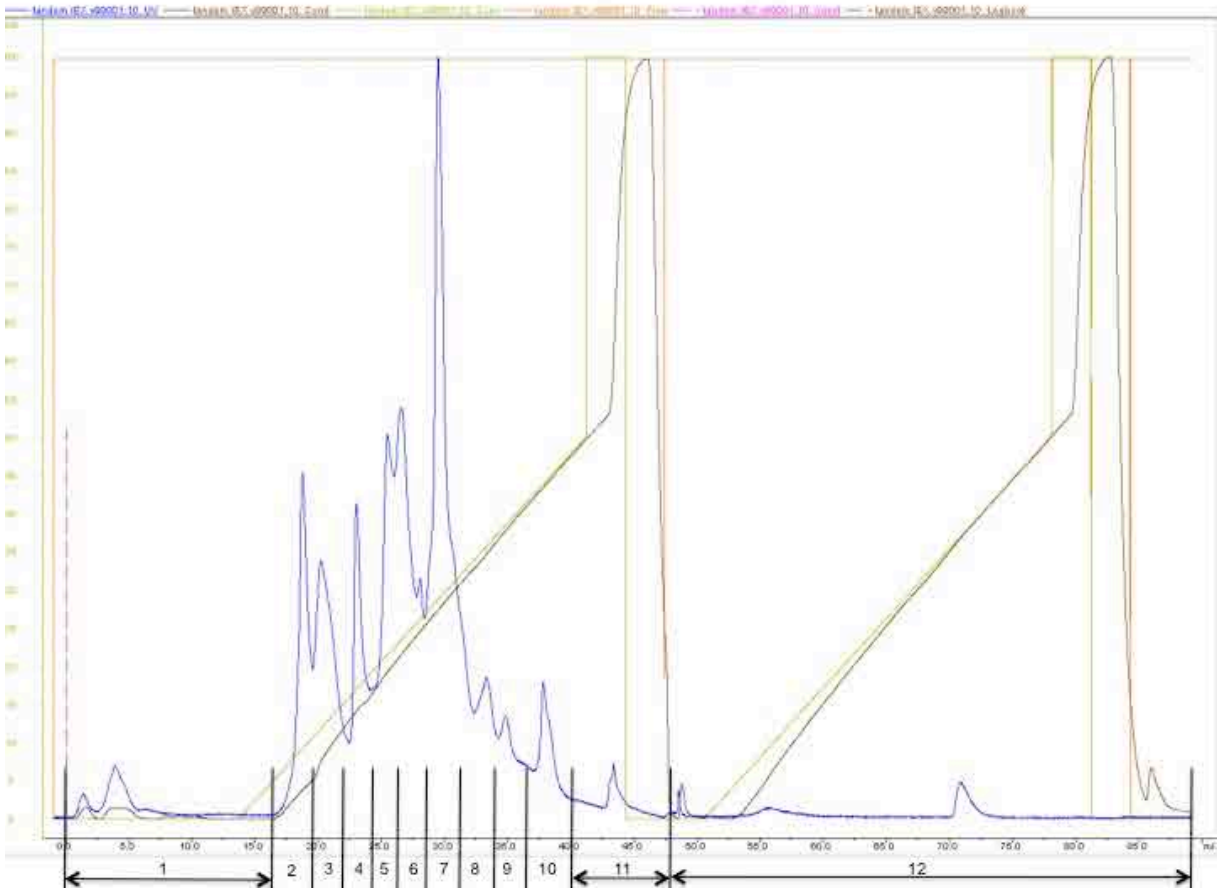


Figure 3.5 PROOF UV-trace of the GenWay depleted plasma proteins under mild conditions (blue) and the pooling of the fractions into 12 groups (black) for digestion and nano-RP-LC-MS/MS analysis. The y-axis represents the gradient for the addition of Buffer B as a percentage (%) of B (green). The x-axis depicts the amount of mobile phase run in ml and identifies where the 12 fractions for MS/MS were pooled based on manual inspection (black) of the UV trace. Fraction 1 is the flow through, 2 – 11 are SCX elution and 12 is SAX elution.

A representative UV trace of the strong cationic exchange chromatography (Std-SCX) of peptides from depleted plasma is shown in Figure 3.6. Inspection of the chromatograms (>10) showed a high degree of reproducibility to the human eye. The total elution time and gradient used was in the order of 130 min, and hence this is where and why we tried a similar gradient to start with for the PROOF. Since we were running two gradients, not one, and we needed to show comparative advantages for the PROOF system over Std-SCX beyond just the number of proteins identified, we made the necessary changes mentioned earlier to make the total time of the two gradient system similar to the single gradient system of Std-SCX on peptides.

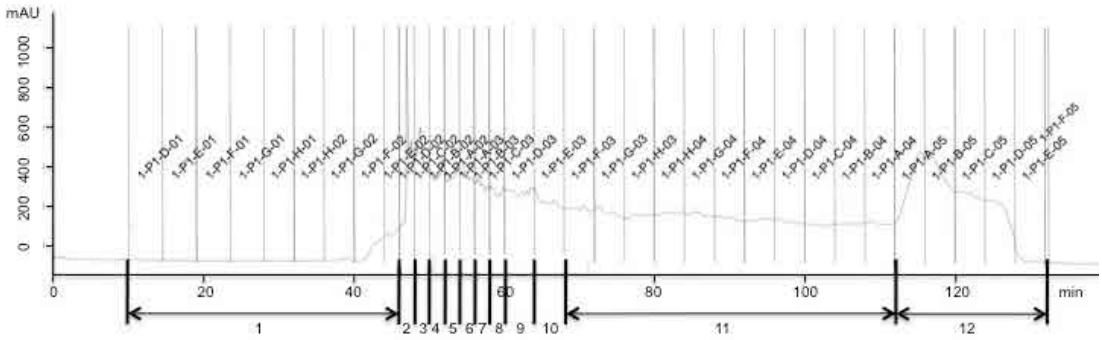


Figure 3.6 UV trace of Std-SCX of peptides from GenWay depleted plasma and the pooling of the fractions into 12 groups for MS/MS analysis. The pooling of the sample into 12 groups was done by manual inspection and interpretation of the UV trace so as to have an equitable amount of peptides in each of the twelve fractions.

3.4.2 Selection Criteria for High Stringency Identifications

When analyzing shotgun proteomic data generated from GPM with X!Tandem, a protein filtering of log e -1 (probability) is generally applied as the primary filter for the generation of a primary list of potential identification candidates with reverse database hits included. The investigator then has a number of options for the generation of a final list of protein identifications with confidence. Commonly one would manually inspect this list to find the reverse database hit with the lowest log e value (highest confidence) and then an initial calculation of the FDR (protein and peptide) is conducted. This is followed by subsequent calculations of the proceeding reverse database hits in order, from highest confidence to lowest confidence, until a protein FDR of between 1-2% is found, which is deemed as an acceptable selection criteria for many shotgun proteomics datasets [245, 390]. At this point all hits above

this value are kept and the rest of the dataset is omitted to generate a list of high confidence protein identifications.

When dealing with replicate datasets (three, for example) an investigator would similarly apply the log e -1 as the primary filter as above and then have the option to undergo a secondary filtering by combining the three to present one list with only identifications that are found in all three replicates, while omitting any identifications that were only made in one or two of the replicates [158]. This protein identification list would also still have the reverse database hits within it. This list is then manually interrogated for the calculation of a protein and peptide FDR as above to confirm the confidence in the dataset and to also draw a line of acceptance within the dataset for inclusion and omission around the FDR line. It has been shown by members of our research group that more often than not, there is no need to truncate the data after the combine filter, because the FDR value is within acceptable limits [391, 392].

The ideology behind such a combine filter is to use a fundamental principle of scientific investigation, that is so as to be confident in any observation then you must see it at least three times. Traditionally, a single peptide identification for the presence of a protein was acceptable [393]. The criteria for a confident identification is constantly evolving, where some would argue that two peptides are sufficient, while others argue that 3 or more are required and some argue a need for the distinction between unique and non-unique peptides be made when stating an identification [394]. The use of a combining filter similar to what is mentioned above tries to use one peptide to identify a protein, provided it is reproducibly seen across three or more events. Any identifications occurring in 3 replicates are by definition highly unlikely to be random, so the combine filter approach minimizes random noise in the data set.

Applying that approach, the ordering of the proteins in the output file from the combined dataset is based on the number of peptides used to make the identification with the lowest possible number being three, representing one peptide for each replicate at the bottom of the list. In some instances there is no need to further filter the dataset after the primary log e -1 and combine filtering is applied, but in some instances there is. This can be conducted by either increasing the stringency of the log e value in the primary filter, to log e – 2 for example, or to re-filter the data after

the combine filter to produce an acceptable FDR. The ultimate goal of any of these methods is to produce the largest possible list of high confidence protein identifications with the lowest possible false discovery rate.

Datasets change from experiment to experiment and across sample types for a multitude of reasons. For example, the variable composition of one proteome to the next between sample types, or the discrepancy between the number of peptides present in a digest of a proteome and the analytical capacity of the LC-MS/MS systems can prevent a perfectly reproducible set of peptides from being identified in repeat analyses of the same sample [395].

The goal of the experiments within this work on the protein fractionation of plasma was not to generate the largest list of protein identifications; it was to generate a list of proteins with the highest level of confidence and stringency so that any interpretation of the proteins identified utilizing the unique graphical representations would be true. In order to achieve this, I compared the ‘filter then combine’ approach against the ‘combine then filter’ approach to see which gave the most protein identifications at the highest confidence.

3.4.3 False Discovery Rate of PROOF and MudPIT

3.4.3.a Filter then Combine

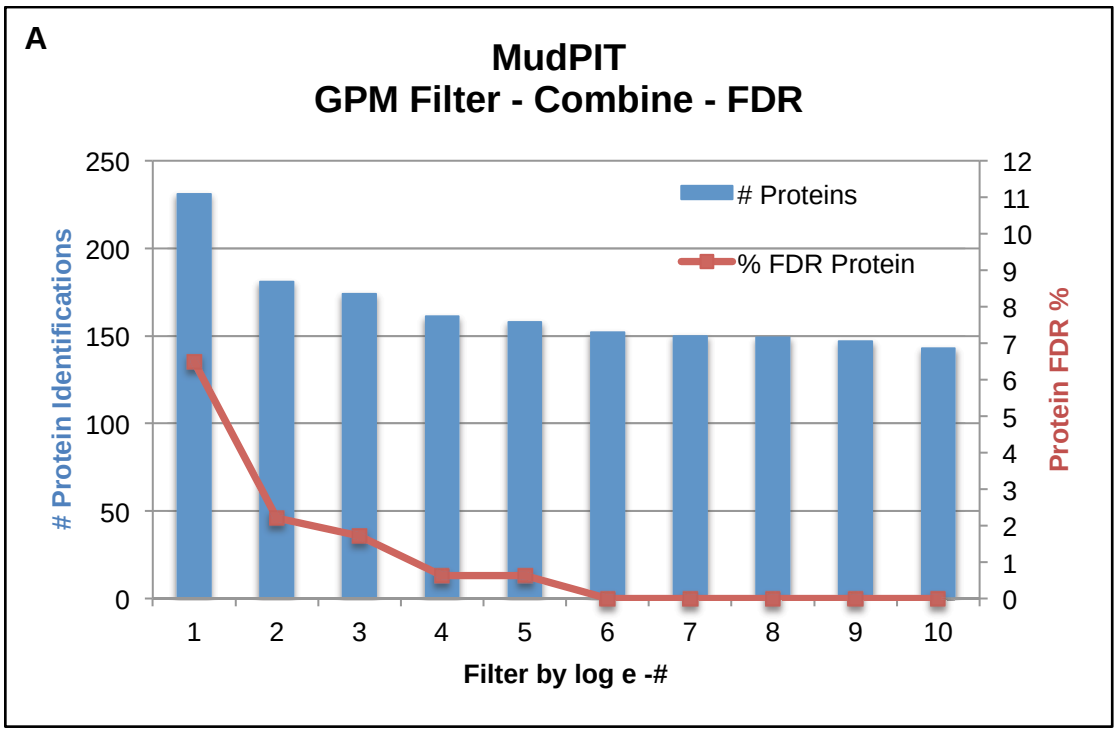
In order to calculate FDRs for the plasma PROOF and MudPIT datasets, I began with lists of proteins with reverse database hits included for all hits with a confidence of greater than log e -1. These were then combined into a single dataset and the protein FDR was calculated. The primary filter was then increased to protein log e value -2 and the combining and FDR steps were repeated. This process was repeated iteratively for both experimental datasets, using protein log e values from -1 to -10. The results of these analyses are shown in Figures 3.7.1 A and B below.

The point on both graphs where there is no FDR to report and little to no change in the number of proteins identified as the log e value increases is at a protein log e value -8 for the primary filter. Inspection of the individual replicate datasets before the

combine filter was applied showed that using a primary filtering of protein log e value -8 gave FDRs for the individual datasets ranging from 2 to 5%, while using a protein log e value of -10 gave primary datasets with no reverse database hits and thus a zero FDR. This indicates that log e -10 was too stringent and log e -8 was not stringent enough.

A protein log e value of -9 for the primary filter produces individual replicate datasets containing 211 +/- 13 identifications with FDR 1.36 +/- 0.83% for MudPIT, and for PROOF 233 +/- 22 identifications with FDR 2.04 +/- 0.77%. When the combine filtering is applied to generate the final high stringency and confidence list of proteins, it produces 147 for MudPIT and 146 for PROOF with no reverse database hits, thus a FDR of zero.

I chose to use log e -9 as the primary filter for datasets for further analysis, to ensure the highest stringency while retaining as much of the data as possible. It was this dataset that was used to establish and generate the unique graphical interpretation reported later in this chapter.



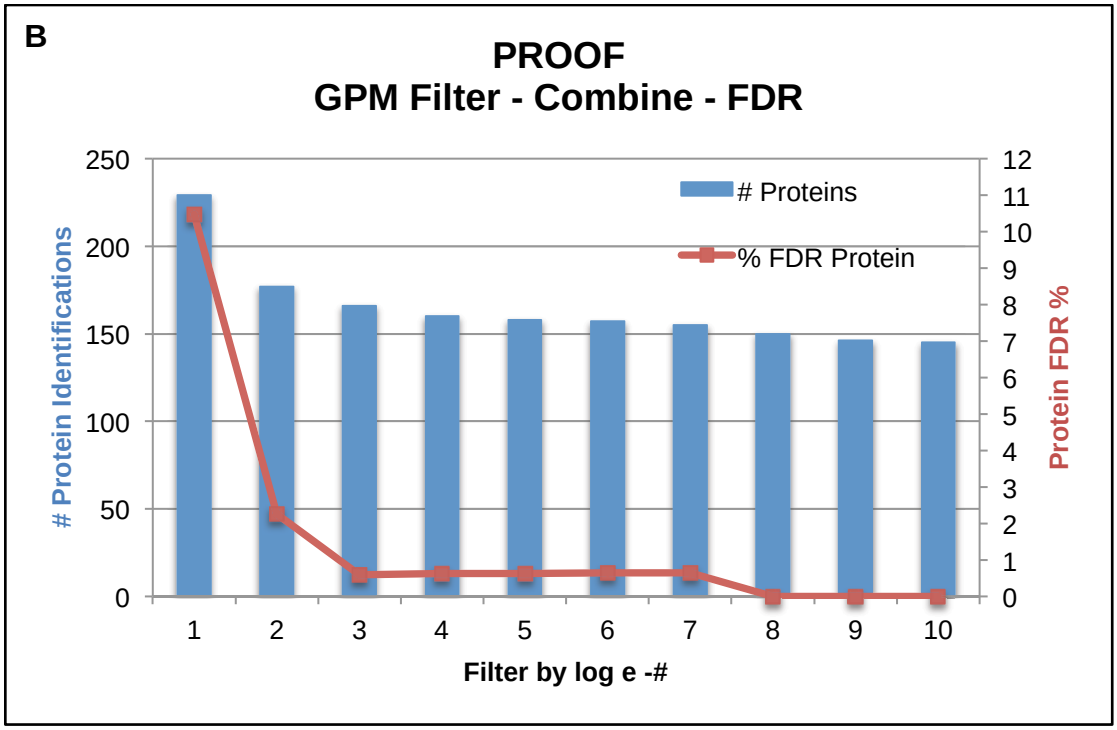


Figure 3.7.1. Graphical representation of the plasma datasets analysed using various protein log e – values (GPM filter) before using the combine filter. The x-axis depicts the log e – values (GPM filter) used prior to applying the combine filter. The total number of protein identified (blue) and %FDR (red) are shown on the two y-axes after the implementation of the combine filter at the various log e - values. The FDR goes to zero at log e -6 and -8 for MudPIT and PROOF, respectively

PROOF		Using the log e -9 cut-off	
PROTEIN FDR	# PROTEIN ID's	# REVERSED PROTEIN ID's	PROTEIN FDR
REPLICATE 1	260	3	1.15%
REPLICATE 2	251	7	2.79%
REPLICATE 3	219	3	1.37%
PEPTIDE FDR	# PEPTIDE ID's	# REVERSED PEPTIDE ID's	PEPTIDE FDR
REPLICATE 1	45252	10	0.02%
REPLICATE 2	42006	23	0.05%
REPLICATE 3	43444	9	0.02%

Std-SCX Peptides		Using the log e -9 cut-off	
PROTEIN FDR	# PROTEIN ID's	# REVERSED PROTEIN ID's	PROTEIN FDR
REPLICATE 1	232	5	2.16%
REPLICATE 2	205	1	0.05%
REPLICATE 3	214	2	0.93%
PEPTIDE FDR	# PEPTIDE ID's	# REVERSED PEPTIDE ID's	PEPTIDE FDR
REPLICATE 1	31949	14	0.04%
REPLICATE 2	25500	4	0.02%
REPLICATE 3	34416	28	0.08%

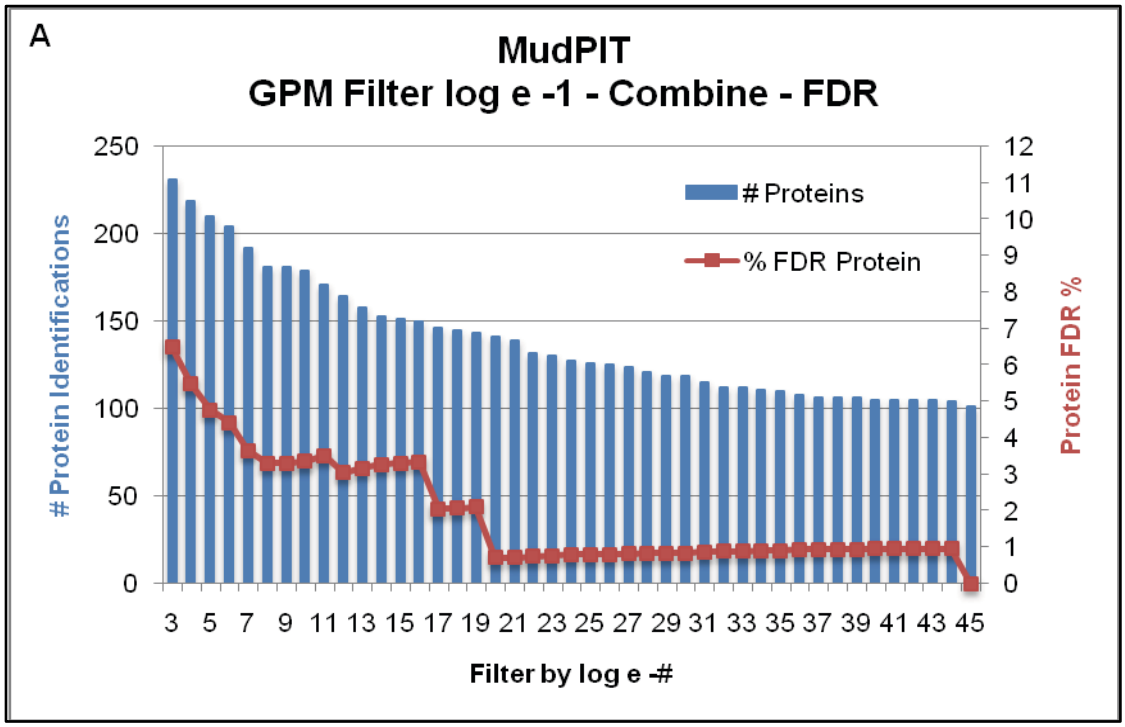
TABLE 3.1 False Discovery Rate (FDR) analysis with reversed data base hits was conducted on both the Std-SCX of peptides and PROOF RP-nano-LC-MS/MS identifications, generated using the X!Tandem algorithm from the Global Proteome Machine (GPM). An exclusion limit of log e -9 was used and created an average of 215 and 240 protein identifications for Std-SCX of peptides and PROOF respectively.

3.4.3.b Combine then Filter

I conducted a second analysis of the same data using the ‘combine then filter’ method of applying a primary filter of protein log e -1 followed by the combine filtering and then the calculation of FDR followed by further filtering if required [391, 392]. This dataset produced 231 protein identifications with 15 reverse database hits to give a protein FDR of 6.50% for the MUDPIT fractionation, and for the PROOF fractionation 229 protein identifications with 24 reverse database hits to give a protein FDR of 10.50%.

I then applied a secondary protein log e value filter on the combined datasets from the PROOF and MudPIT experiments. Graphs 3.7.2 A and B show the relationship between protein log e value used for secondary filtering of the combined datasets and the resulting FDRs.

It is clear that there is an inflection point in the graphs, at around log e -20 for PROOF and log e -16 for MudPIT. Beyond this point, only a small number of reverse database hits are retained, and they can not be removed entirely without significantly reducing the datasets. At those inflection points, the MudPIT dataset contains 143 proteins at FDR of 2.1% and the PROOF dataset contains 142 proteins at 1.41% FDR. This is similar to, but not quite as good as, the datasets produced in the preceding section using the ‘filter then combine’ approach.



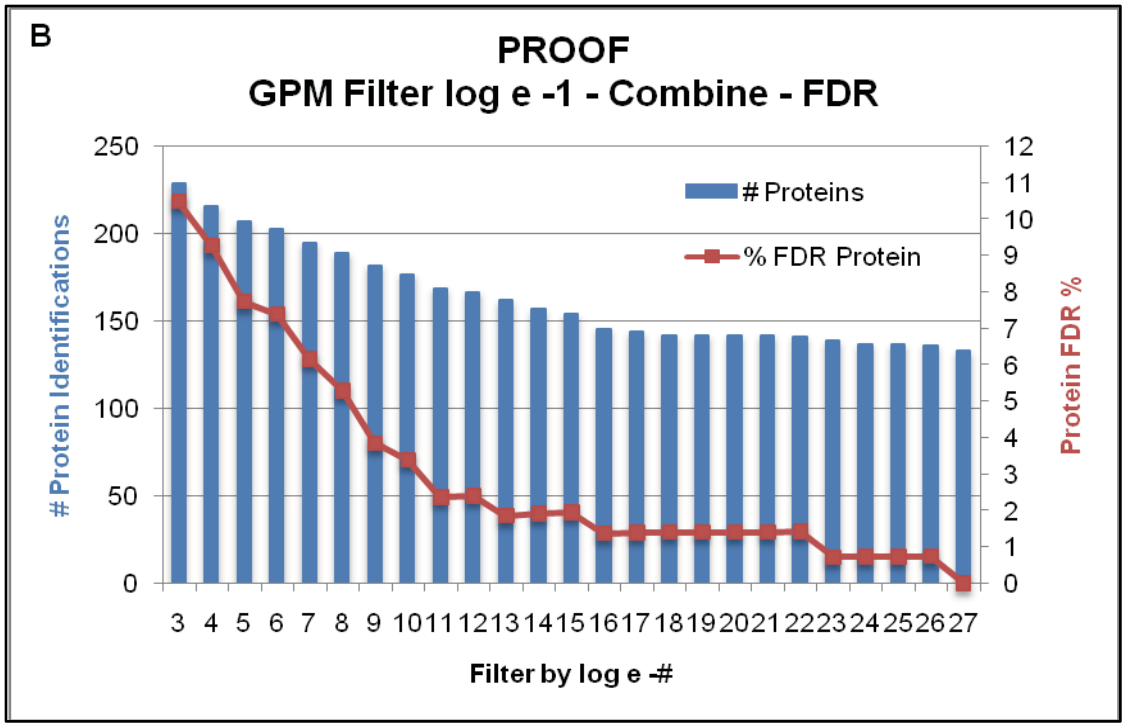


Figure 3.7.2. Graphical representation of the plasma dataset produced after the combine filter and use of the log e -1 as the preliminary filter. The x-axis depicts the secondary log e – value of the 'combine then filter' data. The y-axis depicts the total number of protein identified (blue) and % FDR (red) after implementation of the combine filter at log e -1. The FDR rate goes to zero at protein log e value of -45 and -27 for MudPIT and PROOF respectively.

3.4.4 Comparison of proteins found using PROOF and MudPIT

Despite the distinct nature of the two techniques, one being the fractionation of proteins and the other being the fractionation of peptides, there is approximately 74% similarity in the total number of proteins identified. Hence, the PROOF method has a similar ability to interrogate and identify proteins in a complex system, and achieves similar coverage of a proteome to the MudPIT approach. There is still however, enough difference between the two that would justify employing both techniques in order to maximise protein identifications for further directed experimentation.

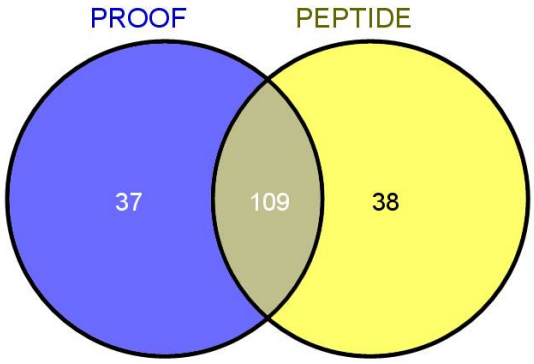


Figure 3.8 This Venn diagram above shows that both methods identified 109 similar proteins. There were 37 and 38 unique protein identifications from the PROOF and Std-SCX of peptides respectively. The final list of protein identifications were 146 and 147 for PROOF and Std-SCX of peptides respectively, after the removal of any identifications that were not found in all three replicates.

The proteins identified uniquely in each approach are presented in Tables 3.2 and 3.3 on the following pages. The proteins identified in both approaches are presented in Table 3.4. It is worth noting that many of the top 12 most abundant plasma proteins were identified in the depleted plasma. This included, alpha-2-macroglobulin, haptoglobin, apolipoprotein A, alpha-1-acid-glycoprotein, alpha-1-antitrypsin, fibrinogen and many immunoglobulin fragments; surprisingly, albumin was not detected. More keratins were identified using the PROOF approach compared to MudPIT. This may be attributed to the extra sample handling that the PROOF approach requires, or it may be a random event.

Proteins identified only in MudPIT			
Identifier	**Mr (kDa)**	**Peptide ID's**	**Description**
IPI00739237.1	44.9	1630	SIMILAR TO COMPLEMENT C3
IPI00947496.1	123.9	636	UNCHARACTERIZED PROTEIN.
IPI00025426.3	163.8	550	ISOFORM 1 OF PREGNANCY ZONE PROTEIN.
IPI00784817.1	52.3	525	ANTI-RHD MONOCLONAL T125 GAMMA1 HEAVY CHAIN.
IPI00021891.5	51.5	516	ISOFORM GAMMA-B OF FIBRINOGEN GAMMA CHAIN.
IPI00816314.1	50.9	508	PUTATIVE UNCHARACTERIZED PROTEIN DKFZP686I15196.
IPI00784842.1	52	508	PUTATIVE UNCHARACTERIZED PROTEIN DKFZP686G11190.
IPI00644018.1	40.7	486	ISOFORM 2 OF ALPHA-1B-GLYCOPROTEIN.
IPI00418153.1	57	324	PUTATIVE UNCHARACTERIZED PROTEIN DKFZP686I15212.
IPI00423462.5	56.4	244	PUTATIVE UNCHARACTERIZED PROTEIN DKFZP686K18196 (FRAGMENT).
IPI00291262.3	52.5	236	ISOFORM 1 OF CLUSTERIN.
IPI00020996.5	66	223	INSULIN-LIKE GROWTH FACTOR-BINDING PROTEIN COMPLEX ACID LABILE SUBUNIT.
IPI00021439.1	41.7	168	ACTIN, CYTOPLASMIC 1.
IPI00383338.1	55.3	158	PRO2769.
IPI00021842.1	36.1	74	APOLIPOPROTEIN E.
IPI00736885.1	12.3	68	IG KAPPA CHAIN V-II REGION TEW.
IPI00830088.1	12.8	49	RHEUMATOID FACTOR RF-IP14.
IPI00007047.1	10.8	36	PROTEIN S100-A8.
IPI00218413.2	61.1	35	BIOTINIDASE.
IPI00029061.3	42.7	35	SELENOPROTEIN P.
IPI00748998.1	25.6	31	SINGLE-CHAIN FV (FRAGMENT).
IPI00022418.2	262.5	29	ISOFORM 1 OF FIBRONECTIN.
IPI00003590.2	82.5	29	ISOFORM 1 OF SULFHYDRYL OXIDASE 1.
IPI00025276.2	464	25	ISOFORM XB OF TENASCIN-X.
IPI00854644.2	16.5	24	UNCHARACTERIZED PROTEIN
IPI00178926.2	18.1	23	IMMUNOGLOBULIN J CHAIN.
IPI00032311.4	53.3	21	LIPOPOLYSACCHARIDE-BINDING PROTEIN.
IPI00783471.1	14.5	21	IMMUNGLOBULIN HEAVY CHAIN VARIABLE REGION (FRAGMENT).
IPI00007199.4	55.1	21	PROTEIN Z-DEPENDENT PROTEASE INHIBITOR.
IPI00298994.6	269.6	18	TALIN-1.
IPI00010471.6	70.2	18	PLASTIN-2.
IPI00022331.1	49.5	17	PHOSPHATIDYLCHOLINE-STEROL ACYLTRANSFERASE.
IPI00029193.1	70.6	16	HEPATOCYTE GROWTH FACTOR ACTIVATOR.
IPI00022432.1	15.9	13	TRANSTHYRETIN.
IPI00027482.1	45.1	13	CORTICOSTEROID-BINDING GLOBULIN.
IPI00021263.3	27.7	12	14-3-3 PROTEIN ZETA/DELTA.
IPI00387120.1	12.6	12	IG KAPPA CHAIN V-IV REGION LEN.
IPI00024825.2	151	11	ISOFORM A OF PROTEOGLYCAN 4.

Table 3.2 List of proteins identified unique to the MudPIT of peptides method. The IPI identifier is column one, column two is the molecular weight, column three is the total number of peptide MS/MS identifications made for each protein identification. The last column is the generic name for the proteins.

Proteins identified only in PROOF			
Identifier	**Mr (kDa)**	**Peptide ID's**	**Description**
IPI00887154.2	192.6	3202	COMPLEMENT COMPONENT 4B.
IPI00645038.1	105.2	1911	INTER-ALPHA (GLOBULIN) INHIBITOR H2.
IPI00218192.3	101.2	1465	ISOFORM 2 OF INTER-ALPHA-TRYPSIN INHIBITOR HEAVY CHAIN H4.
IPI00947137.1	34.7	1179	PROTEIN.
IPI00645363.2	51.7	1128	PUTATIVE UNCHARACTERIZED PROTEIN DKFZP686P15220.
IPI00785084.2	51.1	1126	IGH@ PROTEIN.
IPI00930124.1	52.1	1116	PUTATIVE UNCHARACTERIZED PROTEIN DKFZP686C11235.
IPI00896380.1	51.8	741	ISOFORM 2 OF IG MU CHAIN C REGION.
IPI00969547.1	56.8	521	UNCHARACTERIZED PROTEIN
IPI00816741.1	123.3	376	COMPLEMENT COMPONENT 5 VARIANT (FRAGMENT).
IPI00969620.2	28	317	LIGHT CHAIN FAB
IPI00022434.4	71.7	233	UNCHARACTERIZED PROTEIN.
IPI00719373.2	23	210	IMMUNOGLOBULIN LAMBDA LIKE POLYPEPTIDE
IPI00019359.4	62	175	KERATIN, TYPE I CYTOSKELETAL 9.
IPI00423461.3	54.1	152	PUTATIVE UNCHARACTERIZED PROTEIN DKFZP686C02220 (FRAGMENT).
IPI00021304.1	65.8	90	KERATIN, TYPE II CYTOSKELETAL 2 EPIDERMAL.
IPI00021428.1	42	87	ACTIN, ALPHA SKELETAL MUSCLE.
IPI00329775.7	48.4	74	CARBOXY PEPTIDASE B2
IPI00025204.1	38.1	69	CD5 ANTIGEN-LIKE.
IPI00217963.3	51.2	62	KERATIN, TYPE I CYTOSKELETAL 16.
IPI00879915.1	15.2	52	UNCHARACTERIZED PROTEIN.
IPI00384444.6	51.5	52	KERATIN, TYPE I CYTOSKELETAL 14.
IPI00555752.2	28.3	51	ISOFORM 2 OF C4B-BINDING PROTEIN BETA CHAIN.
IPI00004656.3	13.7	47	BETA-2-MICROGLOBULIN.
IPI00925540.1	81.8	38	HEPATOCYTE GROWTH FACTOR-LIKE PROTEIN PRECURSOR.
IPI00216691.5	15	35	PROFILIN-1.
IPI00009867.3	62.3	28	KERATIN, TYPE II CYTOSKELETAL 5.
IPI00021817.1	52	27	VITAMIN K-DEPENDENT PROTEIN C.
IPI00010295.1	52.3	26	CARBOXYPEPTIDASE N CATALYTIC CHAIN.
IPI00021364.1	51.2	23	PROPERDIN.
IPI00027843.1	44.7	22	ISOFORM 1 OF VITAMIN K-DEPENDENT PROTEIN Z.
IPI00215983.3	28.9	18	CARBONIC ANHYDRASE 1.
IPI00257882.7	54.5	15	XAA-PRO DIPEPTIDASE.
IPI00748955.3	68.9	15	PLATELET GLYCOPROTEIN IB ALPHA CHAIN.
IPI00028030.4	82.8	13	CARTILAGE OLIGOMERIC MATRIX PROTEIN.
IPI00746623.2	62.6	11	HYALURONAN-BINDING PROTEIN 2.
IPI00382440.1	11.5	11	IG LAMBDA CHAIN V-IV REGION HIL.

Table 3.3 List of the proteins identified unique to the ion exchange of proteins method. The IPI identifier is column one, column two is the molecular weight, column three is the total number of peptide MS/MS identifications made for each protein identification. The last column is the generic name for the proteins.

Table 3.4 Page # 1 of 3

1

Proteins identified in both MudPIT and PROOF

Identifier	Mr (kDa)	STD-SCX Peptide ID's	PROOF Peptide ID's	Description
IPI00783987.2	187	12917	13961	COMPLEMENT C3 (FRAGMENT).
IPI00887739.3	144.7	11227	10530	COMPLEMENT C3-LIKE, PARTIAL.
IPI00022488.1	51.6	2756	5548	HEMOPEXIN.
IPI00017601.1	122.1	1822	3736	CERULOPLASMIN.
IPI00550991.3	50.6	2290	2361	CDNA FLJ35730 FIS, CLONE TESTI2003131, HIGHLY SIMILAR TO ALPHA-1-ANTICHYMOTRYPSIN.
IPI00418163.3	192.6	1760	3199	COMPLEMENT C4-B PREPROPROTEIN.
IPI00643525.1	192.6	1905	3192	UNCHARACTERIZED PROTEIN.
IPI00555812.5	52.9	1298	2694	VITAMIN D-BINDING PROTEIN PRECURSOR.
IPI00478003.3	163.2	2073	2099	ALPHA-2-MACROGLOBULIN.
IPI00022229.2	515.3	2028	2358	APOLIPOPROTEIN B-100.
IPI00032179.3	52.6	1904	858	ANTITHROMBIN-III.
IPI00029739.5	139	965	1540	ISOFORM 1 OF COMPLEMENT FACTOR H.
IPI00022895.7	54.2	840	1893	ALPHA-1B-GLYCOPROTEIN PRECURSOR
IPI00019580.1	90.5	913	1838	PLASMINOGEN.
IPI00019591.2	140.9	789	2067	CDNA FLJ55673, HIGHLY SIMILAR TO COMPLEMENT FACTOR B.
IPI00019568.1	70	574	1783	PROTHROMBIN (FRAGMENT).
IPI00292530.1	101.3	777	2017	INTER-ALPHA-TRYPSIN INHIBITOR HEAVY CHAIN H1.
IPI00215894.1	47.9	647	1519	ISOFORM LMW OF KININOGEN-1.
IPI00032328.2	71.9	646	1337	ISOFORM HMW OF KININOGEN-1.
IPI00021727.1	67	451	1050	C4B-BINDING PROTEIN ALPHA CHAIN.
IPI00641737.1	46.7	517	944	HAPTOGLOBIN.
IPI00298971.1	54.3	725	889	VITRONECTIN.
IPI00304273.2	45.4	624	856	APOLIPOPROTEIN A-IV.
IPI00892870.1	51.5	710	825	IMMUNOGLOBULIN HEAVY CONSTANT MU
IPI00298497.3	55.9	665	727	FIBRINOGEN BETA CHAIN.
IPI00298828.3	38.3	557	679	BETA-2-GLYCOPROTEIN 1.
IPI00550731.2	26.2	803	664	PUTATIVE UNCHARACTERIZED PROTEIN.
IPI00784865.1	25.8	749	653	IGK@ PROTEIN.
IPI00022395.1	63.1	489	607	COMPLEMENT COMPONENT C9.
IPI00021885.1	94.9	361	606	ISOFORM 1 OF FIBRINOGEN ALPHA CHAIN.
IPI00032291.2	188.2	491	575	COMPLEMENT C5.
IPI00291866.5	55.1	553	567	PLASMA PROTEASE C1 INHIBITOR.
IPI00218999.3	49.9	333	560	UNCHARACTERIZED PROTEIN.
IPI00220327.4	66	142	525	KERATIN, TYPE II CYTOSKELETAL 1.
IPI00026314.1	85.6	479	508	ISOFORM 1 OF GELSOLIN.
IPI00477597.2	39	279	493	ISOFORM 1 OF HAPTOGLOBIN-RELATED PROTEIN.
IPI00032220.3	53.1	410	477	ANGIOTENSINOGEN.
IPI00291867.3	65.7	203	464	COMPLEMENT FACTOR 1 PRECURSOR
IPI00163207.1	62.2	226	417	ISOFORM 1 OF N-ACETYLMURAMOYL-L-ALANINE AMIDASE.
IPI00022391.1	25.4	299	417	SERUM AMYLOID P-COMPONENT.
IPI00022431.2	46.6	280	413	CDNA FLJ55606, HIGHLY SIMILAR TO ALPHA-2-HS-GLYCOPROTEIN.
IPI00022426.1	39	215	395	PROTEIN AMBP.
IPI00879709.3	105.7	198	361	COMPLEMENT COMPONENT C6 PRECURSOR.
IPI00294004.1	75.1	63	356	VITAMIN K-DEPENDENT PROTEIN S.
IPI00386879.1	53.1	250	353	CDNA FLJ14473 FIS, CLONE MAMMA1001080, HIGHLY SIMILAR TO HOMO SAPIENSSNC73 PROTEIN (SNC73) MRNA.
IPI00022463.1	77	169	349	SEROTRANSFERRIN PRECURSOR

Table 3.4 Page # 2 of 3

2

Proteins identified in both MudPIT and PROOF

Identifier	Mr (kDa)	STD-SCX Peptide ID's	PROOF Peptide ID's	Description
IPI00292950.4	60.1	287	337	SERPIN PEPTIDASE INHIBITOR, CLADE D (HEPARIN COFACTOR), MEMBER 1.
IPI00399007.7	46	218	335	PUTATIVE UNCHARACTERIZED PROTEIN DKFZP686I04196 (FRAGMENT).
IPI00154742.6	24.8	223	333	IGL@ PROTEIN.
IPI00017696.1	76.6	243	329	COMPLEMENT C1S SUBCOMPONENT.
IPI00294395.1	67	179	326	COMPLEMENT COMPONENT C8 BETA CHAIN.
IPI00028413.8	99.8	91	306	ISOFORM 1 OF INTER-ALPHA-TRYPSIN INHIBITOR HEAVY CHAIN H3.
IPI00296165.6	81.8	201	274	CDNA FLJ54471, HIGHLY SIMILAR TO COMPLEMENT C1R SUBCOMPONENT.
IPI00022420.3	23	278	274	RETINOL-BINDING PROTEIN 4.
IPI00022371.1	59.5	235	262	HISTIDINE-RICH GLYCOPROTEIN.
IPI00930442.1	52.4	138	260	PUTATIVE UNCHARACTERIZED PROTEIN DKFZP686M24218.
IPI00019581.1	67.8	75	256	COAGULATION FACTOR X11 PRECURSOR
IPI00296608.6	93.5	151	252	COMPLEMENT COMPONENT C7.
IPI00006114.4	46.3	123	242	PIGMENT EPITHELIUM-DERIVED FACTOR
IPI00303963.1	83.2	102	230	COMPLEMENT C2 (FRAGMENT).
IPI00019943.1	69	139	223	AFAMIN.
IPI00940245.1	44.8	184	218	IMMUNOGLOBULIN HEAVY CHAIN VARIANT (FRAGMENT).
IPI00021841.1	30.8	260	216	APOLIPOPROTEIN A-I.
IPI00011264.2	37.6	130	215	COMPLEMENT FACTOR H-RELATED PROTEIN 1.
IPI00011261.2	22.3	111	203	COMPLEMENT COMPONENT C8 GAMMA CHAIN.
IPI00020986.2	38.4	66	200	LUMICAN.
IPI00022389.1	25	40	198	ISOFORM 1 OF C-REACTIVE PROTEIN.
IPI00011252.1	65.1	96	193	COMPLEMENT COMPONENT C8 ALPHA CHAIN.
IPI00022394.2	25.8	184	192	COMPLEMENT C1Q SUBCOMPONENT SUBUNIT C.
IPI00022429.3	23.5	379	174	ALPHA-1-ACID GLYCOPROTEIN 1.
IPI00479116.1	60.6	77	153	CARBOXYPEPTIDASE N SUBUNIT 2.
IPI00009865.4	58.8	23	149	KERATIN, TYPE I CYTOSKELETAL 10.
IPI00477992.1	26.7	106	134	COMPLEMENT C1Q SUBCOMPONENT SUBUNIT B.
IPI00029863.4	55	126	129	55 KDA PROTEIN.
IPI00909594.1	53.7	98	121	CDNA FLJ58413, HIGHLY SIMILAR TO COMPLEMENT COMPONENT C7.
IPI00553177.1	46.7	216	110	ISOFORM 1 OF ALPHA-1-ANTITRYPSIN.
IPI00916434.1	25.1	57	102	ANTI-(ED-B) SCFV (FRAGMENT).
IPI00783287.1	13.4	67	79	IMMUNGLOBULIN HEAVY CHAIN VARIABLE REGION (FRAGMENT).
IPI00163446.4	47.4	27	75	ISOFORM 2 OF IG DELTA CHAIN C REGION.
IPI00006154.1	30.6	29	69	ISOFORM LONG OF COMPLEMENT FACTOR H-RELATED PROTEIN 2.
IPI00654888.4	71.3	86	68	PLASMA KALLIKREIN.
IPI00296099.6	129.3	45	67	THROMBOSPONDIN-1.
IPI00296176.2	51.7	44	61	COAGULATION FACTOR IX.
IPI00022445.1	13.9	22	61	PLATELET BASIC PROTEIN.
IPI00328609.3	48.5	45	59	KALLISTATIN.
IPI00023673.1	65.3	31	59	GALECTIN-3-BINDING PROTEIN.
IPI00293925.2	32.9	58	58	ISOFORM 1 OF FICOLIN-3.
IPI00019576.1	54.7	20	55	COAGULATION FACTOR X.
IPI00009028.1	22.6	31	54	TETRANECTIN
IPI00027235.1	158.4	16	51	ISOFORM 1 OF ATTRACTIN.
IPI00029260.2	40.1	65	51	MONOCYTE DIFFERENTIATION ANTIGEN CD14.

Table 3.4 Page # 3 of 3

# 3				**Proteins identified in both MudPIT and PROOF**
Identifier	Mr (kDa)	**STD-SCX** Peptide ID's	**PROOF** Peptide ID's	Description
IPI00915820.1	13.3	29	44	ANTI-FOLATE BINDING PROTEIN (FRAGMENT).
IPI00027462.1	13.2	28	42	PROTEIN S100-A9.
IPI00020091.1	23.6	165	40	ALPHA-1-ACID GLYCOPROTEIN 2.
IPI00218732.4	39.7	44	39	SERUM PARAOXONASE/ARYLESTERASE 1.
IPI00022417.4	38.2	13	38	LEUCINE-RICH ALPHA-2-GLYCOPROTEIN.
IPI00009793.4	53.5	14	35	COMPLEMENT C1R SUBCOMPONENT-LIKE PROTEIN.
IPI00003351.2	60.6	37	34	ISOFORM 1 OF EXTRACELLULAR MATRIX PROTEIN 1.
IPI00023019.1	43.8	21	31	ISOFORM 1 OF SEX HORMONE-BINDING GLOBULIN.
IPI00021854.1	11.2	22	29	APOLIPOPROTEIN A-II.
IPI00022937.4	252	36	28	252 KDA PROTEIN.
IPI00816799.1	12.8	34	27	RHEUMATOID FACTOR D5 LIGHT CHAIN (FRAGMENT).
IPI00292946.1	46.3	77	23	THYROXINE-BINDING GLOBULIN.
IPI00025864.5	72.8	22	17	BUTYRYLCHOLINESTERASE, ISOFORM CRA_B.
IPI00395488.2	71.7	5	15	VASORIN.
IPI00171678.4	69	26	14	DOPAMINE BETA-HYDROXYLASE.
IPI00026199.2	25.5	56	14	GLUTATHIONE PEROXIDASE 3.
IPI00064667.4	56.7	27	12	BETA-ALA-HIS DIPEPTIDASE PRECURSOR
IPI00291175.7	116.6	20	8	ISOFORM 1 OF VINCULIN.

Table 3.4 List of the proteins identified similarly by both MudPIT of peptides and PROOF methods. The IPI identifier is column one, column two is the molecular weight, column three and four is the total number of peptide MS/MS identifications made for each protein identification for both Std-SCX of peptides and PROOF respectively. The last column is the generic description name for the proteins.

The protein identification lists generated (Tables 3.2 – 3.4) and the comparative Venn diagram in Figure 3.8 do not present a strong enough argument for the use of the PROOF method as either a complementary or supplementary method to the standard practices of today. However, the extra information gained from the graphical data representation shown in the following sections provides a strong justification for further exploration of this approach as a complementary method.

3.4.5 Novel Graphical Representation of PROOF

As explained in the methods section, protein identification data from the PROOF dataset was re-examined graphically. Graphs were created to display where in the protein level chromatographic fractionation each of the peptides from a particular protein was identified. The number of times each peptide was identified, as a surrogate for relative abundance level, was incorporated into the graphs as well.

In this graphical representation, for each peptide identified a spherical dot appears next to the peptide string and in-line with the particular fraction that it was identified in. The size of each dot is directly related to the number of times that peptide string was identified in that fraction. We also created a distinction between unique peptides in red and non-unique peptides in black.

The main reason that I wanted to try and show the distinction between the unique and non-unique peptides is because at the heart of the protein inference problem is the allocation of non-unique peptides to multiple protein identifications; this also has significant impact on quantitation within a proteome. The quantification side of this equation I will address in Chapter 4, while for this Chapter I am more interested in the correct assignment of identification for the whole protein, species or truncated variants. These graphs also show the potential for the correct assignment of non-unique peptides to proteins. This information could go towards the establishment of a more precise way of weighting peptides for spectral counting quantification for non-unique peptides [257, 396].

3.4.5.a Additional Information Revealed by Graphical Display

Peptide display graphs were generated for all 146 identified proteins and are supplied in the supplemental data 3.2. I will focus here on five selected graphs shown in Figure 3.9 (A-E), which reveal additional potential information at the protein level using the PROOF approach. Each graph represents a single protein identification and each of the dots represents an MS/MS identification of a peptide that was assigned to that protein for its identification. The size of each dot is directly proportional to the number of MS/MS identifications for each peptide, the smallest dots represent a single identity hit and the more identity hits the larger the dot. The numbers of peptides used for each identification are also displayed in Tables 3.2 – 3.4.

The y-axis in Figures 3.9 (A-E) represents each of the peptide sequence strings that were detected by the mass spectrometer and allocated for the identification. The order of the peptide strings in the y-axis are starting at the N-terminus and proceeding toward the C-terminus, from top to bottom, respectively. The y-axis is

also spatially ordered relative to each of the three replicates so that the same peptide string on one replicate is placed in the same height position relative to the other replicates for the same peptide string. This relative positioning of the peptide strings was prepared so that if one of the protein identifications in one of the replicates was more N-terminal centric and the other were more C-terminal centric then it would not skew the data display.

The five proteins highlighted in the proceeding graphs are denoted in the section headings with a ‘unique’ if found by PROOF only and with ‘both’ if identified by both methods: (IPI00423461.3 unique, IPI00298828.3 both, IPI00218192.3 unique, IPI00019591.2 both and IPI00303963.1 both). Three of the five are found in both methods. Inspection of the MudPIT data deeper to see if these two unique proteins above were perhaps missed by the high stringency selection was conducted. Utilising the log e -1 data and the combine filter analysis, which give a protein FDR of 6.5% and 216 protein identifications, did not show the identity of the two proteins IPI00423461.3 and IPI00218192.3. Interestingly, IPI00218192.3 is shown below to be potentially in two forms from the PROOF data and may be an example of the PROOF method in some way extracting or enabling this protein, and similar ones, to be made more available to the trypsin for cleavage and detection by the mass spectrometer.

One may notice when viewing Figures 3.9 (A-E) that it would appear the chromatographic resolution is broad and not well defined at best, due to chromatographic tailing of peptides. Peptides are identified at a start point at a particular fraction and seem to continue eluting for multiple fractions and sometimes throughout the rest of the entire proceeding chromatography. Ion exchange chromatography is known to be low in resolution [311], which is shown in the Std-SCX of peptides UV-trace in Figure 3.6 too. One can observe, however, that the beginning of the identifications, signifying when the protein starts to elute off the column is sharply defined. This trailing effect of peptides in our PROOF graphs in Figures 3.9 (A-E) could however be attributed to non-unique peptides rather than unique peptides, and if this were true then the peptide trails we are seeing are not only from the one protein due to poor chromatographic resolution but due to other proteins that have either been identified in later fractions or from proteins that did not have enough unique peptides identified so as to assign an identification to and were

missed in the standard identification analysis. Additionally, if one was to then recreate the graphs and only show the unique peptides in one graph and then the non-unique in another and then a combined one as the third, one would then theoretically be able to better distinguish between these possibilities.

In my opinion, the ultimate goal within the field of proteomics should be to strive for the highest levels of understanding and truth in the identification of proteins, with the underlying ideology that proteins in complex systems are most likely in more than one form or association and may have more than one function. Hence, if the current identification methods do not distinguish the multiple forms and functions of the proteins, then information is being misinterpreted or lost.

When interpreting the following graphs there are a number of generalised trends that I have observed and tried to apply to the interpretations thereof. One of these trends is that despite ion exchange chromatography being low in resolution and producing a trailing effect of peptides, the point at which the protein first begins to elute off the column seems to be quite a sharp point and is used as a stern line to discriminate one species eluting from the other. The trailing effect can also be used to a lesser extent because it can relate to the different concentrations of the protein species, though they can also relate to non-unique peptides from other proteins or peptides that have a higher probability of ionisation and detection by the mass spectrometer. The distinction of peptides in fraction 12, the SAX column, can mean a large difference in the potential conformation or associations of the protein if it is also found on the SCX column, due to large differences in the pKa for association needed to bind and elute from both. Hence, I use this as one example of a strong argument for the presence of more than one species or protein complex in the sample.

3.4.5.b Specific PROOF example number 1 - Protein IPI 00423461.3 (unique)

The protein IPI00423461.3 DKF2P686c02220 is a fragment of the immunoglobulin heavy constant alpha 2 (A2m marker).

Since Ig alpha is a major immunological class in body secretions, identifying it here in this human plasma sample is not unexpected. The subunit structure is known to be monomeric or polymeric and thus could have many forms and elution points in the chromatography.

Figure 3.9 (A) represents an example of what appears to be a truncated protein that is cleaved toward the N-terminal end of the sequence, resulting in two species that elute off the chromatography differently. The N-terminal end is the smaller fragment that is represented by the two peptides seen in fractions 1 – 3, that being peptide strings {1- QVQLVQSGAEVK} and {2 –QVQLVQSGAEVKKPGASVK}. Both of these peptides are non-unique, the only unique peptide identification is the most C-terminal [8-GETFSCMVDHEALPLAFTQKTIDR], and the unique peptide is identified in only fraction 5 in all three replicates.

Alternatively, these two peptides {1-2} may not be from a fragment of this protein, but perhaps from another protein altogether. It is also possible that these two peptides may be the result of degradation during sample handling, but if this was the case, one would expect to see some of these two peptides in the later fractions to indicate that the protein is also found in the non-degraded form, too. Since we see the fragment more than once and the trend is across all three replicates, this appears unlikely.

Whatever the cause, this protein is a candidate for further investigation regarding its possible endogenous structure and function.

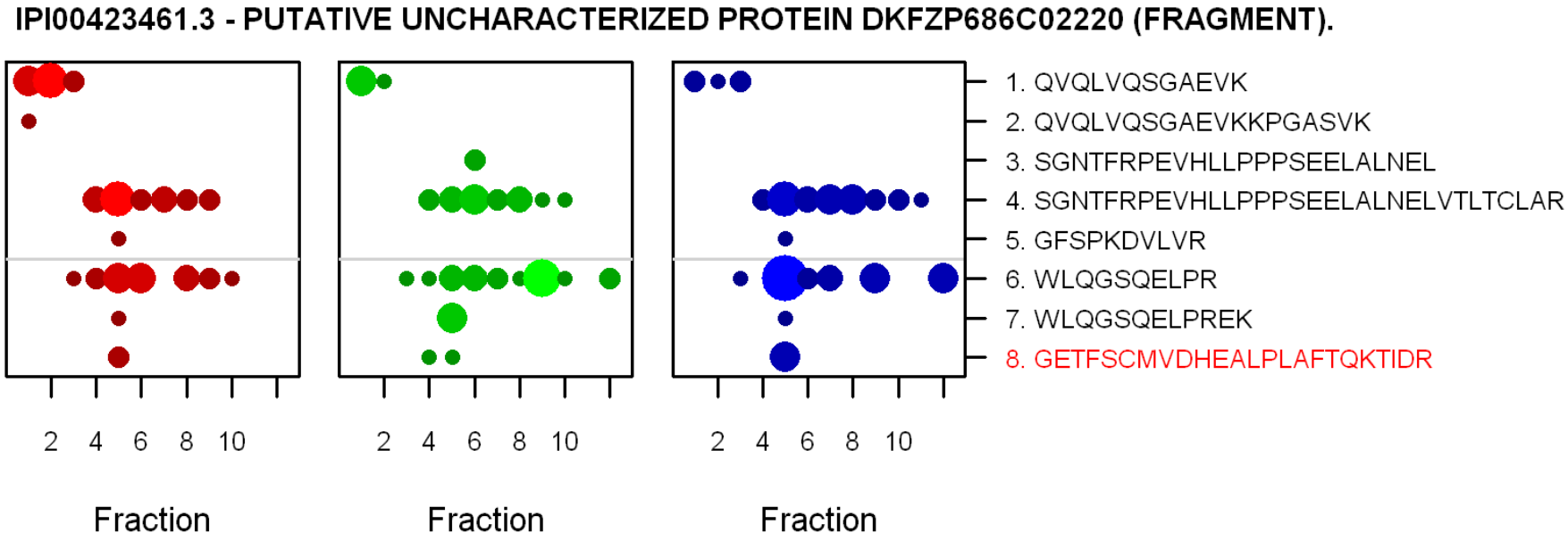


Figure 3.9 (A) IPI00423461.3 - Novel graphical representation of the PROOF data with replicates 1, 2 and 3 are shown from left to right respectively across the three graphs. The dots represent peptide identifications that were allocated in the assignment of the identification of the protein. The sizes of the dots are directly proportional to the number of peptide identifications for that particular peptide in that replicate. The x-axis depicts the pooled fractions of 1 to 12 for where the peptides were identified. The y-axis depicts all the peptide sequences used in the identification of the protein, from N- to C-terminus, top to bottom respectively. The colour of the written peptide sequence on the y-axis is either red (unique) or black (non-unique). It can be inferred from this graph that this protein is endogenously present in a truncated form.

3.4.5.c Specific PROOF example number 2 - Protein IPI00298828.3 (both)

Figure 3.9 (B) represents an example of what looks like to be a protein in either two distinct conformational forms or associations. The majority of peptides used to identify the protein are found around fractions 1-3, then the density of the peptide identifications drop off dramatically and show a strong reappearance in fraction 12, with several additional C-terminal peptides associated to the protein. All the peptides in this identification are unique (red), with no non-unique (black) peptides identified, which lend weight to the point that the entire protein is found in two forms under these conditions.

The protein IPI00298828.3 is apolipoprotein H (Beta-2-glycoprotein 1). Apolipoprotein H has been implicated in many metabolic pathways, including lipoprotein metabolism, coagulation, the production of antiphospholipid autoantibodies, and angiogenesis regulation. Since the protein is known to have a multitude of biological functions, finding it in more than one form is not unexpected. Taking into consideration that it was found eluting off the anionic column as well as the cationic would suggest a major change in form and function, or due to an association with another entity, rather than a cleavage event, since the length of the peptide sequence information is similar in both areas of the graph. Thus, it has been identified here as a candidate of interest for further investigation as to any potential biological function that can be elucidated under these conditions, and a strong example of how this chromatography can illuminate additional information beyond standard proteomic identification from peptides alone.

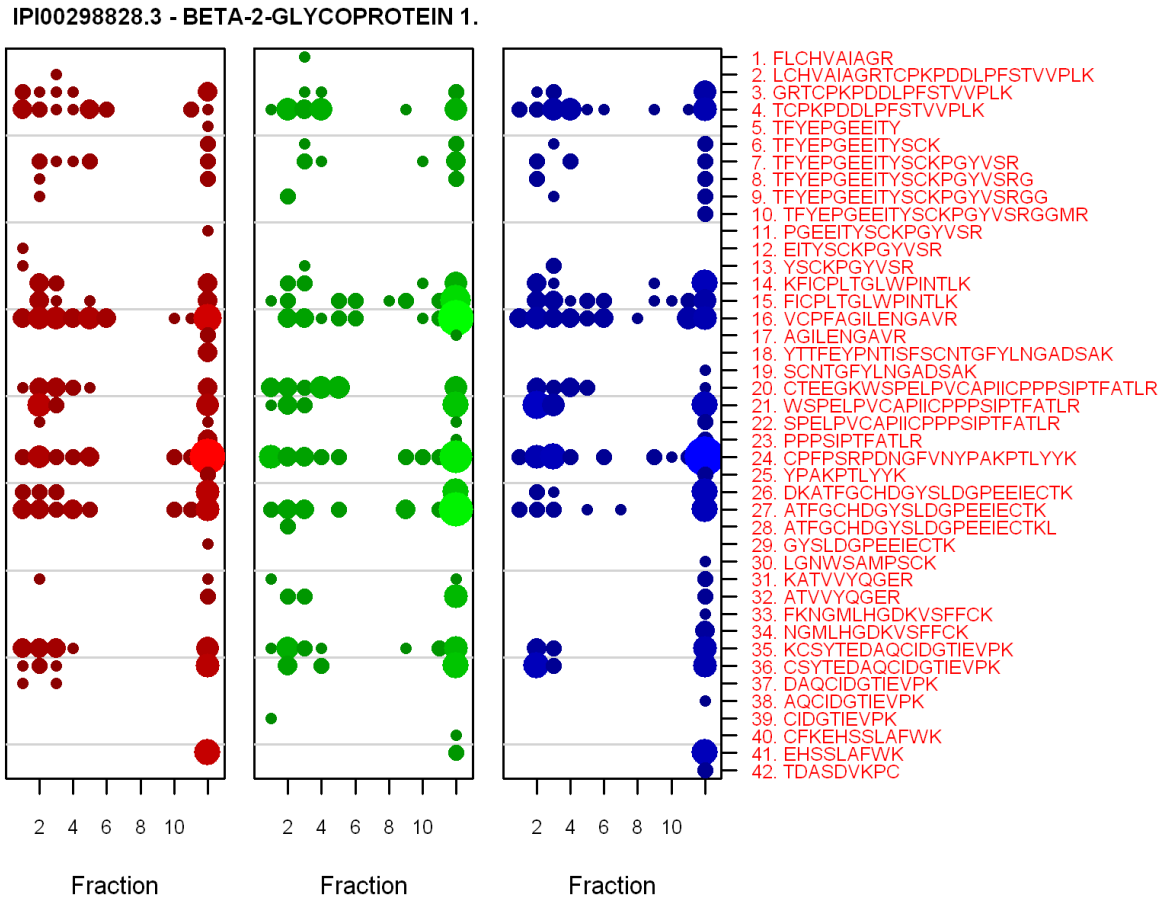


Figure 3.9 (B) IPI00298828.3 - Novel graphical representation of the PROOF data with replicates 1, 2 and 3 are shown from left to right respectively across the three graphs. The dots represent peptide identifications that were allocated in the assignment of the identification of the protein. The sizes of the dots are directly proportional to the number of peptide identifications for that particular peptide in that replicate. The x-axis depicts the pooled fractions of 1 to 12 for where the peptides were identified. The y-axis depicts all the peptide sequences used in the identification of the protein, from N- to C-terminus, top to bottom respectively. The colour of the written peptide sequence on the y-axis is either red or black, representing the unique (red) and non-unique (black) peptide classification generated. It can be inferred from this graph that this protein is endogenously present in at least two distinct conformational forms, representing two possible functionalities for the one identification.

3.4.5.d Specific PROOF example number 3 - Protein IPI00218192.3 (unique)

The protein IPI00218192.3 is inter-alpha (globulin) inhibitor H4 (plasma Kallikrein-sensitive glycoprotein). The protein encoded by this gene is secreted into the blood, where it is cleaved by plasma kallikrein into two smaller forms. This gene is part of a cluster of similar genes on chromosome 3. Two transcript variants encoding different isoforms have been found for this gene. All the peptides in this identification are unique (red), with no non-unique (black) peptides identified.

Figure 3.9 (C) represents an example of what looks like to be a protein cleaved into two fragments at peptide {47 - QLGLPCPPDVPHAA}, due to the clustering of the N-terminal peptides around fractions 5 - 7 and the C-terminal peptides around fraction 2 – 4, centred around fractions 6 and 3, respectively. Upon closer inspection of the first four fractions there are no N-terminal peptides, apart from two single outliers within the first four fractions. However, the C-terminal peptides show a noticeable though reduced presence around the fractions 5 and 6 apart from a few outliers of two peptides further along the elution gradient. This leads me to hypothesise that the protein is found in potentially three forms composed of two major forms (higher concentration) and one minor form (lower concentration). This is only a qualitative estimate regarding concentration, as insufficient information is available to allow any quantitation to be performed. The first two forms are likely to be truncated variants of the original complete protein sequence, with the larger molecular weight unit being the N-terminal fragment and the smaller molecular weight unit being the C-terminal fragment. The minor form seems to be the complete protein sequence not truncated, which seems to be in a lower concentration because the trailing effect is reduced and the sizes of the peptide identification dots are smaller. However, it is also possible that there are only two forms of the protein, that being the two truncated versions of the sequence mentioned above, except that the C-terminal fraction is in either a higher concentration compared to the N-terminal fragment or the peptides are more easily digested or detected by the mass spectrometer to give the elongated trailing effect. Although the visual interpretation of the data is subjective, this protein is definitely a candidate for further investigation based on the unique chromatography observed and the clear cleavage of the protein.

The known structure of this protein as mentioned above includes two forms, and what we are seeing chromatographically is similar. This supports the argument that the PROOF method can isolate proteins and identify truncated forms of a single protein, which is beyond the standard proteomics identification normally presented today from bottom-up methods. This also lends weight to the other interpretations we are making for the identification of interesting candidates for further experimentation based on the unique chromatography of the proteins shown in this type of graphical representation.

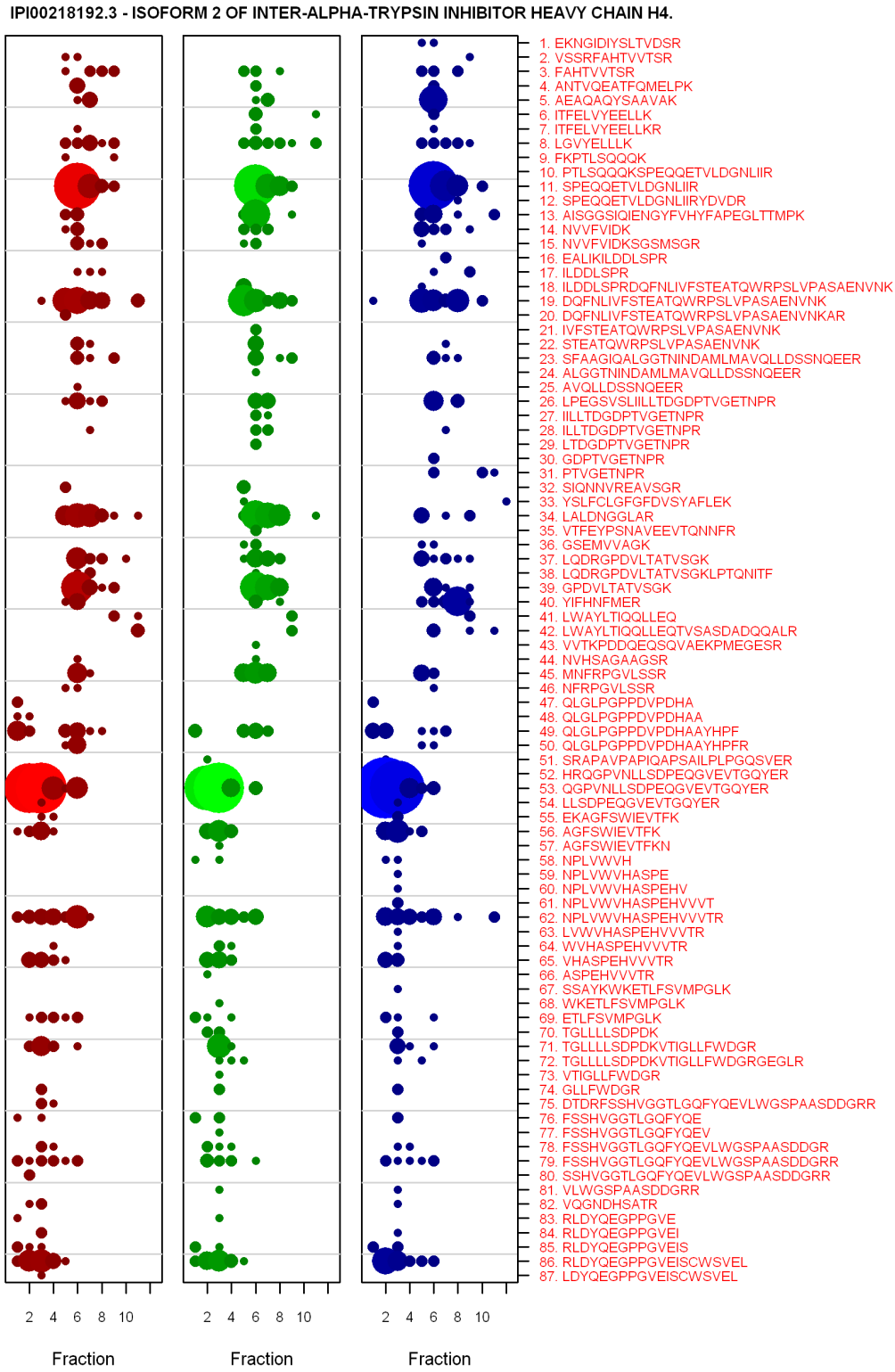


Figure 3.9 (C) IPI00218192.3 - Novel graphical representation of the PROOF data with replicates 1, 2 and 3 are shown from left to right respectively across the three graphs. The dots represent peptide identifications that were allocated in the assignment of the identification of the protein. The sizes of the dots are directly proportional to the number of peptide identifications for that particular peptide in that replicate. The x-axis depicts the pooled fractions of 1 to 12 for where the peptides were identified. The y-axis depicts all the peptide sequences used in the identification of the protein, from N- to C-terminus, top to bottom respectively. The colour of the written peptide sequence on the y-axis is either red or black, representing the unique (red) and non-unique (black) peptide classification generated. It can be inferred from this graph that this protein is endogenously present in at least two forms, one truncated around peptide #47, the other in the intact form, representing three proteins as opposed to one identification.

3.4.5.e Specific PROOF example number 4 - Protein IPI00019591.2 (both)

Figure 3.9 (D) represents an example of what appears to be a protein that has been cleaved at the N-terminus creating two protein species. The first fragment is a smaller molecular weight species that elutes around fractions 2 – 3 and the larger molecular weight species from the C-terminus elutes around fractions 4 – 5.

The protein IPI00019591.2 cDNA FLJ55673 is highly similar to complement factor B. This gene encodes complement factor B, a component of the alternative pathway of complement activation. Factor B circulates in the blood as a single chain polypeptide. Upon activation of the alternative pathway, it is cleaved by complement factor D yielding the non-catalytic chain B-a and the catalytic subunit B-b.

This protein and the resultant graphical representation is a strong example of how the non-unique (black) peptides and unique (red) classification can help with the interpretation of the data. If one was to draw attention to what looks like the clustering of unique and non-unique peptide allocations on the y-axis, it is evidence of conserved regions (black) of peptide sequences that are within the dataset amongst other proteins. The N-terminal fragment shown is composed of entirely non-unique (black) peptides and the C-terminal fragment begins with the unique (red) peptides, representing a clear line of distinction. My interpretation of this data is that there is one large molecular weight species from the C-terminus around peptide {24 – TPWSLARPQGSCSLEGVEIK} and the apparent smaller N-terminal fragment {peptides 1 – 23} is not a fragment, but it is from another protein within the dataset because of the entirety of continuous non-unique peptides clumped together.

In order to support this idea I searched for another protein that also had the non-unique (black) peptides that ranged from the sequence data above from {1 – TYSCPQGLPSPASR} to {23 – DHENELLNKQSVPAHFVALNGSKLNINLK}, with a similar peptide elution patter, this was found to be protein IPI 00303963.1 as shown in Figure 3.9 (E).

3.4.5.f Specific PROOF example number 5 - Protein IPI 00303963.1 (both)

Figure 3.9 (E) shows protein IPI00303963.1 Complement C2 fragment protein, showing peptide sequences eluted across fractions 2 – 4 and some into 5. Importantly it has seven peptides that are unique (red) and four of these peptides are at the N-terminal end of this C2 fragment protein.

This data leads me to believe that cDNA FLJ55673 (IPI00019591.2) protein in Figure 3.9 (D) which is highly similar to Complement B, is in fact cleaved at peptides {23 – DHENELLNKQSVPAHFVALNGSKLNINLK} and {24 – TPWSLARPQGSCSLEGVEIK} and is endogenously found as only the C-terminal fragment shown in Figure 3.9 (D) that elutes around fractions 4 – 5 and not the complete sequence that the standard bottom-up proteomics identification from peptides would have one believe.

Finally, what was first thought to be the presence of a smaller N-terminal molecular fragment of Complement B in Figure 3.9 (D) is not likely; rather it appears to be from the protein Complement C2 in Figure 3.9 (E) that elutes across 2 – 4 and not a separate fragment of Complement B.

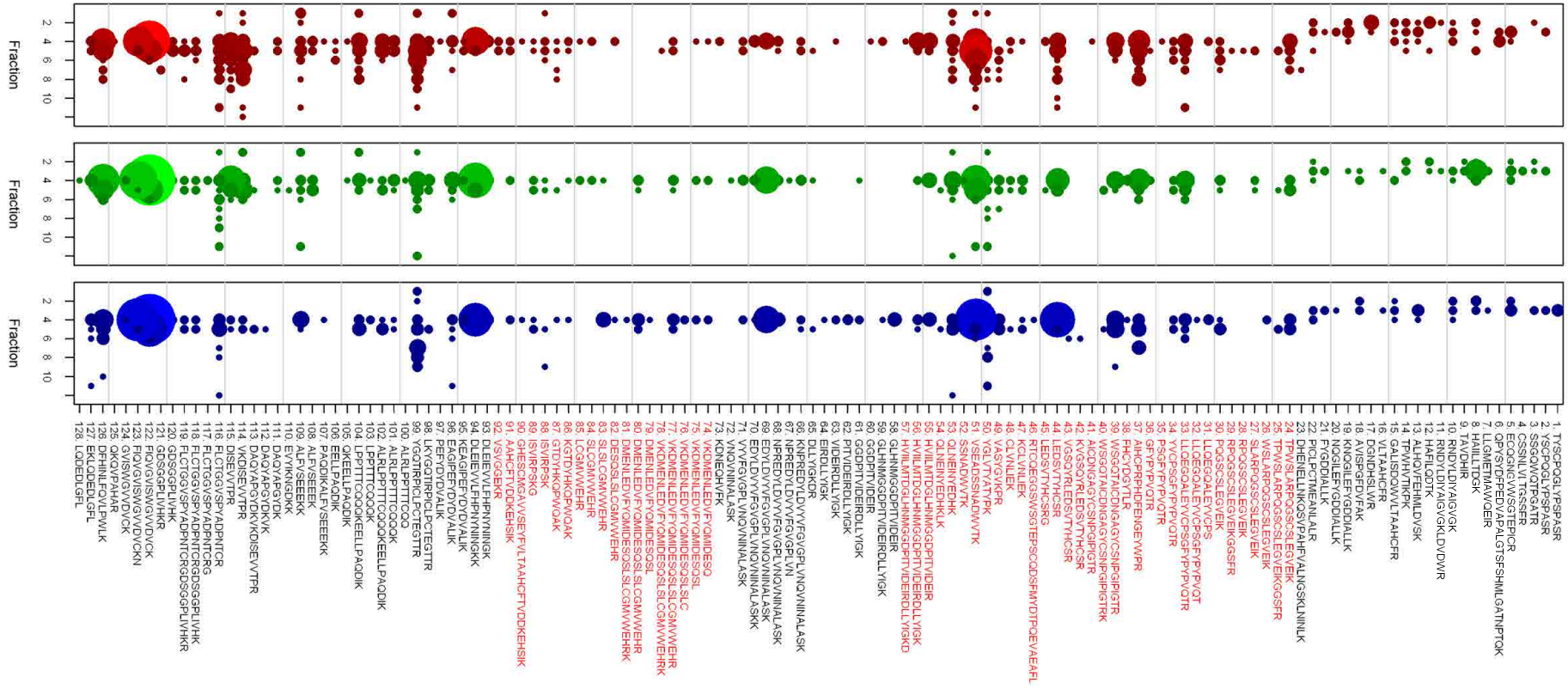


Figure 3.9 (D) IPI00019591.2 - Novel graphical representation of the PROOF data with replicates 1, 2 and 3 are shown from left to right respectively across the three graphs. The dots represent peptide identifications that were allocated in the assignment of the identification of the protein. The sizes of the dots are directly proportional to the number of peptide identifications for that particular peptide in that replicate. The x-axis depicts the pooled fractions of 1 to 12 for where the peptides were identified. The y-axis depicts all the peptide sequences used in the identification of the protein, from N- to C-terminus, top to bottom respectively. The colour of the written peptide sequence on the y-axis is either red or black, representing the unique (red) and non-unique (black) peptide classification generated. It can be inferred from this graph that this protein is cleaved at peptide # 23-24, creating two proteins from the one identification. However, when taking into consideration figure 3.9 (E), there is another possibility; that protein (D) is only one protein truncated at peptide # 23-24 to give the protein from peptide # 24-128. The fragment peptide # 1-23 could possibly by identified as protein (E).

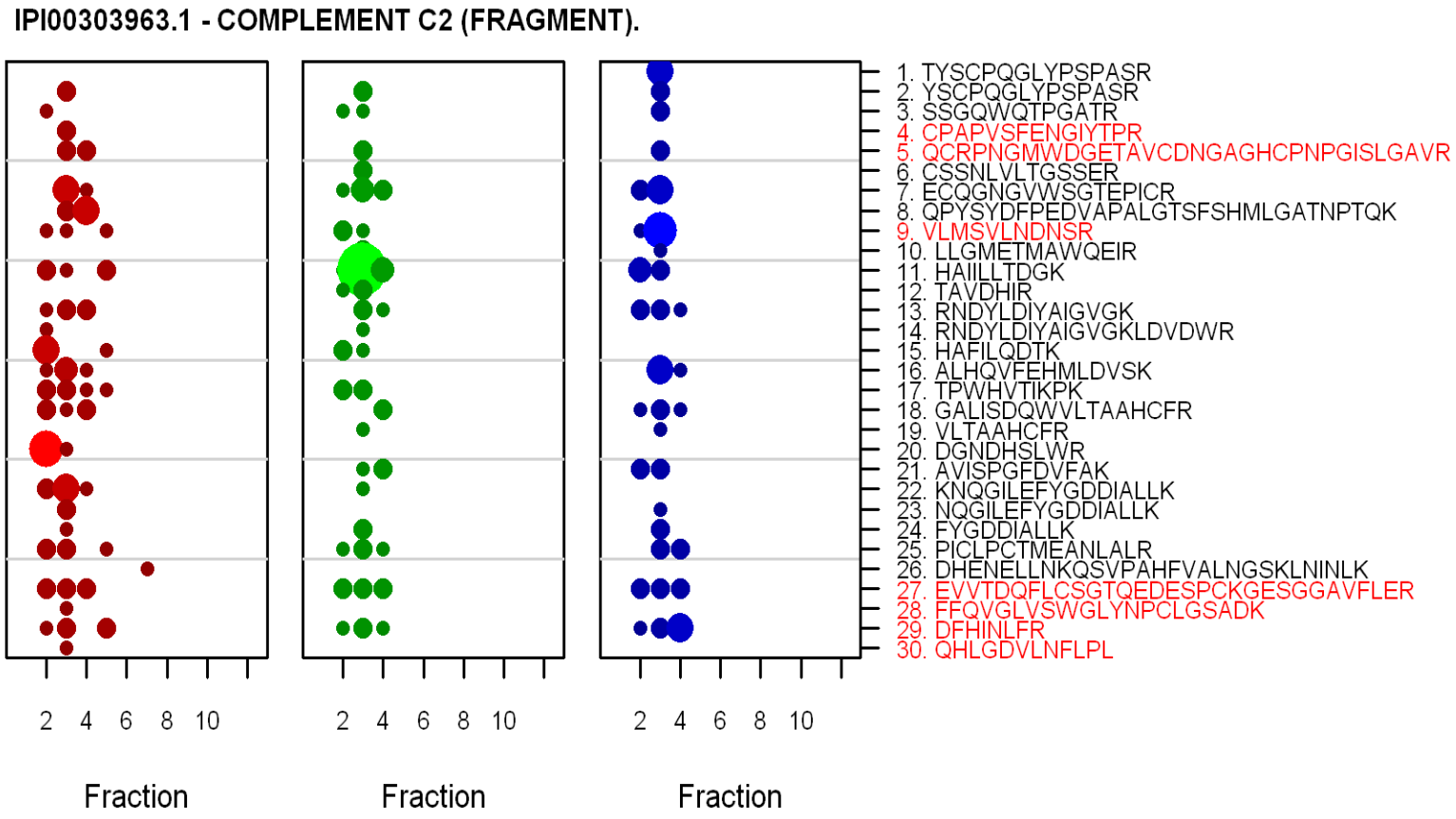


Figure 3.9 (E) IPI00303963.1 - Novel graphical representation of the PROOF data with replicates 1, 2 and 3 are shown from left to right respectively across the three graphs. The dots represent peptide identifications that were allocated in the assignment of the identification of the protein. The sizes of the dots are directly proportional to the number of peptide identifications for that particular peptide in that replicate. The x-axis depicts the pooled fractions of 1 to 12 for where the peptides were identified. The y-axis depicts all the peptide sequences used in the identification of the protein, from N- to C-terminus, top to bottom respectively. The colour of the written peptide sequence on the y-axis is either red or black, representing the unique (red) and non-unique (black) peptide classification generated. This graph supports the possibility that the N-terminal of protein (D) is not that, but it is protein (E) above.

3.5 FUTURE VIEWS AND DIRECTIONS

The information generated from these PROOF graphs is highly subjective and qualitative and would need to be developed further, both at the physical method development level and interpretative graphics areas of the methodology. However, it is clear that information from the literature for some of the proteins discussed above agree with the results of our experiments. This reinforces the notion that proteomic identifications based purely on peptides alone can lead to a misrepresentation of the actual proteome. The employment of protein fractionation under mild conditions with enhanced graphical representation to relate the peptides back to the intact mass of the proteins can generate information about the multiple tertiary and quaternary structural elements or any protein-protein interactions.

The system outlined here is not limited to Ion Exchange Chromatography (IEX), for I could incorporate Hydrophobic Interaction Chromatography (HIC), or Affinity Chromatography, with or without the IEX for a multi-Dimensional chromatographic fractionation system for proteins if desired. It would be useful to collect an increased number of fractions for the pooling of the SAX column, rather than the single fraction that was done here, to see if there are any kind of chromatographic resolution of species not visible to the UV detector but distinguishable by the mass spectrometer. Alternatively, resolution could be improved by testing minor changes in the physiological pH range, to see if there are more optimal pH ranges for enhanced resolution in the SAX and or the SCX. Due to resource constraints it was decided not to undertake this enhanced analysis at this stage. Additionally, I would like to elongate the gradient to try and enhance the chromatographic resolution, especially for the SCX column, and then take more fractions of smaller pooled amounts for MS/MS analysis, to see if any more functional, or structural information can be obtained. Lastly, it would be interesting to examine, in addition, non-depleted plasma and the flow through from the GenWay column and use the differences in the chromatographic peaks in the ÄKTÄ system to identify and quantify the sub-proteome of the albumin associated entities [397].

The PROOF approach has been applied to a complex proteome of depleted human plasma, generating comparable information when viewed from a standard proteomic

view of the total number of identifications of a proteome. Though, when comparing the techniques from a non-standard view integrating the novel graphical representations, PROOF in many respects generates potentially more tangible and unique information compared to the MudPIT analysis of peptides alone. The protein inference problem remains one of the largest problems within the field of proteomics today, both from an identification and quantification perspective. The use of protein fractionation by liquid chromatography should help with solving a piece of the protein inference puzzle, this being the more correct assignment of peptide identifications to the correct protein endogenous form. Further development within the field of protein fractionation and associated bioinformatics should enhance the information being generated by the proteomics field. This will aid in the evolution of proteomics from generating lists of identifications, to a field that offers insight into the tertiary and quaternary structure, protein-protein interactions, the truncated bioactive variants of proteins, and the endogenous biological activity and function [398].

Traditionally, biomarker development and the field of proteomics has tried to come up with ways in which to identify proteins at lower and lower concentrations, in the hope that a diagnostic tool can be produced that will identify specific proteins earlier in the disease state [275]. The ‘early detection of disease’ mentality is well established in the medical field today and has been shown to have positive effects in disease management [399]. Though, the idea that early detection will come only from identifying proteins at lower concentrations is perhaps flawed and is possibly why there is still not one stand alone diagnostic marker from standard proteomic identification of proteins from peptides at low concentrations [298]. The work I have presented here has the potential to help scientific investigators, medical diagnostic professionals or disease management teams, in a tangible way by identifying known proteins in their true endogenous forms.

The medical relevance of a potential ideological shift in proteomics can be shown by an ideological shift in the treatment of cancer that was historically thought either not possible or not relevant. This pertains to the identification of known protein forms, at known and detectable limits, and then the ability to distinguish between these entities as being cleaved or conjugated with direct application towards cancer therapy, rather than trying to drill deeper into the proteome to find unfathomable proteins at miniscule concentrations. Professor Judah Folkman, who pioneered the field of

angiogenesis, has postulated, amongst many things during his time, that if we could control the angiogenesis process with antiangiogenic therapy one could not only reduce the size of cancerous growths (tumours), one can turn an acute condition into a chronic and manageable condition [400-402]. Rather than trying to kill a tumour completely, as is the majority of the focus within the medical field today, one would try to identify and manage the disease situation and induce a type of controlled apoptosis of the tumour to a manageable size creating a chronic condition [400-402]. O'Reilly *et al.* showed in a Science publication in 1999, that when antithrombin, which is a member of the serpin family and functions as an inhibitor of thrombin and other enzymes, is cleaved at the C-terminal loop, that this conformational change induces it to have potent antiangiogenic and antitumour activity in mouse models [400]. The authors postulated that the presence of peptide inhibitors of angiogenesis within well-known proteins such as plasminogen, thrombospondin, platelet factor 4, kininogen, prothrombin and antithrombin, may not only offer precise regulation of angiogenesis at sites of microvascular injury, they may offer potential for improved efficacy and diminished toxicity in the treatment of cancer [400]. I have identified four out of these six proteins here in these experiments and I have shown how I can identify truncated or cleaved elements, which is at the heart of the antiangiogenic process. Hence, I postulate that this technique could play a role in the future development of tools for the management of such disease states.

A major reason for the current state of proteomic analysis being done on peptides and not on proteins can be attributed to protein fractionation traditionally being seen to be too difficult, identifying less and dissimilar proteins to fractionation of peptides. I believe I have shown here that this is no longer a justification for not doing protein fractionation of complex systems with liquid chromatography. It has been shown by others since this work was first conducted that top-down proteomics and protein fractionation is compatible with chromatographic time scales, and detecting high mass proteins and isoforms justifies further development in this field [134, 201, 389, 403].

Despite the aforementioned changes to the chromatography that could be implemented and optimised on the current system, I believe there would be benefits in the implementation of an appropriate mass spectrometer linked to the fractionation system to conduct top-down proteomics. Thus, having both a bottom-up and top-

down analysis combined with the graphical representation data shown or developed further into a two-dimensional map, somewhat analogous to the visualisation of a two-dimensional gel.

In addition to protein fractionation, further developments within the analytical field of mass spectrometry so as to conduct the analysis of the proteome at the protein level rather than the peptide level, also will play a major role in solving the current protein inference problem [280]. Not every investigator can use an FT-ICR mass spectrometer for top-down proteomics, although developments within this field are making top-down proteomics easier and more accessible. This strengthens the need for protein fractionation with liquid chromatography, so as to further enhance the quality and validity of the data generated by the proteomics field as a whole.

Casado-Vela *et. al.* stated in a review in 2011 on proteomics that the identification and characterisation of protein species using these (proteomics) technologies are still fraught with limitations and require improvement [184]. Thus, it is my belief that until the field of proteomics works principally at the protein level for both the fractionation (sample preparation) and analysis (mass spectrometry) of the biological samples, there will always be information holes and certainty questions. The field will be inadvertently obscuring the accurate understanding of the proteome, and the problems in perception of the field will remain. It is viewed by many to be an area of great promise, though it is only to be considered as a first stage and is not considered suitable to answer biological questions of a complex nature. Given sufficient development it (proteomics) has the potential to become a stand alone scientific method in its own right, able to answer the most difficult complex biological questions.

CHAPTER 4: QUANTITATIVE PROTEOMICS OF TEMPERATURE STRESS IN RICE LEAF

4.1 ABSTRACT

It is likely that in the near future we, as a society, will not be able to produce enough rice to satisfy the increasing demand without modification in current agricultural practices and generation of fresh cultivars before the end of the 21st century. This pressure is increased by population growth, reduced land availability, and variability of environmental conditions due to climate change. A detailed understanding of the molecular responses to thermal stresses is essential for developing cultivars that can adapt or have a greater resistance to low or high temperature stresses. This study investigated the proteomics response of rice seedlings under cold stress of 12-14°C after 48, 72 and 96 h exposure. Label-free and iTRAQ quantitative approaches were both employed to assess the global proteome expression and also the complementarity of the two methodologies for use in plant proteomics. Both approaches delivered similar insights into the biological response to cold stress despite the disparity in proteins identified. The label-free approach identified 236 cold-responsive proteins compared to 85 from iTRAQ, with only 24 in common. Functional analysis revealed differential expression of proteins involved in transport, photosynthesis, generation of precursor metabolites and energy. Intriguingly, histones and vitamin B biosynthetic proteins were observed to be affected by cold stress in rice leaf.

4.2 PREAMBLE

Partial content of this chapter was published in;

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

In order to try and keep some form of continuity with respect to my project theme, that being the development or enhancement of emerging proteomics techniques, my supervisor and I decided to undertake a series of quantitative proteomics experiments to ascertain a best practice for his research group and the greater scientific community at large. He is a renowned expert in plant proteomics, and a leader within the field of shotgun proteomics applied to flora. Thus, we decided to work on a model system that is widely used within his group, that being rice and the response of the proteome to abiotic stress. His group was already conducting label-free spectral counting quantification proteomics and basing their analysis pipeline around this method. We wanted to make sure that we were not missing out on any potentially valuable information by not using iTRAQ or similar labelling techniques, and to see if there was any justification in the added cost of running iTRAQ compared to spectral counting.

Karlie Neilson and I decided to work together for the data presented within the chapter, due primarily to her plant growth experience and my proteomic analytical experience. Thus, Karlie conducted most of the plant growth, temperature stressing and harvesting of the leaves and I conducted the extraction and purification of the proteins from the leaf material and the differential sample preparation for iTRAQ and label-free. APAF staff and I conducted the mass spectrometry for iTRAQ together, while I conducted the mass spectrometry for the label-free myself. Karlie and I conducted the data analysis collaboratively, with credit going to Karlie for manually interpolating and identifying the histone and vitamin B biosynthetic expression relevance.

At the time of these experiments in 2007 there had only been two reported works conducted on plants with iTRAQ, with none on rice [404, 405] and the majority of techniques employed within were considered anecdotal in a review by Jorrin *et al.*

[406] even though the iTRAQ method was considered an established leader in the field of multiplexed isotope labelling quantification in non-plant systems with numerous citations since its emergence in 2004 [227, 407]. My major obstacle was to try and develop a method so that the rice leaf proteins extracted were clean of compounds that would potentially react with the iTRAQ label reagents, which if present would result in reduced (or no) conjugation of the iTRAQ to the rice peptides and hence a failure in quantification.

At the same time of these experiments another competing isotope labelling method was just emerging called EXACTAG™ (PerkinElmer Life and Analytical Science, USA) [408]. We decided to use this as an alternative isotope labelling system in our comparison, too, primarily because Exact-Tag did not require a high resolution instrument like iTRAQ, meaning that we could analyse the samples on our LTQ-XL ion trap as we did with the label-free experiments. This was of great interest to us since we could use the same instrument for the comparison, removing the instrument variability for the comparison, as well as not having to manage the usual extra cost and time constraints associated with outsourcing any mass spectrometry analysis.

Briefly, the Exact-Tag data could not be analysed because, at the time we were ready to conduct the computational analysis, the product was discontinued and no longer supported so we had to terminate this aspect of the comparison. However, it did seem from the initial visual inspection of the mass spectra at the time of running the samples and on first inspection of the data that the labelling system had conjugated to the sample, though to what extent and at what quantitative significance was not able to be addressed. Thus, no data is presented within this chapter.

Based on my earlier work in Chapter 3 on protein fractionation, we considered the idea of implementing a protein pre-fractionation step with liquid chromatography so as to try and apply the method to a biological system undergoing stress and known change. We finally decided to be conservative in our method development aspirations for these sets of experiments and did not undertake any protein pre-fractionation with liquid chromatography for this comparison. Although my method development was conducted with the potential use of protein fractionation in mind for future experiments, hence the omission of the ubiquitous acetone precipitation step, as one example to try and maintain protein structure and function in the sample. It is

interesting to note that these experiments highlighted a potential justification for the implementation of this type of technology, discussed within for proteins PDX1.1 and PDX 1.2. It was finally decided that the best course of action was to implement minimal changes and apply it to a relevant biological system, abiotic stress in rice, thus bringing together methods from established techniques and modifying them as necessary. The ultimate goal was making the iTRAQ chemistry work in plant tissue for a true comparison with label-free spectral counting and to see if any further biological insights would be forthcoming.

4.3 MATERIALS AND METHODS

4.3.1 Materials and Reagents

All chemicals and reagents were purchased from Sigma-Aldrich (St. Louis, Missouri, USA), unless otherwise stated.

4.3.2 Rice Growth and Leaf Sampling Conditions

Rice (*Oryza sativa* L. cv. Nipponbare) seedlings were germinated in soil and grown in trays in a temperature controlled growth cabinet with day/night cycle of 12 h/12 h and day/night temperatures of 28°C/12°C. Once the seedlings reached the 4-5 leaf stage, four of the five trays were placed into a separate incubator with temperature maintained at day/night 14°C/12°C. Both the control and the stressed plants were kept hydrated with the same level of water in water baths. The non-senescent mid-section of leaves were harvested from the plants daily at 9 am (3 h after light) at four time points: 0 (control), 48, 72, and 96 h after cold stress imposition. Leaf material was ground to a fine powder in a mortar and pestle under liquid nitrogen and prepared according to Breci *et al.* [409]. Briefly, 40 ml resuspension buffer (10% trichloroacetic acid, 0.07% (v/v) mercaptoethanol in acetone) was added to 4 g ground leaf material, mixed thoroughly, and placed at -20°C for 45 min. Resuspended leaf powder was centrifuged for 15 min at 35,000 x g. The supernatant was discarded and the pellet was washed with EDTA was solution (0.07% (v/v) mercaptoethanol, 2 mM EDTA, in acetone) three times or until the leaf tissue was no longer green. The leaf tissue protein pellet was lyophilised and stored at -20°C until needed.

4.3.3 Protein Extraction and Preparation

Soluble protein was extracted by adding 6 ml 100 mM Tris-HCl, pH 8.5 to 400 mg lyophilised leaf powder followed by 3 min incubation at 4°C with vortexing and centrifugation at 18,000 x g for 10 min. The supernatant was collected and insoluble

protein was extracted from the remaining leaf powder three times with 8 ml of 8 M urea. For each extraction, the sample was incubated with urea for 1 min at 4°C with vortexing, and the supernatant was collected after centrifugation at 18,000 x g for 10 min. The supernatants from all extraction steps were pooled and concentrated to 4 ml by centrifugation at 4000 rpm at 10°C using a 3 kDa molecular weight cut-off Amicon Ultra -15 Centrifugal Filter Device (Millipore, USA). The protein sample was buffer exchanged with ultrapure water, and total protein was quantified using the FluoroProfile Quantification Kit. Based on the quantification reading obtained, the protein extracts were aliquoted into three equal amounts of 100 µg for each of the four time points: 0, 48, 72 and 96 h. The solution was lyophilised and placed at -20°C until needed for labelling (iTRAQ) or label free digestion.

4.3.4 Label-free Tryptic Digestion

Protein extracts (100 µg each) were dissolved in 40 µl of 6 M urea in 50 mM NH_4HCO_3 (pH 8) and heated to 95°C for 5 min. The samples were reduced with 1 µl of 0.5 mM DTT in 50 mM NH_4HCO_3 (pH 8) and incubated for 60 min at 37°C before alkylating in the dark with 2 µl of 200 mM iodoacetamide in 50 mM NH_4HCO_3 (pH8) for 60 min at room temperature. Urea was diluted by adding 360 µl of 50 mM NH_4HCO_3 and incubated overnight at 37°C with 2 µg trypsin (Promega, USA). The reaction was quenched with the addition of 100 µl of 1% (v/v) formic acid. Samples were dried under vacuum and reconstituted in 0.1% (v/v) formic acid and 2% (v/v) acetonitrile (ACN). Peptides were stored at -20°C until required for SCX fractionation.

4.3.5 iTRAQ Labeling and Tryptic Digestion

Protein extracts (100 µg each) were resuspended in 200 ml of 0.25 M triethylammonium bicarbonate (TEAB), pH 8.5, 0.1% SDS, reduced with 1 mM Tris (2-carboxyethyl)-phosphine (60°C, 1 h) and alkylated with 1 mM methyl methane thiosulfate (room temperature, 10 min). Proteins were digested overnight at 37°C with 5 µg trypsin (Promega, USA). The samples were dried down the following day in a vacuum centrifuge and reconstituted with 30 µl of 500 mM TEAB. Each sample was labelled using a 4-plex iTRAQ kit as per the manufactures instructions (Applied

Biosystems, USA). iTRAQ reagent labels (114, 115, 116 and 117) were resuspended in a final concentration of 70% (v/v) ethanol, added to the respective samples (0. 48, 72, and 96 h), and incubated at room temperature for 1 h. The reaction was quenched by adding 100 µl of ultrapure water to each sample. The iTRAQ labelled samples were mixed in equal ratios, dried in a vacuum centrifuge and stored at -20°C. MALDI-MS was conducted on a subset of the sample to confirm successful conjugation of the label to the peptide mixture.

4.3.6 Strong Cation Exchange (SCX) Chromatography

Identical methods for SCX chromatography were employed for the label-free and iTRAQ labelled peptide samples using an Agilent 1100 quaternary HPLC pump with a PolyLC polysulfoethyl aspartamide column (200 mm x 2.1 mm, 5 µm, 200 Å). The column was equilibrated with 5 mM sodium phosphate with 25% (v/v) ACN, pH 2.7, which was also used for sample resuspension, sample injection, and peptide adsorption to the column. Peptide elution was achieved with a 70 min gradient to 100% 5 mM sodium phosphate with 25% (v/v) ACN, 350 mM KCl, pH 2.7, at a flow rate of 300 µl/min. Peptides were collected into 12 or 17 fractions for label-free and iTRAQ experiments, respectively, dried in a vacuum centrifuge, and stored at -20°C until required for LC-MS/MS.

4.3.7 Label-free nanoflow LC-MS/MS

Each of the 12 reconstituted fractions of triplicate sets of four different time points were analysed by nanoflow LC-MS/MS using a LTQ-XL linear ion trap mass spectrometer (Thermo, USA). Each of the peptide extracts were desalted using C18 tips (Omix, Varian, Inc., CA) and the eluate was dried using a vacuum centrifuge followed by resuspension in 0.1% (v/v) formic acid. The methods of nanoflow LC-MS/MS, previously followed by Chick *et al.* [243], were adopted. In a fused silica capillary with an integrated electrospray tip, reversed-phase columns were packed in-house to approximately 7 cm (100 µm id) using 100 Å, 5 mm Zorbax C18 resin (Agilent Technologies, USA). An electrospray voltage of 1.8 kV was applied via a liquid junction upstream of the C18 column. Samples were injected onto the column

using a Surveyor Autosampler, which was followed by an initial was step with buffer A (5% ACN, 0.1% formic acid) for 10 min at 1 µl/min. Then, peptides were eluted from the column with 0-50% buffer B (95% ACN, 0.1% formic acid) for 58 min at 500 nl/min. The column eluate was directed into a nanospray ionization source of the mass spectrometer. Spectra in positive ion mode were scanned over the range of 400-1500 amu, and using Xcalibur software (Version 2.06, Thermo, USA), automated peak recognition, dynamic exclusion and MS/MS of the top six most intense precursor ions at 35% normalization collisions energy were performed.

4.3.8 iTRAQ nanoflow LC-MS/MS

The 17 SCX fractions were resuspended in 60 µl 0.1% (v/v) trifluoroacetic acid (TFA), 2% (v/v) ACN and peptides were loaded onto the C18 column with 0.1% (v/v) formic acid, 2% (v/v) ACN. Peptides were eluted with three linear gradient steps increased from 5-95% (v/v) ACN of 95 min. Mass spectra were acquired on a QStar XL hybrid quadrupole time-of-flight (TOF) mass spectrometer (Applied Biosystems, USA). The reverse phase nanoLC eluent was subjected to positive ion nanoflow electrospray analysis in an information dependent acquisition mode. The TOF MS survey scan spectra from m/z 380-1600 were acquired for each fraction every 0.5 s, with the three most intense multiply charged ion (counts >50) sequentially subjected to MS/MS analysis. MS/MS spectra were accumulated for 2 s in the mass range m/z 100-1600 with a modified Enhanced All Q2 transition setting favouring low mass ions so that the reporting iTRAQ tag ion (114, 115, 116 and 117 m/z) intensities were enhanced for quantification.

4.3.9 Label-free Database Searching for Protein Identification

The LTQ-XL raw output files were converted into mzXML format and searched against the NCBI *O. sativa* Reference Sequence (RefSeq) database (26,938 proteins, December 2008), using the global proteome machine (GPM) software (version 2.11) and the X!Tandem algorithm. The 12 fractions of each replicate were processed sequentially with output files generated for each fraction, and a merged, non-redundant output file for protein identification with log(e) values < -1 was produced. Peptide identification was determined using 0.4 Da fragment mass error.

Carbamidomethyl was considered as a complete modification, and partial modifications considered included oxidation of methionine and tryptophan. Reverse database searching was used for estimating false discovery rates (FDRs). Complete protein and peptide identification information, including spectra (converted to PRIDE-xml using the PRIDE converter tool) [410], is available from the PRIDE database (www.ebi.ac.uk/Pride) using accession numbers 15854-15865 (spectra) and 15873-15884 (protein identifications) and in the supplemental data 4.1 section of this thesis.

4.3.10 Label-free Data Processing and Quantification

The 12 lists of proteins identified, representing the three replicates at each of the four time points were filtered based on two criteria: (1) a protein was retained if it was present reproducibly in all replicates of at least one time point and (2) had a total spectral count of ≥ 4. Peptide and protein FDRs were subsequently calculated. Peptide FDR was calculated by 2 x (total number of peptides identified for reversed protein hits/total number of peptides identified for all proteins in the list) x 100 and protein FDR was calculated by (number of reversed protein hits/total number of proteins in the list) x 100[386]. NSAF for each protein was calculated according to Zybailov *et al.* [240]. Briefly, the NSAF for a protein is given by the number of spectral counts (SpC, the total number of MS/MS spectra) identifying a protein's length (L), divided by the sum of SpC/L for all proteins in the experiment. A fraction of a spectral count (0.2 on average) was added to all counts to compensate for null values [240].

4.3.11 Statistical Analysis of Label-free Differentially Expressed Proteins

The natural log NSAF data were examined by generating overlapped kernel density plots for all samples, to assure that the data for each sample were normally distributed. A two-sample student's *t*-test was performed to determine statistical significance in differential abundance between proteins identified in the control (0 h) samples and proteins identified in the 48, 72 and 96 h cold stress time points. Proteins with *p*-values < 0.05 were selected as differentially expressed. All data processing, including the data normality assessments and *t*-tests, was carried out

using functionality from the stats package of the R statistical programming environment, (http://www.R-project.org) [411]

4.3.12 iTRAQ Data Analysis

The QStar XL output .wiff files were submitted in ProteinPilot (Applied Biosystems, v 3.0) for data processing, protein identification, and quantification using the Paragon algorithm [412]. Data were searched against the NCBI *O. sativa* RefSeq database, as for label-free experiments, but for consistency, the database was modified to contain the common repository of adventitious proteins information as is used in the GPM. A reverse database was also searched to estimate the FDR. The following search parameters were selected: sample type (iTRAQ peptide labelled), cys alkylation (MMTS), digestion (trypsin), instrument (QStar ESI), special factors (nil selected), ID focus (biological modifications), search effort (thorough ID), and specify processing (quantitation, background and bias corrections). The detected protein threshold (unused ProtScore) was set to > 1.3 (95% confidence or better) and a *p*-value < 0.05 ensured that protein identification and subsequent quantitation were not based on single peptide hits. Spectral information (converted to PRIDE-xml using the PRIDE converter tool) [410] is available from the PRIDE database (www.ebi.ac.uk/Pride) using accession numbers 15866-15871. Protein identification information from Protein Pilot was not compatible with PRIDE submission, but it is supplied in the supplemental data 4.1 section of this thesis.

4.3.13 Functional Classification from Gene Ontology Information

Gene Ontology (GO) information was used to categorise the biological processes of identified proteins from both the label-free and iTRAQ datasets. GO annotations were extracted from the Uni-Prot database and matched to corresponding gene locus identifiers embedded in the NCBI *O. sativa* RefSeq database. Proteins were then classified based on their biological process using Web Gen Ontology Annotation Plot (WEGO) (wego.genomics/org.cn/cgi-bin/wego/index/pl) [413].

4.4 RESULTS AND DISCUSSION

4.4.1 Label-free proteomic analysis

Samples were run in triplicate (technical replicates) for each of the time points, t = 0, 48, 72 and 96 h. Only proteins that were present in all three replicates with a total spectral count (SpC) ≥4 were retained for quantification. The low level of variability in the dataset (2.04-11.73% RSD) indicates a high degree of reproducibility across and between the replicates (Table 4.1). The false discovery rate (FDR) was calculated after combining the replicate data and was found to be <0.7% at the protein level and <0.4% at the peptide level, thus no further filtering of results was deemed necessary. The total number of unique proteins included in the dataset from all of the four time points was 1050. A summary of the number of proteins and peptides identified, along with the FDRs in the label-free experiment, is shown in Table 4.1.

The spectral counts (SpCs) of proteins from the label-free dataset were converted to normalised spectral abundance factors (NSAFs) to normalise for the amino acid sequence length of the proteins and total amount of protein in a replicate [241, 244]. Log NSAF values were checked for normal distribution before a students *t*-test was used to statistically assess abundance changes of proteins between the control and each cold stress time point. The list of differentially expressed proteins and *p*-values is included in supplemental data 4.1. A summarised version of the differential expression data for both the label-free and iTRAQ experiments are shown in Table 4.2.

For the label-free data, considering the control sample as a reference: after 48 h of cold stress, 64 proteins were found to be up-regulated and 36 down-regulated; after 72 h of cold stress, 77 proteins were found to be up-regulated and 39 down-regulated; and after 96 h of cold stress, 70 proteins were found to be up-regulated and 35 down regulated. In total, there were 236 proteins that showed differential expression (150 uniquely up-regulated and 88 uniquely down-regulated) in response to cold stress. Overall, there were 563 proteins found at the three time points that were unchanged by cold stress or were outside the acceptable range of the *t*-test for

significance, supplemental data 4.2 shows the proteins that were deemed to be unchanged by the label-free method.

Hours after cold stress	No. proteins common to the 3 replicates	No. peptides			Total no. peptides	Average no. peptides ±%RSD	Protein FDR (%)	Peptide FDR (%)
		R1	R2	R3				
0	689	13395	12605	11303	37303	12434.3±8.50	0.58	0.21
48	686	12834	12863	10403	36100	12033±11.73	0.29	0.18
72	750	15168	14556	13804	43528	14509±4.71	0.67	0.31
96	733	13056	12537	12758	38351	12784±2.04	0.27	0.21
Total	2858	54453	52561	48268	155282	51761±6.12	0.45	0.23

Table 4.1 Summary of the label-free results. The number of proteins identified and quantified across each time point and in triplicate range from 686 proteins for the 48 h time point and 750 proteins for the 72 h time point. All protein identifications had a protein FDR of <0.7% and an SpC of ≥4. Relative standard deviation is abbreviated to (RSD) and false discovery rate to (FDR).

	Label-free	iTRAQ
Total no. proteins (<1% FDR)	1050	1269
Total differentially expressed (n.r)	236	85
48 h up-regulated	64	15
48 h down-regulated	36	20
72 h up-regulated	77	12
72 h down-regulated	39	16
96 h up-regulated	70	18
96 h down-regulated	35	15
Up-regulated at all time points	16	1
Down-regulated at all time points	4	0
Unchanged common	563	436

Table 4.2 Summary of the label-free and iTRAQ results. The total number of proteins identified by iTRAQ was 1269 though only 521 could be used for quantification, which is less than the total number of proteins identified and quantified by label-free, 1050. Non-redundant is abbreviated to (n.r) and false discovery rate to (FDR).

Four proteins were found to be significantly down-regulated across all three time points, and 16 proteins were found to be significantly down-regulated across two time points. Sixteen proteins were found to be significantly up-regulated across all three time points, and 30 proteins were found to be significantly up-regulated across two time points. Two proteins were found to be both significantly down- and up-regulated: gi115489174, a putative pathogenesis-related protein, was found to be down-regulated after 72 h of stress but up-regulated at 48 and 96 h time points, and gi115439533, glycine dehydrogenase P protein, was found to be up-regulated after 72 h of stress and down-regulated after 96 h.

4.4.2 iTRAQ proteomic analysis

An iTRAQ experiment was performed using identical leaf material as for the label-free experiment. Overall, 1269 proteins were identified with a protein FDR of <1%. Of the 1269 proteins, 748 were not used for quantification, due to incomplete labelling or inconsistent detection of one or more isobaric tag from the mass spectra produced. The software associated with the iTRAQ method uses the Paragon algorithm for picking a “winner” protein out of several potential entries for the same protein [412]. In the *O. sativa* FASTA file there are many proteins that have the same name and very similar, but not identical, amino acid sequences. For example, Rubisco small chain is associated with the gene identifiers gi|115488234, gi|115488240, gi|115488238, and gi|115488144. Paragon allocated gi|115488234 as the “winner” protein, which was used for quantification, and any unused spectra were allocated to non-winner proteins. This selection process should not be confused with shared (non-unique) peptides of homologous proteins; the Paragon algorithm does not perform quantification on peptides that are shared between homologous sequences of different proteins.

Determination of a “threshold line”, or “point”, in the dataset for the classification of significance for down- and up-regulation is a debatable idea. The range fluctuates due to changing versions of the iTRAQ analysis software and ongoing debate regarding what an isobaric tag ratio number represents in terms of a fold change in protein expression. We used parameters proposed by Keshamouni *et al.* and Ali *et al.* [414, 415] which are based on two criteria: the protein must have a 20% fold

change compared to the control, i.e. an expression ratio of >1.20 or <0.83; and the expression value must have a *p*-value of <0.05. A summarised version of the differential expression data for both iTRAQ and label-free is presented in Table 4.2. Of the 1269 proteins identified only 521 proteins were quantified, 85 showed significant differential expression and 436 were considered unchanged; supplemental data 4.1 displays the full details of the iTRAQ results.

For the iTRAQ data, considering the control sample as a reference: after 48 h of cold stress, 15 proteins were found to be up-regulated and 20 down-regulated; after 72 h, 12 proteins were found to be up-regulated and 16 down-regulated; and after 96 h, 18 proteins were found to be up-regulated and 15 down-regulated. No proteins were found to be significantly down-regulated across the three time points, although 4 proteins were found to significantly down-regulated across any two time points. One protein was found to be significantly up-regulated across the three time points, and 4 proteins were found significantly up-regulated across any two time points. One protein was found to be significantly down- and up-regulated: gi|115477659, a putative RNA-binding protein, was found to be down-regulated at 48 h and up-regulated at 96 h time points after cold stress.

4.4.3 Similarities between Label-free and iTRAQ

Based on the gene identifier numbers for the proteins identified with a FDR of <1%, there was a high degree of overlap between label-free and iTRAQ datasets, with a total of 751 proteins identified by both techniques, representing 72 and 59% of proteins for the label-free and iTRAQ datasets respectively, as shown in Figure 4.1. We incorporated non-winner proteins from the iTRAQ data when determining the amount of overlap between the label-free and iTRAQ datasets. The large proportion of overlap between the datasets is consistent with a previous study [272] which also compared label-free and iTRAQ approaches. They found that when examining proteins identified by >1 peptide, a substantial amount of proteins identified by iTRAQ were also identified by the label-free method; 33% of the label-free data was shared with iTRAQ data but 79% of the proteins found by iTRAQ were also found by label-free.

There were 24 proteins that were found to be commonly differentially expressed in both the label-free and iTRAQ datasets, and these are displayed in Table 4.3, which represents approximately 10 and 28% of the label-free and iTRAQ differentially expressed datasets, respectively. There was a strong convergence between the expression patterns indicated by the two approaches, with 20 of the 24 proteins showing similar changes in expression. This was a proportionally lower overlap than identified in previous studies comparing label-free and labelling approaches [254, 272]. However, when a comparison is made between the differentially expressed proteins identified in one technique with all the proteins in the complementary approach, an intriguing result was revealed. All but 8 of the 85 differentially expressed proteins identified using iTRAQ were also identified reproducibly in the label-free approach. Five of the 8 proteins were identified using the label-free method in independent MS runs (but not in triplicates of one time point) and so were not included in the merged <1% FDR cut-off dataset. Conversely, when comparing proteins differentially expressed in the label-free data with the entire iTRAQ dataset, 71 proteins were found uniquely, and reproducibly, by the label-free approach.

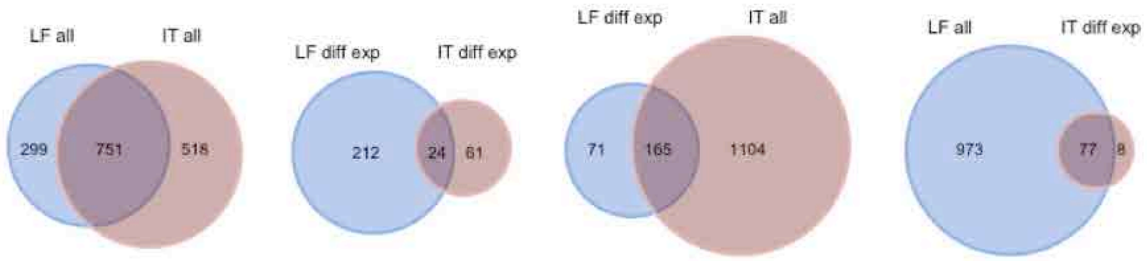


Figure 4.1 Venn diagrams depicting the amount of overlap between label-free and iTRAQ datasets. LF all and IT all indicate the total number of proteins identified at FDR <1% in label-free and iTRAQ respectively. LF differential expression and IT differential expression indicate the total number of non-redundant significantly differentially expressed proteins in label-free and iTRAQ datasets, respectively.

Protein name	Gene identifier	Method	48 h	72 h	96 h
Os07g0469100 protein	gi|115472001|	LF		↓	
		IT		↓	
Histone H3.2	gi|115484259|	LF			↓
		IT		↓	
Succinyl-CoA ligase [GDP-forming] subunit beta, Succinyl-CoA synthetase beta chain	gi|115447367|	LF			↓
		IT			↓
(S)-2-hydroxy-acid oxidase, glycolate oxidase, putative	gi|115470621|	LF			↓
		IT			↓
Os12g0189300 protein (92% homology with an O. sativa Carboxyvinyl-carboxyphosphonate phosphorylmutase, putative, expressed Q2QWN6)	gi|115487692|	LF			↓
		IT	↑		
Harpin binding protein 1, putative	gi|115486133|	LF		↓	↓
		IT	↑		
Calcium-binding protein, Calmodulin-like protein 7, probable	gi|115474531|	LF	↓	↓	↓
		IT	↓		
Chlorophyll a-b binding protein	gi|115453971|	LF	↑		
		IT	↑	↑	
Os04g0640700 protein (84% homology with a H. vulgare Alpha-L-arabinofuranosidase/beta-D-xylosidase isoenzyme ARA-I Q8W012)	gi|115460876|	LF	↑		
		IT		↑	
Mannose-specific jacalin-related lectin, putative	gi|115488016|	LF	↑		
		IT		↓	
NADP-specific isocitrate dehydrogenase	gi|115438939|	LF		↑	
		IT	↑		
Alanine aminotransferase, putative	gi|115470235|	LF		↑	
		IT		↑	
Os05g0406000 protein (Fragment)	gi|115463823|	LF		↑	
		IT	↓		
Fructose-bisphosphate aldolase cytoplasmic isozyme, Gravity-specific protein	gi|115463789|	LF		↑	
		IT			↑
Coproporphyrinogen-III oxidase	gi|115460466|	LF			↑
		IT			↑
Cell division protease ftsH homolog 1, chloroplastic	gi|115470052|	LF		↑	↑
		IT		↑	
Pathogenesis-related protein 5, putative, expressed	gi|115489174|	LF	↑	↓	↑
		IT			↑
Os04g0111200 protein (94% homology with a Z. mays Bifunctional 3-phosphoadenosine 5-phosphosulfate synthetase 2 B6SRJ5)	gi|115456862|	LF	↑		↑
		IT	↑		
Drought-induced S-like ribonuclease	gi|115480399|	LF	↑		↑
		IT		↑	↑
Cell division protease ftsH homolog 2, chloroplastic	gi|115469444|	LF	↑	↑	↑
		IT	↑	↑	↑
Chlorophyll a/b-binding protein, putative	gi|115467828|	LF	↑	↑	↑
		IT	↑	↑	
Nucleic acid-binding protein, putative	gi|115480705|	LF	↑	↑	↑
		IT			↑
Abscisic acid-and stress-induced protein-rice	gi|115484359|	LF	↑	↑	↑
		IT	↑		
29 kDa ribonucleoprotein A, putative	gi|115473531|	LF	↑	↑	↑
		IT	↑		

Table 4.3 Overview of the differentially expressed proteins common to the label-free and iTRAQ methodologies. Unknown proteins were BLAST searched and homology to other sequences is shown in brackets in the protein name field. Methods described are LF = label-free and IT = iTRAQ. Significant differential expression is indicated by the up (↑) or down (↓) arrows for up- and down-regulation at the particular time point compared to the control at time point 0 h.

4.4.4 Biological Insights

Proteins that were significantly differentially expressed from either approach were functionally categorised using WEGO, a freely available online application that groups proteins based on their GO annotations [413]. WEGO allows the option to categorise the proteins based on biological process, molecular function or cellular compartment; we presented the data within using biological process at a GO annotation level of 5, and 21 functional categories selected. Some proteins have been counted more than once if they were assigned to more than one category. Using GO annotations is not currently a comprehensive method for functional classification, because only approximately 50% of the proteins are annotated. Nevertheless, it is a standardised, unbiased method for viewing global trends at the cellular level for protein and cellular function. An overview of the functional categories and the percentage of proteins allocated to each function are shown in supplemental data 4.3. Up- and down-regulated groups can be compared with the unchanged group to assess whether a protein group was intrinsically greater in abundance, or if it had been up- or down-regulated in response to cold stress. Functional groups that were noticeably changed in response to cold stress include transport, generation of precursor metabolites and energy metabolic process, and photosynthesis. Despite the large disparity between the proteins identified as significantly differentially expressed, similar trends in functional categories were seen across label-free and iTRAQ datasets. Figure 4.2 outlines these trends observed from the WEGO analysis. Conspicuous trends in this graph include the up-regulation of carbohydrate and protein metabolic processes and the down-regulation of proteins involved with oxidation and reduction.

A comprehensive overview of changes occurring in specific molecular pathways could not be obtained using this approach. We believe this is due to several factors: firstly, the presence of Rubisco, which masks proteins of lower abundance and is troublesome for MS data acquisition in a data-dependent mode, similar to albumin in human plasma; and secondly that the functional categories from WEGO are not complete and provided limited scope, hence further investigations were conducted manually.

In earlier studies of cold stress on rice leaf there have been many reported contradictions regarding which proteins were up- or down-regulated [336]. This study has identified similar proteins to other studies with regard to the up- and down-regulation due to cold stress, including photosynthetic proteins (Rubisco large chain, chloroplastic ATP synthase and oxygen evolving complex protein of photosystem II), ribosomal proteins, elongation factors, heat shock proteins, calreticulin, ascorbate peroxidase and UDP-glucose pyrophosphorylase [358, 416-420]. Since others have already discussed these proteins in detail with regard to cold stress response, we will focus on proteins not commonly identified as cold-responsive in rice plants.

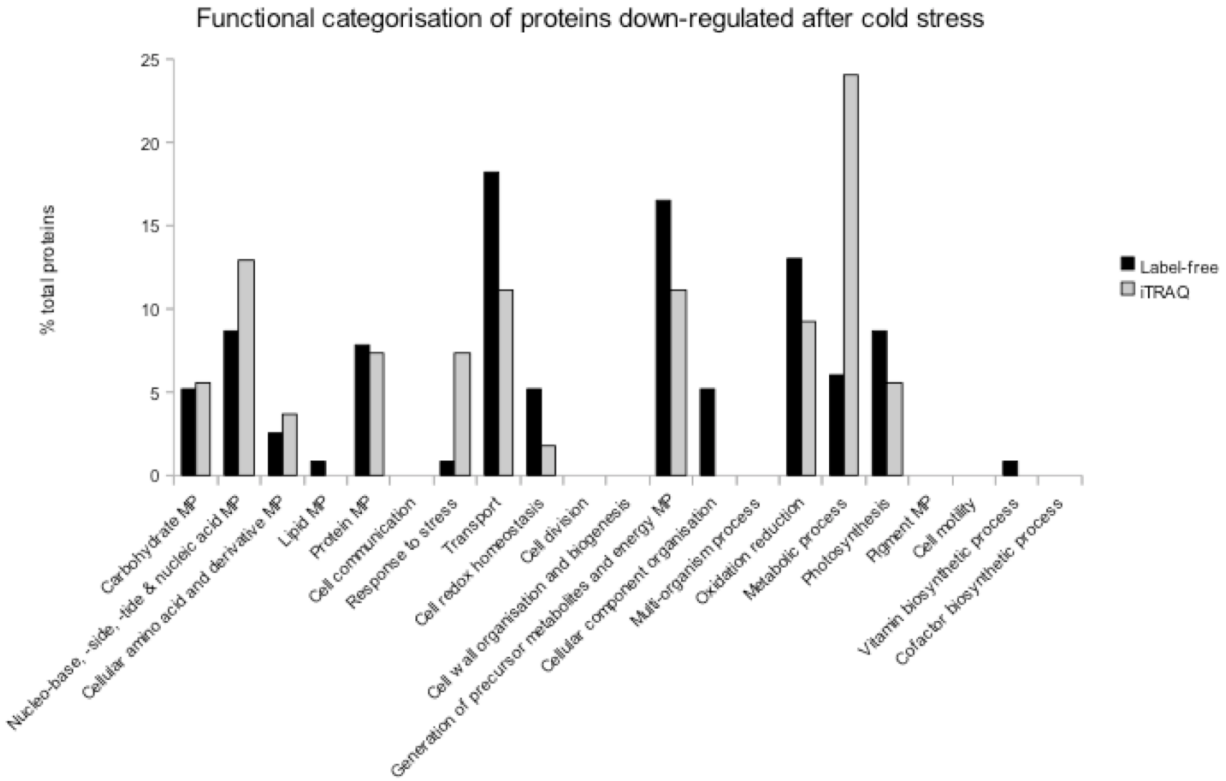


Figure 4.2 (A) Functional categorisation of differentially expressed proteins that were identified as down-regulated for the comparison between label-free (black) and iTRAQ (grey) methodologies. WEGO was use to generate these functional classifications and graphs that were found to be cold responsive in rice leaf. Some proteins with GO annotations were automatically assigned to one or more of the 21 categories based on the WEGO algorithm for functional classification.

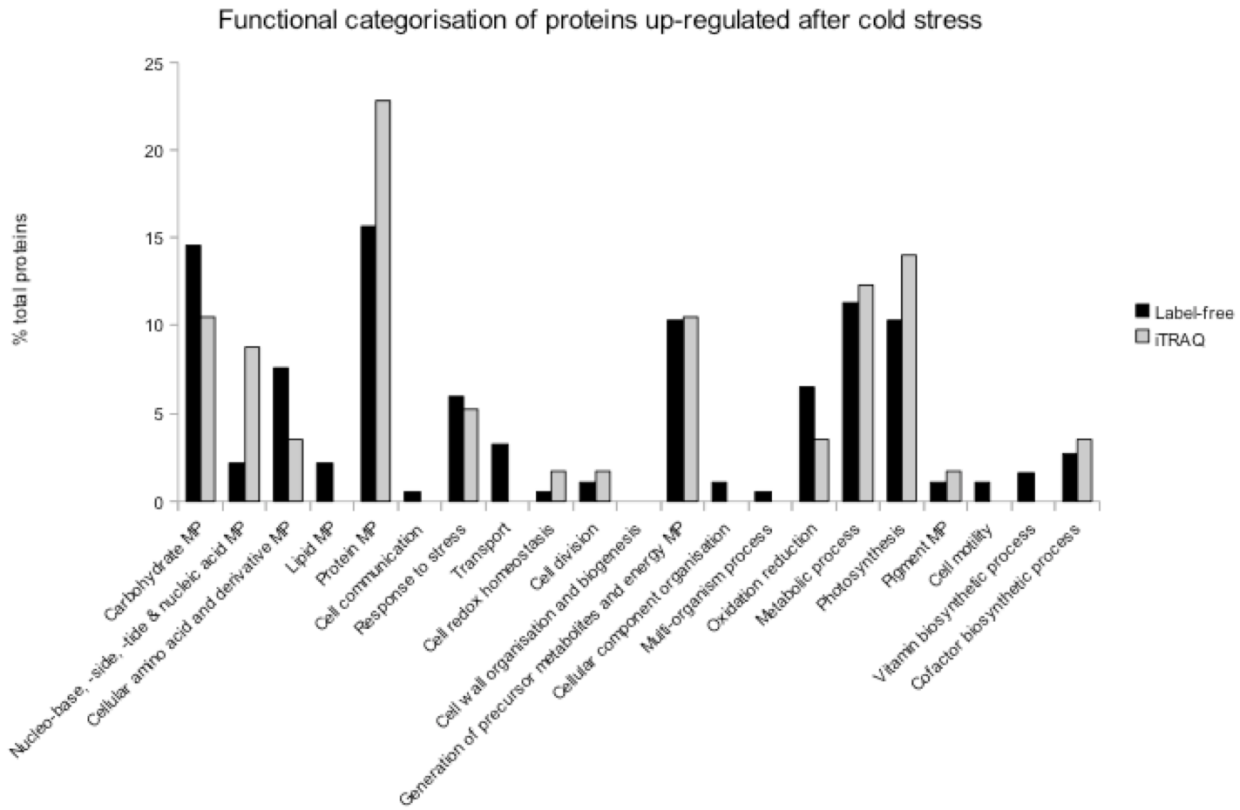


Figure 4.2 (B) Functional categorisation of differentially expressed proteins that were identified as up-regulated for the comparison between label-free (black) and iTRAQ (grey) methodologies. WEGO was use to generate these functional classifications and graphs that were found to be cold responsive in rice leaf. Some proteins with GO annotations were automatically assigned to one or more of the 21 categories based on the WEGO algorithm for functional classification.

Both label-free and iTRAQ methods suggest that histone regulation plays a role in the cold stress response. The principal histones, H4, H2B.9, H3.2, and linker histones, H12 and H5, were observed to be down-regulated, while H2A.3 was found to be up-regulated after 72 h of cold stress using the label-free approach. Previously, histone H1 has been reported as being up-regulated at the transcript level in Arabidopsis in response to cold, salt and drought stresses [421]. Histones are prone to reversible post-translational modifications such as methylation, phosphorylation, glycosylation, acetylation and ubiquitination, allowing the protein to respond flexibly to stimuli. Histone modification is a key regulator of gene expression in eukaryotic cells and has been implicated in plant stress response, including response to low temperatures [422-425]. Proteins known to modify histones include histone methyltransferases or acetyltransfases and WD-40 repeat proteins. Curiously, a WD-40 repeat protein was found to be up-regulated after 96 h of cold stress using the label-free method. It is not understood why histones were down-regulated in this study, nevertheless, our data suggest that a quantitative study of post translational modifications of histones may reveal more information in the regulation of cold tolerance.

The label-free approach suggests that vitamin B biosynthesis proteins may be altered in response to cold stress. Earlier studies have suggested that vitamin B has anti-oxidant activity in plants [426, 427]. Several vitamin B biosynthetic proteins were identified as being differentially expressed by the label-free approach. One of these, thiamine biosynthetic enzyme (vitamin B1), was expressed at all cold stress time points except for the control sample. Proteins involved in the synthesis of folate and riboflavin, vitamins B9 and B2 respectively, were down-regulated. Differential expression of vitamin B biosynthetic proteins in the wild rice species *O. meridionalis*, have recently been proposed to enhance thermal tolerance and our data provides additional evidence to support this hypothesis [428].

Lastly, the pyridoxal biosynthesis protein, (PDX; vitamin B6) was also identified in the label-free datasets, however, it is inconclusive as to whether this protein is up- or down-regulated. This protein has several similar candidates for identification and PDX1.1 and PDX1.2 were both identified. These homologues differ by only a few amino acids throughout their sequence with the majority of peptides used for identification and quantification being shared (non-unique) peptides, which can skew

the quantification data. In earlier studies on Arabidopsis, PDX1.1 and PDX1.3 were found to be responsive to oxidative stress while PDX1.2 was not [429, 430].

In this study we have shown that the label-free and iTRAQ methodologies may be considered complementary as they reveal similar trends in biological function for differentially expressed proteins in response to cold stress, despite the number of distinct peptides and proteins identified by each methodology. Nevertheless, there seemed to be no extra information obtainable from the iTRAQ data that was not already in the label-free data, while the label-free data appeared to reveal several insights not visible from the iTRAQ dataset.

4.5 FUTURE VIEWS AND DIRECTIONS

The implementation of abundant protein removal, namely Rubisco for plants, analogous to albumin removal in plasma, should increase the total number of identifications and quantifications of low abundance proteins, as it has in other studies to date [416-418]. This should increase the information and biological insights generated for future experiments in a rice leaf system.

I propose that if supplemental studies were conducted on these rice samples in a top-down proteomic manner, or at the very least similar to the work I conducted in Chapter 3 with protein fractionation followed by LC-MS/MS of the peptides, then a clearer distinction between PDX1.1 and PDX1.2 may be made. This would also provide more accurate information in a quantitative sense too and not just with respect to identification alone, as it has in other top-down proteomic studies [204, 431]. For example, if the minimal amino acid differences of PDX1.1 and PDX1.2 change the conformation or charge dynamics of the protein enough to enable differentiation within a liquid chromatographic environment in the protein form, this should facilitate differential chromatographic elution and distinction, whether these proteins are conjugated to different proteins, or they form super-molecular complexes as a dimers, trimers, or higher order multimers. Thus, the peptides that are shared (non-unique) between the two homologues of PDX1.1 and PDX1.2 could be confidently apportioned to the correct form, and the label-free quantification would be able to deliver a more accurate representation of the quantitation of protein homologues than is currently achievable.

Since the Paragon algorithm does not undertake quantification on peptides that are shared (non-unique) while the label-free method does not discriminate, this may account for the large disparity in protein identification numbers generated by iTRAQ for quantification compared to label-free. This may mean that the reduced number of proteins able to be used for quantification by iTRAQ, even though it identified more proteins (1269 compared to 1050 for label-free), is due to the ideology behind the manipulation of the data, rather than any concerns with the successful conjugation of label to the available peptides in the sample. Conversely, the label-free data may be over or under representing differentially expressed proteins due to the inclusion of

these shared (non-unique) peptides. This is an area that needs further investigation and development before it can really be considered a mature technology.

An intermediate solution to the protein inference problem for quantitative proteomics, rather than the implementation of protein fractionation, top-down proteomics or the exclusion of shared (non-unique) peptides in the analysis, is perhaps a more sophisticated form of the current NSAF. The current NSAF takes into consideration the length of the protein and the abundance, but it does not take into consideration the difference between unique and non-unique peptides. I propose that, on organisms with a completed genome, a theoretical digest is conducted on the theoretical transcribed genome. Each non-unique peptide is allocated a weighted number, representing the total number of times it can be theoretically produced and how many proteins compared to the total number of peptides and proteins in the genome. This is also done for the unique peptides. Spectral counting of one unique peptide and three non-unique may have an adjusted count of 1.6 as an arbitrary example, rather than 4 in the non-weighted version. This weighting could bring the relative differential expression figures generated by label-free proteomics closer to those of absolute quantification or labelled techniques, increasing one’s confidence in the data. There is a potential flaw in this argument being that at any one point in time the entire genome is not being translated or transcribed. Thus, the weighted numbers based on the whole genome may skew the adjusted figures without appropriate compensatory mathematical models.

CHAPTER 5: CONCLUSION

This thesis concludes with respect to:

1) The use of self assembled monolayer chemistries orientated in concentric circles attached to metal plates (SCSC-MALDI) for use as a concentrator of peptide samples on a MALDI plate can afford a practitioner increases in limits of detection between 10-100 fold for identical sample concentrations compared to standard MALDI practices, in the attomole/µl range. Compared to the AnchorChip^{TM} for single protein digests, both techniques performed similarly for the identification of the proteins from digests, while outperforming the standard MALDI technique. The variability of the concentration event for the SCSC-MALDI made it difficult to achieve reproducibility without a multitude of replicates making the technique problematic at best, be this due to issues with respect to sample, manufacturing variabilities, or altered surface chemistry. The ability to conduct the removal of salts pre-concentration on the surface of the plate was shown to be achievable on simple peptide standards, though it was problematic at the time of the experiments, particularly with respect to gel plug extracts, and needs further investigation. The ability to selectively capture phosphorylated peptides from standard protein digests could be achieved at the low femtomole level. However, the changing of the surface chemistry to enable affinity capture removed the ability of the biochip to concentrate the sample and needs further development before it is ready for use as a viable technique that can conduct both affinity capture and concentration.
2) The development of the PROOF method highlighted the ability to fractionate a complex protein sample (depleted human plasma) with ease and to identify similar proteins as a standard shotgun proteomic method, with ~74% similarity in proteins identified. Additionally, the ability to relate peptide MS/MS data

back to the original and structural forms of the protein with respect to truncated or cleaved elements of the proteome was shown through the novel graphical representation system developed. This highlighted the potential misidentification of current bottom-up experiments and the need for further development of this (PROOF) or particular systems that enable proteomics researchers to relate MS/MS peptide information back to the speciation of the proteome with ease. This would allow correct assignment of identification, form and function of the proteins and provides a compelling argument for the development of top-down proteomics strategies for the future of the field.

3) Comparison of the protein expression levels in leaf material of cold temperature stressed rice with both label-free and labeling (iTRAQ) quantitation proteomics, highlighted the burgeoning capabilities and arguable superiority of the label-free quantification techniques, in particular spectral counting. Label-free identified and quantified 236 cold responsive proteins, while iTRAQ identified and quantified 85, with only 24 proteins in common between the methods employed. Both methods are complementary for use in plants, delivering similar insights into the biological response, despite the disparity in total proteins quantified. The functional analysis did reveal differential expression of proteins involved in transport, photosynthesis, generation of precursor metabolites and energy. Additionally, histones and vitamin B biosynthetic proteins were observed to be affected by cold stress in rice leaf only in the label-free experiments. This provides a positive argument for the use of spectral counting rather than iTRAQ quantification for quantifying protein expression levels in plant systems.

In closing, it has been approximately 8 years from the conception of this thesis within proteomics (2004) to the eventual completion (2012). Taking into consideration two factors, the freshness of the field and the rapid growth of methodologies since mapping of the human genome at the turn of the century, I have observed many precipitous changes for the advancement of the field of proteomics and hold great optimism for the part it will play in the betterment of both scientific knowledge and the quality of our society in the near and distant future.

BIBLIOGRAPHY

[1] James, P., Protein identification in the post-genome era: The rapid rise of proteomics. *Q. Rev. Biophys.* **1997**, *30*, 279-331.

[2] Lamarck, J. B., *Système des animaux sans vertèbres*, Bavarian State Library **1801**.

[3] Coleman, W. L., *Biology in the Nineteenth Century: Problems of form, function, and transformation*, Cambridge University Press, Cambridge **1977**.

[4] Lindberg, D. C., *The Beginnings of Western Science: The European scientific tradition in philisophical, religious, and institutional context, prehistory to A.D. 1450*, The University of Chicago Press, London **1992**.

[5] Simpson, G. G., Biology and the nature of science. *Science* **1963**, *139*, 81-88.

[6] Croft, W. J., *Under the Microscope: A brief history of microscopy*, World Scientific Publishing, Singapore **2006**.

[7] Ruska, E., in: Grandin, K. (Ed.), *Les Prix Nobel*, The Nobel Prize in Physics 1986: Ernst Ruska, Gerd Binnig, Heinrich Rohrer, The Royal Swedish Academy of Sciences, Stockholm **1986**.

[8] Bowman, S., *Interpreting the Past: Radio carbon dating*, University of California Press, Berkely Los Angeles **1990**.

[9] Libby, W. F., in: Grandin, K. (Ed.), *Les Prix Nobel*, The Nobel Prize in Physics 1960: Willard Frank Libby, The Royal Swedish Academy of Sciences, Stockholm **1960**.

[10] Saunders, M., Wishnia, A., Kirkwood, J. G., The nuclear magnetic resonce spectrum of ribonuclease 1. *J. Am. Chem. Soc.* **1957**, *79*, 3289-3290.

[11] Wagner, G., Wüthrich, K., Sequential resonance assignments in protein 1H nuclear magnetic resonance spectra: Basic pancreatic trypsin inhibitor. *J. Mol. Biol.* **1982**, *155*, 347-366.

[12] Pryde, A., *Applications of High Performance Liquid Chromatography*, Chapman and Hall, London **1979**.

[13] Strain, H. H., Sherma, J., Michael Tswett's contributions to sixty years of chromatography. *J. Chem. Educ.* **1967**, *44*, 235.

[14] Tswett, M., Physikalisch-chemische studien über das chlorophyll. Die adsorption. (Physical-chemical studies of chlorophyll. Adsorption.). *Berichte der Deutschen Botanischen Gesellschaft* **1906**, *24*, 316-326.

[15] Tswett, M., Adsorptionanalyse und chromatographische methode. Anwendung auf die chemie des chlorophylls. (Adsorption analysis and chromatographic method. Application to the chemistry of chlorophyll.). *Berichte der Deutschen Botanischen Gesellschaft* **1906**, *24*, 384-393.

[16] Watson, T. J., *Introduction to Mass Spectrometry: Third edition*, Lippincott-Raven Philadelphia **1997**.

[17] Thomson, J. J., *Rays of Positive Electricity and Their Application to Chemical Analyses*, Longmans, Green and Co., London **1913**.

[18] Barlett, J. M. S., Stirling, D., in: Barlett, J. M. S., Stirling, D. (Eds.), *PCR Protocols*, A Short History of the Polymerase Chain Reaction, Humana Press **2003**, pp. 3-6.

[19] Saiki, R. K., Scharf, S., Faloona, F., Mullis, K. B., Horn, G. T., Erlich, H. A., Arnheim, N., Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* **1985**, *230*, 1350-1354.

[20] Crowfoot, D., X-Ray single crystal photographs of insulin. *Nature* **1935**, *135*, 591-592.

[21] Kendrew, J. C., Bodo, G., Dintzis, H. M., Parrish, R. G., Wyckoff, H., Phillips, D. C., A three-dimensional model of the myoglobin molecule obtained by X-Ray analysis. *Nature* **1958**, *181*, 662-666.
[22] Watson, J. D., Crick, F. H. C., Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. *Nature* **1953**, *171*, 737-738.
[23] Kenrick, K. G., Margolis, J., Isoelectric focusing and gradient gel electrophoresis: A two-dimensional technique. *Anal. Biochem.* **1970**, *33*, 204-207.
[24] O'Farrell, P. H., High resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.* **1975**, *250*, 4007-4021.
[25] Klose, J., Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues. A novel approach to testing for induced point mutations in mammals. *Humangenetik* **1975**, *26*, 231-243.
[26] Perrett, D., From ‘protein’ to the beginnings of clinical proteomics. *Proteomics Clin. Appl.* **2007**, *1*, 720-738.
[27] Wasinger, V. C., Cordwell, S. J., Cerpa-Poljak, A., Yan, J. X., Gooley, A. A., Wilkins, M. R., . . . Humphery-Smith, I., Progress with gene-product mapping of the Mollicutes: Mycoplasma genitalium. *Electrophoresis* **1995**, *16*, 1090-1094.
[28] Yates, J. R., 3rd, Mass spectrometry: From genomics to proteomics. *Trends Genet.* **2000**, *16*, 5-8.
[29] Karas, M., Bachmann, D., Bahr, U., Hillenkamp, F., Matrix-assisted ultraviolet laser desorption of non-volatile compounds. *Int. J. Mass Spectrom* **1987**, *78*, 53-68.
[30] Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., Whitehouse, C. M., Electrospray ionization for mass spectrometry of large biomolecules. *Science* **1989**, *246*, 64-71.
[31] Aston, F. W., in: Grandin, K. (Ed.), *Les Prix Nobel*, The Nobel Prize in Chemistry 1922: Francis W. Aston, The Royal Swedish Academy of Sciences, Stockholm **1922**.
[32] Aston, F. W., A magnetic mass spectrometer. *Philisophical Magazine* **1919**, *38*, 707-712.
[33] Ekman, R., Silberring, J., Westman-Brinkmalm, A. M., Kraj, A., *Mass Spectrometry: Instrumentation, interpretation and applications*, Wiley, Hoboken, New Jersey **2009**.
[34] Karas, M., Bachmann, D., Hillenkamp, F., Influence of the wavelength in high-irradiance ultraviolet laser desorption mass spectrometry of organic molecules. *Anal. Chem.* **1985**, *57*, 2935-2939.
[35] Tanaka, K., Waki, H., Ido, Y., Akita, S., Yoshida, Y., Yoshida, T., Matsuo, T., Protein and polymer analyses up to m/z 100 000 by laser ionization time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* **1988**, *2*, 151-153.
[36] Tanaka, K., in: Grandin, K. (Ed.), *Les Prix Nobel*, The Royal Swedish Academy of Sciences, Stockholm **2002**.
[37] Dole, M., Mack, L. L., Hines, R. L., Mobley, R. C., Ferguson, L. D., Alice, M. B., Molecular beams of macroions. *J. Chem. Phys.* **1968**, *49*, 2240-2249.
[38] Taylor, G., Disintegration of water drops in an electric field. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* **1964**, *280*, 383-397.
[39] Yamashita, M., Fenn, J. B., Electrospray ion source. Another variation on the free-jet theme. *J. Phys. Chem.* **1984**, *88*, 4451-4459.
[40] Fenn, J. B., in: Grandin, K. (Ed.), *Les Prix Nobel*, The Royal Swedish Academy of Sciences, Stockholm **2002**.
[41] Knochenmuss, R., Zenobi, R., MALDI ionization: The role of in-plume processes. *Chem. Rev.* **2003**, *103*, 441-452.

[42] Dreisewerd, K., The desorption process in MALDI. *Chem. Rev.* **2003**, *103*, 395-426.
[43] Zenobi, R., Knochenmuss, R., Ion formation in MALDI mass spectrometry. *Mass Spectrom. Rev.* **1998**, *17*, 337-336.
[44] Strupat, K., Karas, M., Hillenkamp, F., 2,5-Dihydroxybenzoic acid: A new matrix for laser desorption-ionization mass spectrometry. *Int. J. Mass Spectrom* **1991**, *111*, 89-102.
[45] Beavis, R. C., Chaudhary, T., Chait, B. T., α-Cyano-4-hydroxycinnamic acid as a matrix for matrixassisted laser desorption mass spectromtry. *Org. Mass Spectrom.* **1992**, *27*, 156-158.
[46] Karas, M., Kruger, R., Ion formation in MALDI: the cluster ionization mechanism. *Chem. Rev.* **2003**, *103*, 427-440.
[47] Ehring, H., Karas, M., Hillenkamp, F., Role of photoionization and photochemistry in ionization processes of organic-molecules and relevance for matrix-assisted laser desorption ionization mass-spectrometry. *Org. Mass Spectrom.* **1992**, *27*, 472-480.
[48] Chang, W. C., Huang, L. C. L., Wang, Y.-S., Peng, W.-P., Chang, H. C., Hsu, N. Y., . . . Chen, C. H., Matrix-assisted laser desorption/ionization (MALDI) mechanism revisited. *Anal. Chim. Acta* **2007**, *582*, 1-9.
[49] Kuzyk, M. A., Ohlund, L. B., Elliott, M. H., Smith, D., Qian, H., Delaney, A., . . . Borchers, C. H., A comparison of MS/MS-based, stable-isotope-labeled, quantitation performance on ESI-quadrupole TOF and MALDI-TOF/TOF mass spectrometers. *Proteomics* **2009**, *9*, 3328-3340.
[50] Wilm, M., Mann, M., Analytical properties of the nanoelectrospray ion source. *Anal. Chem.* **1996**, *68*, 1-8.
[51] Wilm, M. S., Mann, M., Electrospray and Taylor-Cone theory, Dole's beam of macromolecules at last? *Int. J. Mass Spectrom* **1994**, *136*, 167-180.
[52] Aleksandrov, M. L., Gall, L. N., Krasnov, N. V., Nikolaev, V. I., Pavlenko, V. A., Shkurov, V. A., Ion extraction from solutions at atmospheric pressure: method of mass-spectrometric analysis of bio-organic substances. *Dokl. Akad. Nauk SSSR* **1984**, *277*, 379-383.
[53] Wilm, M., Principles of electrospray ionization. *Mol. Cell. Proteomics* **2011**, *10*, 1-8.
[54] Veenstra, T. D., Electrospray ionization mass spectrometry in the study of biomolecular non-covalent interactions. *Biophys. Chem.* **1999**, *79*, 63-79.
[55] Bolbach, G., Matrix-assisted laser desorption/ionization analysis of non-covalent complexes: Fundamentals and applications. *Curr. Pharm. Des.* **2005**, *11*, 2535-2557.
[56] Strupat, K., Rogniaux, H., Van Dorsselaer, A., Roth, J., Vogl, T., Calcium-induced noncovalently linked tetramers of MRP8 and MRP14 are confirmed by electrospray ionization-mass analysis. *J. Am. Soc. Mass Spectrom.* **2000**, *11*, 780-788.
[57] Vogl, T., Roth, J., Sorg, C., Hillenkamp, F., Strupat, K., Calcium-induced noncovalently linked tetramers of MRP8 and MRP14 detected by ultraviolet matrix-assisted laser desorption/ionization mass spectrometry. *J. Am. Soc. Mass Spectrom.* **1999**, *10*, 1124-1130.
[58] Westerman-Brinkmalm, A., Brinkmalm, G., *Mass Spectrometry and Hyphenated Techniques in Neuropeptide Research*, John Wiley and Sons, New York **2002**.
[59] Urban, P. L., Amantonico, A., Zenobi, R., Lab-on-a-plate: Extending the functionality of MALDI-MS and LDI-MS targets. *Mass Spectrom. Rev.* **2011**, *30*, 435-478.

[60] Navare, A., Nouzova, M., Noriega, F. G., Hernandez-Martinez, S., Menzel, C., Fernandez, F. M., On-chip solid-phase extraction pre-concentration/focusing substrates coupled to atmospheric pressure matrix-assisted laser desorption/ionization ion trap mass spectrometry for high sensitivity biomolecule analysis. *Rapid Commun. Mass Spectrom.* **2009**, *23*, 477-486.
[61] Nordhoff, E., Schürenberg, M., Thiele, G., Lübbert, C., Kloeppel, K.-D., Theiss, D., . . . Gobom, J., Sample preparation protocols for MALDI-MS of peptides and oligonucleotides using prestructured sample supports. *Int. J. Mass Spectrom.* **2003**, *226*, 163-180.
[62] Hutchens, T. W., Yip, T. T., New desorption strategies for the mass spectrometric analysis of macromolecules. *Rapid Commun. Mass Spectrom.* **1993**, *7*, 576-580.
[63] Issaq, H. J., Veenstra, T. D., Conrads, T. P., Felschow, D., The SELDI-TOF MS approach to proteomics: Protein profiling and biomarker identification. *Biochem. Biophys. Res. Commun.* **2002**, *292*, 587-592.
[64] Tang, N., Tornatore, P., Weinberger, S. R., Current developments in SELDI affinity technology. *Mass Spectrom. Rev.* **2004**, *23*, 34-44.
[65] Zhang, J., Kinsel, G. R., Quantification of protein-polymer interactions by matrix-assisted laser desorption/ionization mass spectrometry. *Langmuir* **2002**, *18*, 4444-4448.
[66] Zhang, J., Kinsel, G. R., Evaluation of surface-protein binding constants by matrix-assisted laser desorption ionization mass spectrometry. *Langmuir* **2003**, *19*, 3531-3534.
[67] Röhl, H., Goethel, S., Reihs, K., MPep MALDI chips for high-sensitivity and high-throughput peptide analysis by MALDI-TOF MS. *Nat. Methods* **2005**, *2*.
[68] Love, J. C., Estroff, L. A., Kriebel, J. K., Nuzzo, R. G., Whitesides, G. M., Self-assembled monolayers of thiolates on metals as a form of nanotechnology. *Chem. Rev.* **2005**, *105*, 1103-1170.
[69] Nuzzo, R. G., Allara, D. L., Adsorption of bifunctional organic disulfides on gold surfaces. *J. Am. Chem. Soc.* **1983**, *105*, 4481-4483.
[70] Porter, M. D., Bright, T. B., Allara, D. L., Chidsey, C. E. D., Spontaneously organized molecular assemblies. 4. Structural characterization of n-alkyl thiol monolayers on gold by optical ellipsometry, infrared spectroscopy, and electrochemistry. *J. Am. Chem. Soc.* **1987**, *109*, 3559-3568.
[71] Dubois, L. H., Nuzzo, R. G., Synthesis, structure, and properties of model organic surfaces. *Annu. Rev. Phys. Chem.* **1992**, *43*, 437-463.
[72] Bain, C. D., Whitesides, G. M., Molecular-level control over surface order in self-assembled monolayer films of thiols on gold. *Science* **1988**, *240*, 62-63.
[73] Laibinis, P. E., Whitesides, G. M., Allara, D. L., Tao, Y. T., Parikh, A. N., Nuzzo, R. G., Comparison of the structures and wetting properties of self-assembled monolayers of n-alkanethiols on the coinage metal surfaces, copper, silver, and gold. *J. Am. Chem. Soc.* **1991**, *113*, 7152-7167.
[74] Dubois, L. H., Zegarski, B. R., Nuzzo, R. G., Molecular ordering of organosulfur compounds on Au(111) and Au(100): Adsorption from solution and in ultrahigh vacuum. *J. Chem. Phys.* **1993**, *98*, 678.
[75] Poirier, G. E., Pylant, E. D., The self-assembly mechanism of alkanethiols on Au(111). *Science* **1996**, *272*, 1145-1148.
[76] Laibinis, P. E., Whitesides, G. M., .omega.-Terminated alkanethiolate monolayers on surfaces of copper, silver, and gold have similar wettabilities. *J. Am. Chem. Soc.* **1992**, *114*, 1990-1995.
[77] Covey, T. R., Huang, E. C., Henion, J. D., Structural characterization of protein tryptic peptides via liquid chromatography/mass spectrometry and collision-

induced dissociation of their doubly charged molecular ions. *Anal. Chem.* **1991**, *63*, 1193-1200.
[78] Colorado, R., Lee, T. R., Wettabilities of self-assembled monolayers on gold generated from progressively fluorinated alkanethiols. *Langmuir* **2003**, *19*, 3288-3296.
[79] Leggett, G. J., Friction force microscopy of self-assembled monolayers: Probing molecular organisation at the nanometre scale. *Anal. Chim. Acta* **2003**, *479*, 17-38.
[80] Houston, J. E., Kim, H. I., Adhesion, friction, and mechanical properties of functionalized alkanethiol aelf-assembled monolayers. *Accounts Chem. Res.* **2002**, *35*, 547-553.
[81] Jennings, G. K., Yong, T.-H., Munro, J. C., Laibinis, P. E., Structural effects on the barrier properties of self-assembled monolayers formed from long-chain ω-alkoxy-n-alkanethiols on copper. *J. Am. Chem. Soc.* **2003**, *125*, 2950-2957.
[82] Burleigh, T. D., Gu, Y., Donahey, G., Vida, M., D.H. Waldeck, Tarnish protection of silver using a hexadecanethiol self-assembled monolayer and descriptions of accelerated tarnish tests. *Corrosion* **2001**, *57*, 1066-1074.
[83] Becker, H., Chips, money, industry, education and the "killer application". *Lab. Chip* **2009**, *9*, 1659-1660.
[84] Anderson, N. G., Anderson, N. L., Twenty years of two-dimensional electrophoresis: Past, present and future. *Electrophoresis* **1996**, *17*, 443-453.
[85] Ahmed, F. E., Mining the oncoproteome and studying molecular interactions for biomarker development by 2DE, ChIP and SPR technologies. *Expert Rev. Proteomics* **2008**, *5*, 469-496.
[86] Palagi, P. M., Hernandez, P., Walther, D., Appel, R. D., Proteome informatics I: Bioinformatics tools for processing experimental data. *Proteomics* **2006**, *6*, 5435-5444.
[87] Ahmed, F. E., Utility of mass spectrometry for proteome analysis: Part II. Ion-activation methods, statistics, bioinformatics and annotation. *Expert Rev. Proteomics* **2009**, *6*, 171(127).
[88] Spengler, B., Kirsch, D., Kaufmann, R., Jaeger, E., Peptide sequencing by matrix-assisted laser-desorption mass spectrometry. *Rapid Commun. Mass Spectrom.* **1992**, *6*, 105-108.
[89] Rouse, J. C., Yu, W., Martin, S. A., A comparison of the peptide fragmentation obtained from a reflector matrix-assisted laser desorption-ionization time-of-flight and a tandem four sector mass spectrometer. *J. Am. Soc. Mass Spectrom.* **1995**, *6*, 822-835.
[90] Chaurand, P., Luetzenkirchen, F., Spengler, B., Peptide and protein identification by matrix-assisted laser desorption ionization (MALDI) and MALDI-post-source decay time-of-flight mass spectrometry. *J. Am. Soc. Mass Spectrom.* **1999**, *10*, 91-103.
[91] Medzihradszky, K. F., Campbell, J. M., Baldwin, M. A., Falick, A. M., Juhasz, P., Vestal, M. L., Burlingame, A. L., The characteristics of peptide collision-induced dissociation using a high-performance MALDI-TOF/TOF tandem mass spectrometer. *Anal. Chem.* **2000**, *72*, 552-558.
[92] Wahlander, A., Arrigoni, G., Snel, M., Hellman, U., James, P., Parallel post-source decay for increasing protein identification confidence levels from 2-D gels. *Proteomics* **2008**, *8*, 1771-1779.
[93] Domon, B., Aebersold, R., Mass spectrometry and protein analysis. *Science* **2006**, *312*, 212-217.
[94] Jennings, K. R., The changing impact of the collision-induced decomposition of ions on mass spectrometry. *Int. J. Mass Spectrom.* **2000**, *200*, 479-486.

[95] Roepstorff, P., Fohlman, J., Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biol. Mass Spectrom.* **1984**, *11*, 601-601.
[96] Sleno, L., Volmer, D. A., Ion activation methods for tandem mass spectrometry. *J. Mass Spectrom.* **2004**, *39*, 1091-1112.
[97] Zubarev, R. A., Kelleher, N. L., McLafferty, F. W., Electron Capture Dissociation of multiply charged protein cations. A nonergodic process. *J. Am. Chem. Soc.* **1998**, *120*, 3265-3266.
[98] Zubarev, R. A., Reactions of polypeptide ions with electrons in the gas phase. *Mass Spectrom. Rev.* **2003**, *22*, 57-77.
[99] Li, W., Hendrickson, C. L., Emmett, M. R., Marshall, A. G., Identification of intact proteins in mixtures by alternated capillary liquid chromatography electrospray ionization and LC ESI infrared multiphoton dissociation fourier transform ion cyclotron resonance mass spectrometry. *Anal. Chem.* **1999**, *71*, 4397-4402.
[100] Price, W. D., Schnier, P. D., Williams, E. R., Tandem mass spectrometry of large biomolecule ions by blackbody infrared radiative dissociation. *Anal. Chem.* **1996**, *68*, 859-866.
[101] March, R. E., Quadrupole ion trap mass spectrometry: theory, simulation, recent developments and applications. *Rapid Commun. Mass Spectrom.* **1998**, *12*, 1543-1554.
[102] Yu, L. R., Zhu, Z., Chan, K. C., Issaq, H. J., Dimitrov, D. S., Veenstra, T. D., Improved titanium dioxide enrichment of phosphopeptides from HeLa cells and high confident phosphopeptide identification by cross-validation of MS/MS and MS/MS/MS spectra. *J. Proteome Res.* **2007**, *6*, 4150-4162.
[103] Fröhlich, T., Arnold, G. J., Proteome research based on modern liquid chromatography - tandem mass spectrometry: separation, identification and quantification. *J. Neural Transm.* **2006**, *113*, 973-994.
[104] Ahmed, F. E., Utility of mass spectrometry for proteome ana lysis: Part I. Conceptual and experimental approaches. *Expert Rev. Proteomics* **2008**, *5*, 841(824).
[105] Stephens, W. E., A pulsed mass spectrometer is time disperson. *Phys. Rev. Lett.* **1946**, *69*, 691.
[106] McDaniel, E. W., Martin, D. W., Barnes, W. S., Drift-tube mass spectrometer for studies of low-energy ion-molecule reactions. *Rev. Sci. Instrum.* **1962**, *33*, 2-12.
[107] Paul, W., Steinwedel, H., Ein nues massenspektrometer ohne magnetfield. *Z. Naturforsh* **1953**, *A8*, 448-450.
[108] Paul, W., Steinwedel, H., Method of operating ion trap detector in MS/MS mode, **1960**, U.S. Patent No. 2939952.
[109] Schwartz, J. C., Senko, M. W., Syka, J. E., A two-dimensional quadrupole ion trap mass spectrometer. *J. Am. Soc. Mass Spectrom.* **2002**, *13*, 659-669.
[110] Kingdom, K. H., A method for the neutralization of electron space charge by positive ionization at very low gas pressure. *Physical Reviews* **1923**, *21*, 2717-2728.
[111] Makarov, A. A., Mass spectrometer, **1999**, U.S. Patent No. 5886346.
[112] Makarov, A., Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis. *Anal. Chem.* **2000**, *72*, 1156-1162.
[113] Lawerence, E. O., Edlefsen, N. E., On the production of high speed protons. *Science* **1930**, *72*, 376-377.
[114] Lawerence, E. O., Livingston, M. S., The production of high speed light ions without the use of high voltages. *Phys. Rev. Lett.* **1932**, *40*, 19-35.
[115] Sommer, H., Thomas, H. A., Hipple, J. A., A precise method of determining the faraday by magnetic resonance. *Phys. Rev. Lett.* **1949**, *76*, 1877-1878.
[116] Sommer, H., Thomas, H. A., Hipple, J. A., The measurement of e/m by cyclotron resonance. *Phys. Rev. Lett.* **1951**, *82*, 697.

[117] Scigelova, M., Hornshaw, M., Giannakopulos, A., Makarov, A., Fourier transform mass spectrometry. *Mol. Cell. Proteomics* **2011**.
[118] Dehmelt, H. G., in: Grandin, K. (Ed.), *Les Prix Nobel*, The Royal Swedish Academy of Sciences, Stockholm **1989**.
[119] Paul, W., in: Grandin, K. (Ed.), *Les Prix Nobel*, The Royal Swedish Academy of Sciences, Stockholm **1989**.
[120] Koppenaal, D. W., Barinaga, C. J., Denton, M. B., Sperline, R. P., Hieftje, G. M., Schilling, G. D., . . . Barnes, J. H. t., MS detectors. *Anal. Chem.* **2005**, *77*, 418A-427A.
[121] Molléa, D., Jardin, J., Piot, M., Pasco, M., Léonila, J., Gagnaire, V., Comparison of electrospray and matrix-assisted laser desorption ionization on the same hybrid quadrupole time-of-flight tandem mass spectrometer: Application to bidimensional liquid chromatography of proteins from bovine milk fraction. *J. Chromatogr. A* **2009**, *1216*, 2424-2432.
[122] Wisniewski, J. R., Zougman, A., Nagaraj, N., Mann, M., Universal sample preparation method for proteome analysis. *Nat. Methods* **2009**, *6*, 359-362.
[123] Graumann, J., Hubner, N. C., Kim, J. B., Ko, K., Moser, M., Kumar, C., . . . Mann, M., Stable isotope labeling by amino acids in cell culture (SILAC) and proteome quantitation of mouse embryonic stem cells to a depth of 5,111 proteins. *Mol. Cell. Proteomics* **2008**, *7*, 672-683.
[124] Chait, B. T., Chemistry. Mass spectrometry: bottom-up or top-down? *Science* **2006**, *314*, 65-66.
[125] Ge, Y., Lawhorn, B. G., ElNaggar, M., Strauss, E., Park, J. H., Begley, T. P., McLafferty, F. W., Top down characterization of larger proteins (45 kDa) by electron capture dissociation mass spectrometry. *J. Am. Chem. Soc.* **2002**, *124*, 672-678.
[126] Calligaris, D., Villard, C., Lafitte, D., Advances in top-down proteomics for disease biomarker discovery. *J. Proteomics* **2011**, *74*, 920-934.
[127] Garcia, B. A., What does the future hold for top down mass spectrometry? *J. Am. Soc. Mass Spectrom.* **2010**, *21*, 193-202.
[128] Jungblut, P., Thiede, B., Zimny-Arndt, U., Muller, E. C., Scheler, C., Wittmann-Liebold, B., Otto, A., Resolution power of two-dimensional electrophoresis and identification of proteins from gels. *Electrophoresis* **1996**, *17*, 839-847.
[129] Jungblut, P. R., Holzhutter, H. G., Apweiler, R., Schluter, H., The speciation of the proteome. *Chem. Cent. J.* **2008**, *2*, 16.
[130] Schluter, H., Apweiler, R., Holzhutter, H. G., Jungblut, P. R., Finding one's way in proteomics: A protein species nomenclature. *Chem. Cent. J.* **2009**, *3*, 11.
[131] Young, N. L., Plazas-Mayorca, M. D., Garcia, B. A., Systems-wide proteomic characterization of combinatorial post-translational modification patterns. *Expert Rev. Proteomics* **2010**, *7*, 79-92.
[132] Jensen, O. N., Interpreting the protein language using proteomics. *Nat. Rev. Mol. Cell Biol.* **2006**, *7*, 391-403.
[133] Huang, T.-y., McLuckey, S. A., Top-down protein characterization facilitated by ion/ion reactions on a quadrupole/time of flight platform. *Proteomics* **2010**, *10*, 3577-3588.
[134] Tran, J. C., Zamdborg, L., Ahlf, D. R., Lee, J. E., Catherman, A. D., Durbin, K. R., . . . Kelleher, N. L., Mapping intact protein isoforms in discovery mode using top-down proteomics. *Nature* **2011**.
[135] Johnson, W. C., Protein secondary structure and circular dichroism: A practical guide. *Proteins* **1990**, *7*, 205-214.
[136] Coons, A., Creech, H., Jones, R., Immunological properties of an antibody containing a fluorescent group. *Proc. Soc. Exp. Biol. Med.* **1941**, *47*, 200-202.

[137] Hunter, R. L., Markert, C. L., Histochemical demonstration of enzymes separated by zone electrophoresis in starch gels. *Science* **1957**, *125*, 1294-1295.
[138] Burger, A., Richterich, R., Aebi, H., The heterogenicity of creatine kinase. *Biochem. Z.* **1964**, *339*, 305-314.
[139] Eppenberger, H. M., Dawson, D. M., Kaplan, N. O., The comparative enzymology of creatine kinases. I. Isolation and characterization from chicken and rabbit tissues. *The Journal of Biological Chemistry* **1967**, *242*, 204-209.
[140] Dawson, D. M., Eppenberger, H. M., Kaplan, N. O., The comparative enzymology of creatine kinases. II. Physical and chemical properties. *The Journal of Biological Chemistry* **1967**, *242*, 210-217.
[141] Shapiro, A. L., Vinuela, E., Maizel, J. V., Molecular weight estimation of polypeptide chains by electrophoresis in SDS-polyacrylamide gels. *Biochem. Biophys. Res. Commun.* **1967**, *28*, 815–820.
[142] Laemmli, U. K., Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **1970**, *227*, 680-685.
[143] Shevchenko, A., Tomas, H., Havlis, J., Olsen, J. V., Mann, M., In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat. Protoc.* **2006**, *1*, 2856-2860.
[144] Walls, D., Loughran, S. T., *Protein Chromatography: Methods and Protocols*, Springer, New York **2011**.
[145] Link, A. J., Eng, J., Schieltz, D. M., Carmack, E., Mize, G. J., Morris, D. R., . . . Yates, J. R., Direct analysis of protein complexes using mass spectrometry. *Nat. Biotechnol.* **1999**, *17*, 676.
[146] Washburn, M. P., Wolters, D., Yates, J. R., 3rd, Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol.* **2001**, *19*, 242-247.
[147] Yates, J. R., 3rd, Ruse, C. I., Nakorchevsky, A., Proteomics by mass spectrometry: approaches, advances, and applications. *Annu. Rev. Biomed. Eng.* **2009**, *11*, 49-79.
[148] Westermeier, R., Naven, T., Hopker, H.-R., *Proteomics in Practice*, Wiley-VCH, Weinheim **2008**.
[149] Rabilloud, T., Membrane proteins and proteomics: Love is possible, but so difficult. *Electrophoresis* **2009**, *30 Suppl 1*, S174-180.
[150] Wilkins, M. R., Gasteiger, E., Sanchez, J. C., Bairoch, A., Hochstrasser, D. F., Two-dimensional gel electrophoresis for proteome projects: The effects of protein hydrophobicity and copy number. *Electrophoresis* **1998**, *19*, 1501-1505.
[151] McDonald, W. H., Yates, J. R., 3rd, Shotgun proteomics and biomarker discovery. *Dis. Markers* **2002**, *18*, 99-105.
[152] Henzel, W. J., Watanabe, C., Stults, J. T., Protein identification: The origins of peptide mass fingerprinting. *J. Am. Soc. Mass Spectrom.* **2003**, *14*, 931-942.
[153] Medina, M. A., Haynes, P. A., Breci, L., Francisco, W. A., Analysis of secreted proteins from Aspergillus flavus. *Proteomics* **2005**, *5*, 3153-3161.
[154] Breci, L., Hattrup, E., Keeler, M., Letarte, J., Johnson, R., Haynes, P. A., Comprehensive proteomics in yeast using chromatographic fractionation, gas phase fractionation, protein gel electrophoresis, and isoelectric focusing. *Proteomics* **2005**, *5*, 2018-2028.
[155] Armirotti, A., Damonte, G., Achievements and perspectives of top-down proteomics. *Proteomics* **2010**, *10*, 3566–3576.
[156] Meng, F., Cargile, B. J., Patrie, S. M., Johnson, J. R., McLoughlin, S. M., Kelleher, N. L., Processing complex mixtures of intact proteins for direct analysis by mass spectrometry. *Anal. Chem.* **2002**, *74*, 2923-2929.

[157] Casado-Vela, J., Rodriguez-Suarez, E., Iloro, I., Ametzazurra, A., Alkorta, N., Garcia-Velasco, J. A., . . . Elortza, F., Comprehensive proteomic analysis of human endometrial fluid aspirate. *J. Proteome Res.* **2009**, *8*, 4622-4632.
[158] Neilson, K. A., Mariani, M., Haynes, P. A., Quantitative proteomic analysis of cold-responsive proteins in rice. *Proteomics* **2011**, *11*, 1696-1706.
[159] Richey, J., FPLC: A comprehensive separation technique for biopolymers. *Am. Lab.* **1982**, *14*, 104-129.
[160] Plank, J., Andres, P. R., Krause, I., Winter, C., Gram scale separation of casein proteins from whole casein on a Source 30Q anion-exchange resin column utilizing fast protein liquid chromatography (FPLC). *Protein Expr. Purif.* **2008**, *60*, 176-181.
[161] Boesze-Battaglia, K., Kong, F., Lamba, O. P., Stefano, F. P., Williams, D. S., Purification and light-dependent phosphorylation of a candidate fusion protein, the photoreceptor cell peripherin/rds. *Biochemistry (Mosc.)* **1997**, *36*, 6835-6846.
[162] Miyazaki, O., Fukamachi, I., Mori, A., Hashimoto, H., Kawashiri, M. A., Nohara, A., . . . Kobayashi, J., Formation of prebeta1-HDL during lipolysis of triglyceride-rich lipoprotein. *Biochem. Biophys. Res. Commun.* **2009**, *379*, 55-59.
[163] Zeng, C.-M., Liao, J.-L., Nakazato, K. i., Hjert√©n, S., Hydrophobic-interaction chromatography of proteins on continuous beds derivatized with isopropyl groups. *J. Chromatogr. A* **1996**, *753*, 227-234.
[164] Moon, P. G., Hwang, H. H., Boo, Y. C., Kwon, J., Cho, J. Y., Baek, M. C., Identification of rat urinary glycoproteome captured by three lectins using gel and LC-based proteomics. *Electrophoresis* **2008**, *29*, 4324-4331.
[165] Cabrera, K., Applications of silica-based monolithic HPLC columns. *J. Sep. Sci.* **2004**, *27*, 843-852.
[166] Tangvarasittichai, S., Tangvarasittichai, O., Jermnim, N., Comparison of fast protein liquid chromatography (FPLC) with HPLC, electrophoresis & microcolumn chromatography techniques for the diagnosis of beta-thalassaemia. *The Indian Journal of Medical Research* **2009**, *129*, 242-248.
[167] Croley, T. R., Ultra-performance liquid chromatography-mass spectrometry of proteins. *Methods Mol. Biol.* **2011**, *681*, 419-429.
[168] MacNair, J. E., Opiteck, G. J., Jorgenson, J. W., Moseley, M. A., 3rd, Rapid separation and characterization of protein and peptide mixtures using 1.5 microns diameter non-porous silica in packed capillary liquid chromatography/mass spectrometry. *Rapid Commun. Mass Spectrom.* **1997**, *11*, 1279-1285.
[169] Eschelbach, J. W., Jorgenson, J. W., Improved protein recovery in reversed-phase liquid chromatography by the use of ultrahigh pressures. *Anal. Chem.* **2006**, *78*, 1697-1706.
[170] Mellors, J. S., Jorgenson, J. W., Use of 1.5-microm porous ethyl-bridged hybrid particles as a stationary-phase support for reversed-phase ultrahigh-pressure liquid chromatography. *Anal. Chem.* **2004**, *76*, 5441-5450.
[171] Cummins, P. M., Dowling, O., O'Connor, B. F., Ion-exchange chromatography: Basic principles and application to the partial purification of soluble mammalian prolyl oligopeptidase. *Methods Mol. Biol.* **2011**, *681*, 215-228.
[172] Fritz, J. S., Early milestones in the development of ion-exchange chromatography: A personal account. *Journal of Chromatography. A* **2004**, *1039*, 3-12.
[173] Lucy, C. A., Evolution of ion-exchange: From Moses to the Manhattan project to modern times. *J. Chromatogr. A* **2003**, *1000*, 711-724.
[174] Kent, U. M., Purification of antibodies using ion-exchange chromatography. *Methods Mol. Biol.* **1999**, *115*, 19-22.

[175] Levison, P. R., Large-scale ion-exchange column chromatography of proteins. Comparison of different formats. *J. Chromatogr. B Analyt. Technol. Biomed. Life. Sci.* **2003**, *790*, 17-33.
[176] Cummins, P. M., O'Connor, B., Bovine brain pyroglutamyl aminopeptidase (type-1): Purification and characterisation of a neuropeptide-inactivating peptidase. *IJBCB* **1996**, *28*, 883-893.
[177] Wu, B., Wu, L., Chen, D., Yang, Z., Luo, M., Purification and characterization of a novel fibrinolytic protease from *Fusarium sp*. CPCC 480097. *J. Ind. Microbiol. Biotechnol.* **2009**, *36*, 451-459.
[178] Knudsen, H. L., Fahrner, R. L., Xu, Y., Norling, L. A., Blank, G. S., Membrane ion-exchange chromatography for process-scale antibody purification. *J. Chromatogr. A* **2001**, *907*, 145-154.
[179] Mant, C. T., Hodges, R. S., Mixed-mode hydrophilic interaction/cation-exchange chromatography: Separation of complex mixtures of peptides of varying charge and hydrophobicity. *J. Sep. Sci.* **2008**, *31*, 1573-1584.
[180] Fekkes, D., Voskuilen-Kooyman, A., Jankie, R., Huijmans, J., Precise analysis of primary amino acids in urine by an automated high-performance liquid chromatography method: Comparison with ion-exchange chromatography. *J. Chromatogr. B Biomed. Sci. Appl.* **2000**, *744*, 183-188.
[181] Yang, Y., Hebron, H. R., Hang, J., High performance DNA purification using a novel ion exchange matrix. *J. Biomol. Tech.* **2008**, *19*, 205-210.
[182] Bonn, G., High-performance liquid chromatographic isolation of 14C-labelled gluco-oligosaccharides, monosaccharides and sugar degradation products on ion-exchange resins. *J. Chromatogr.* **1987**, *387*, 393-398.
[183] Hajos, P., Nagy, L., Retention behaviours and separation of carboxylic acids by ion-exchange chromatography. *J. Chromatogr. B Biomed. Sci. Appl.* **1998**, *717*, 27-38.
[184] Casado-Vela, J., Cebrián, A., Gómez del Pulgar, M. T., Sánchez-López, E., Vilaseca, M., Menchén, L., . . . Lacal, J. C., Lights and shadows of proteomic technologies for the study of protein species including isoforms, splicing variants and protein post-translational modifications. *Proteomics* **2011**, *11*, 590-603.
[185] Nesvizhskii, A. I., Aebersold, R., Interpretation of shotgun proteomic data: The protein inference problem. *Mol. Cell. Proteomics* **2005**, *4*, 1419-1440.
[186] Bielow, C., Gröpl, C., Kohlbacher, O., Reinert, K., in: Mayer, B. (Ed.), *Bioinformatics for Omics Data: Methods in Molecular Biology*, Bioinformatics for Qualitative and Quantitative Proteomics, Springer Science **2011**, pp. 331-349.
[187] Perkins, D. N., Pappin, D. J., Creasy, D. M., Cottrell, J. S., Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **1999**, *20*, 3551-3567.
[188] Yates, J. R., 3rd, Eng, J. K., McCormack, A. L., Schieltz, D., Method to correlate tandem mass spectra of modified peptides to amino acid sequences in the protein database. *Anal. Chem.* **1995**, *67*, 1426-1436.
[189] Matthiesen, R., Trelle, M. B., Hojrup, P., Bunkenborg, J., Jensen, O. N., VEMS 3.0: Algorithms and computational tools for tandem mass spectrometry based identification of post-translational modifications in proteins. *J. Proteome Res.* **2005**, *4*, 2338-2347.
[190] Craig, R., Beavis, R. C., TANDEM: Matching proteins with tandem mass spectra. *Bioinformatics* **2004**, *20*, 1466-1467.
[191] Geer, L. Y., Markey, S. P., Kowalak, J. A., Wagner, L., Xu, M., Maynard, D. M., . . . Bryant, S. H., Open mass spectrometry search algorithm. *J. Proteome Res.* **2004**, *3*, 958-964.

[192] Nesvizhskii, A. I., Keller, A., Kolker, E., Aebersold, R., A statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem.* **2003**, *75*, 4646-4658.
[193] Domon, B., Aebersold, R., Challenges and opportunities in proteomics data analysis. *Mol. Cell. Proteomics* **2006**, *5*, 1921-1926.
[194] Bogdanov, B., Smith, R. D., Proteomics by FTICR mass spectrometry: Top down and bottom up. *Mass Spectrom. Rev.* **2005**, *24*, 168-200.
[195] Kelleher, N. L., Lin, H. Y., Valaskovic, G. A., Aaserud, D. J., Fridriksson, E. K., McLafferty, F. W., Top down versus bottom up protein characterization by tandem high-resolution mass spectrometry. *J. Am. Chem. Soc.* **1999**, *121*, 806-812.
[196] Hernandez, H., Dziembowski, A., Taverner, T., Seraphin, B., Robinson, C. V., Subunit architecture of multimeric complexes isolated directly from cells. *EMBO Rep.* **2006**, *7*, 605-610.
[197] Han, X., Jin, M., Breuker, K., McLafferty, F. W., Extending top-down mass spectrometry to proteins with masses greater than 200 kilodaltons. *Science* **2006**, *314*, 109-112.
[198] Vellaichamy, A., Tran, J. C., Catherman, A. D., Lee, J. E., Kellie, J. F., Sweet, S. M., . . . Kelleher, N. L., Size-sorting combined with improved nanocapillary liquid chromatography-mass spectrometry for identification of intact proteins up to 80 kDa. *Anal. Chem.* **2010**, *82*, 1234-1244.
[199] Siuti, N., Kelleher, N. L., Decoding protein modifications using top-down mass spectrometry. *Nat. Methods* **2007**, *4*, 817-821.
[200] Strader, M. B., Verberkmoes, N. C., Tabb, D. L., Connelly, H. M., Barton, J. W., Bruce, B. D., . . . Hurst, G. B., Characterization of the 70S Ribosome from Rhodopseudomonas palustris using an integrated "top-down" and "bottom-up" mass spectrometric approach. *J. Proteome Res.* **2004**, *3*, 965-978.
[201] Parks, B. A., Jiang, L., Thomas, P. M., Wenger, C. D., Roth, M. J., Boyne, M. T., 2nd, . . . Kelleher, N. L., Top-down proteomics on a chromatographic time scale using linear ion trap fourier transform hybrid mass spectrometers. *Anal. Chem.* **2007**, *79*, 7984-7991.
[202] Durbin, K. R., Tran, J. C., Zamdborg, L., Sweet, S. M., Catherman, A. D., Lee, J. E., . . . Kelleher, N. L., Intact mass detection, interpretation, and visualization to automate top-down proteomics on a large scale. *Proteomics* **2010**, *10*, 3589-3597.
[203] Redondo, P., Prieto, J., Munoz, I. G., Alibes, A., Stricher, F., Serrano, L., . . . Montoya, G., Molecular basis of xeroderma pigmentosum group C DNA recognition by engineered meganucleases. *Nature* **2008**, *456*, 107-111.
[204] Mazur, M. T., Cardasis, H. L., Spellman, D. S., Liaw, A., Yates, N. A., Hendrickson, R. C., Quantitative analysis of intact apolipoproteins in human HDL by top-down differential mass spectrometry. *Proceedings of the National Academy of Sciences* **2010**, *107*, 7728-7733.
[205] Hernandez, H., Makarova, O. V., Makarov, E. M., Morgner, N., Muto, Y., Krummel, D. P., Robinson, C. V., Isoforms of U1-70k control subunit dynamics in the human spliceosomal U1 snRNP. *PLoS ONE* **2009**, *4*, e7202.
[206] Suckau, D., Resemann, A., T3-sequencing: Targeted characterization of the N- and C-termini of undigested proteins by mass spectrometry. *Anal. Chem.* **2003**, *75*, 5817-5824.
[207] Horn, D. M., Zubarev, R. A., McLafferty, F. W., Automated *de novo* sequencing of proteins by tandem high-resolution mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 10313-10317.
[208] Syka, J. E., Coon, J. J., Schroeder, M. J., Shabanowitz, J., Hunt, D. F., Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 9528-9533.

[209] Tran, J. C., Doucette, A. A., Rapid and effective focusing in a carrier ampholyte solution isoelectric focusing system: A proteome prefractionation tool. *J. Proteome Res.* **2008**, *7*, 1761-1766.
[210] Tran, J. C., Doucette, A. A., Gel-eluted liquid fraction entrapment electrophoresis: An electrophoretic method for broad molecular weight range proteome separation. *Anal. Chem.* **2008**, *80*, 1568-1573.
[211] Lee, J. E., Kellie, J. F., Tran, J. C., Tipton, J. D., Catherman, A. D., Thomas, H. M., . . . Kelleher, N. L., A robust two-dimensional separation for top-down tandem mass spectrometry of the low-mass proteome. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 2183-2191.
[212] Seidler, J., Zinn, N., Boehm, M. E., Lehmann, W. D., *De novo* sequencing of peptides by MS/MS. *Proteomics* **2010**, *10*, 634-649.
[213] Xu, P., Peng, J., Characterization of polyubiquitin chain structure by middle-down mass spectrometry. *Anal. Chem.* **2008**, *80*, 3438-3444.
[214] Garcia, B. A., Siuti, N., Thomas, C. E., Mizzen, C. A., Kelleher, N. L., Characterization of neurohistone variants and post-translational modifications by electron capture dissociation mass spectrometry. *Int. J. Mass Spectrom.* **2007**, *259*, 184-196.
[215] Wu, S., Tolick, N., Tian, Z., Robinson, E. W., Pasa-Tolic, L., in: Wu, C. H., Chen, C. (Eds.), *Methods in Molecular Biology Methods and Protocols: Bioinformatics for Comparative Proteomics*, An Integrated Top-Down and Bottom-Up Strategy for Characterization of Protein Isoforms and Modifications, Springer Science **2011**, pp. 291-304.
[216] Schiess, R., Wollscheid, B., Aebersold, R., Targeted proteomic strategy for clinical biomarker discovery. *Mol. Oncol.* **2009**, *3*, 33-44.
[217] Faca, V., Krasnoselsky, A., Hanash, S., Innovative proteomic approaches for cancer biomarker discovery. *Biotechniques* **2007**, *43*, 279, 281-273, 285.
[218] Polanski, M., Anderson, N. L., A list of candidate cancer biomarkers for targeted proteomics. *Biomark Insights* **2007**, *1*, 1-48.
[219] Gygi, S. P., Rochon, Y., Franza, B. R., Aebersold, R., Correlation between protein and mRNA abundance in Yeast. *Mol. Cell. Biol.* **1999**, *19*, 1720-1730.
[220] Hatzimanikatis, V., Lee, K. H., Dynamical analysis of gene networks requires both mRNA and protein expression information. *Metab. Eng.* **1999**, *1*, 275-281.
[221] Ong, S. E., Foster, L. J., Mann, M., Mass spectrometric-based approaches in quantitative proteomics. *Methods* **2003**, *29*, 124-130.
[222] Ong, S. E., Mann, M., Mass spectrometry-based proteomics turns quantitative. *Nat. Chem. Biol.* **2005**, *1*, 252-262.
[223] Steen, H., Pandey, A., Proteomics goes quantitative: Measuring protein abundance. *Trends Biotechnol.* **2002**, *20*, 361-364.
[224] Cheng, D., Hoogenraad, C. C., Rush, J., Ramm, E., Schlager, M. A., Duong, D. M., . . . Peng, J., Relative and absolute quantification of postsynaptic density proteome isolated from rat forebrain and cerebellum. *Mol. Cell. Proteomics* **2006**, *5*, 1158-1170.
[225] Cox, J., Mann, M., Quantitative, high-resolution proteomics for data-driven systems biology. *Annu. Rev. Biochem.* **2011**, *80*, 273-299.
[226] Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H., Aebersold, R., Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* **1999**, *17*, 994-999.
[227] Ross, P. L., Huang, Y. N., Marchese, J. N., Williamson, B., Parker, K., Hattan, S., . . . Pappin, D. J., Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol. Cell. Proteomics* **2004**, *3*, 1154-1169.

[228] Thompson, A., Schafer, J., Kuhn, K., Kienle, S., Schwarz, J., Schmidt, G., . . . Hamon, C., Tandem Mass Tags: A novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal. Chem.* **2003**, *75*, 1895-1904.
[229] Ong, S. E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A., Mann, M., Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* **2002**, *1*, 376-386.
[230] Ibarrola, N., Molina, H., Iwahori, A., Pandey, A., A novel proteomic approach for specific identification of tyrosine kinase substrates using [13C]tyrosine. *J. Biol. Chem.* **2004**, *279*, 15805-15813.
[231] Martinovic, S., Veenstra, T. D., Anderson, G. A., Pasa-Tolic, L., Smith, R. D., Selective incorporation of isotopically labeled amino acids for identification of intact proteins on a proteome-wide level. *J. Mass Spectrom.* **2002**, *37*, 99-107.
[232] Oda, Y., Huang, K., Cross, F. R., Cowburn, D., Chait, B. T., Accurate quantitation of protein expression and site-specific phosphorylation. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 6591-6596.
[233] Bondarenko, P. V., Chelius, D., Shaler, T. A., Identification and relative quantitation of protein mixtures by enzymatic digestion followed by capillary reversed-phase liquid chromatography-tandem mass spectrometry. *Anal. Chem.* **2002**, *74*, 4741-4749.
[234] Chelius, D., Bondarenko, P. V., Quantitative profiling of proteins in complex mixtures using liquid chromatography and mass spectrometry. *J. Proteome Res.* **2002**, *1*, 317-323.
[235] Liu, H., Sadygov, R. G., Yates, J. R., 3rd, A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal. Chem.* **2004**, *76*, 4193-4201.
[236] Bantscheff, M., Schirle, M., Sweetman, G., Rick, J., Kuster, B., Quantitative mass spectrometry in proteomics: A critical review. *Anal. Bioanal. Chem.* **2007**, *389*, 1017-1031.
[237] Neilson, K. A., Ali, N. A., Muralidharan, S., Mirzaei, M., Mariani, M., Assadourian, G., . . . Haynes, P. A., Less label, more free: Approaches in label-free quantitative mass spectrometry. *Proteomics* **2011**, *11*, 535-553.
[238] Lundgren, D. H., Hwang, S.-I., Wu, L., Han, D. K., Role of spectral counting in quantitative proteomics. *Expert Rev. Proteomics* **2010**, *7*, 39-53.
[239] Florens, L., Carozza, M. J., Swanson, S. K., Fournier, M., Coleman, M. K., Workman, J. L., Washburn, M. P., Analyzing chromatin remodeling complexes using shotgun proteomics and normalized spectral abundance factors. *Methods* **2006**, *40*, 303-311.
[240] Zybailov, B., Mosley, A. L., Sardiu, M. E., Coleman, M. K., Florens, L., Washburn, M. P., Statistical analysis of membrane proteome expression changes in *Saccharomyces cerevisiae*. *J. Proteome Res.* **2006**, *5*, 2339-2347.
[241] Zybailov, B. L., Florens, L., Washburn, M. P., Quantitative shotgun proteomics using a protease with broad specificity and normalized spectral abundance factors. *Mol. Biosyst.* **2007**, *3*, 354-360.
[242] Pavelka, N., Fournier, M. L., Swanson, S. K., Pelizzola, M., Ricciardi-Castagnoli, P., Florens, L., Washburn, M. P., Statistical similarities between transcriptomics and quantitative shotgun proteomics data. *Mol. Cell. Proteomics* **2008**, *7*, 631-644.
[243] Chick, J. M., Haynes, P. A., Bjellqvist, B., Baker, M. S., A combination of immobilised pH gradients improves membrane proteomics. *J. Proteome Res.* **2008**, *7*, 4974-4981.

[244] Mosley, A. L., Florens, L., Wen, Z., Washburn, M. P., A label free quantitative proteomic analysis of the Saccharomyces cerevisiae nucleus. *J. Proteomics* **2009**, *72*, 110-120.
[245] Gammulla, C. G., Pascovici, D., Atwell, B. J., Haynes, P. A., Differential metabolic response of cultured rice (*Oryza sativa*) cells exposed to high- and low-temperature stress. *Proteomics* **2010**, *10*, 3001-3019.
[246] Voelckel, C., Mirzaei, M., Reichelt, M., Luo, Z., Pascovici, D., Heenan, P. B., . . . Lockhart, P. J., Transcript and protein profiling identify candidate gene sets of potential adaptive significance in New Zealand *Pachycladon*. *BMC Evol. Biol.* **2010**, *10*, 151.
[247] Sardiu, M. E., Cai, Y., Jin, J., Swanson, S. K., Conaway, R. C., Conaway, J. W., . . . Washburn, M. P., Probabilistic assembly of human protein interaction networks from label-free quantitative proteomics. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 1454-1459.
[248] Zhao, Y., Denner, L., Haidacher, S. J., LeJeune, W. S., Tilton, R. G., Comprehensive analysis of the mouse renal cortex using two-dimensional HPLC - tandem mass spectrometry. *Proteome Sci* **2008**, *6*, 15.
[249] Zhu, W., Smith, J. W., Huang, C. M., Mass spectrometry-based label-free quantitative proteomics. *J Biomed Biotechnol* **2010**, *2010*, 840518.
[250] Wiese, S., Reidegeld, K. A., Meyer, H. E., Warscheid, B., Protein labeling by iTRAQ: A new tool for quantitative mass spectrometry in proteome research. *Proteomics* **2007**, *7*, 340-350.
[251] Schmidt, A., Kellermann, J., Lottspeich, F., A novel strategy for quantitative proteomics using isotope-coded protein labels. *Proteomics* **2005**, *5*, 4-15.
[252] Black, D. L., Protein diversity from alternative splicing: A challenge for bioinformatics and post-genome biology. *Cell* **2000**, *103*, 367-370.
[253] Podwojski, K., Eisenacher, M., Kohl, M., Turewicz, M., Meyer, H. E., Rahnenfuhrer, J., Stephan, C., Peek a peak: A glance at statistics for quantitative label-free proteomics. *Expert Rev. Proteomics* **2010**, *7*, 249-261.
[254] Usaite, R., Wohlschlegel, J., Venable, J. D., Park, S. K., Nielsen, J., Olsson, L., Yates Iii, J. R., Characterization of global yeast quantitative proteome data generated from the wild-type and glucose repression saccharomyces cerevisiae strains: The comparison of two quantitative methods. *J. Proteome Res.* **2008**, *7*, 266-275.
[255] Choi, H., Fermin, D., Nesvizhskii, A. I., Significance analysis of spectral count data in label-free shotgun proteomics. *Mol. Cell. Proteomics* **2008**, *7*, 2373-2385.
[256] Liu, W. L., Coleman, R. A., Grob, P., King, D. S., Florens, L., Washburn, M. P., . . . Tjian, R., Structural changes in TAF4b-TFIID correlate with promoter selectivity. *Mol. Cell* **2008**, *29*, 81-91.
[257] Zybailov, B., Rutschow, H., Friso, G., Rudella, A., Emanuelsson, O., Sun, Q., van Wijk, K. J., Sorting signals, N-terminal modifications and abundance of the chloroplast proteome. *PLoS ONE* **2008**, *3*, e1994.
[258] Jin, S., Daly, D. S., Springer, D. L., Miller, J. H., The effects of shared peptides on protein quantitation in label-free proteomics by LC/MS/MS. *J. Proteome Res.* **2008**, *7*, 164-169.
[259] Feng, J., Naiman, D. Q., Cooper, B., Probability model for assessing proteins assembled from peptide sequences inferred from tandem mass spectrometry data. *Anal. Chem.* **2007**, *79*, 3901-3911.
[260] Zhang, Y., Wen, Z., Washburn, M. P., Florens, L., Refinements to label free proteome quantitation: How to deal with peptides shared by multiple proteins. *Anal. Chem.* **2010**, *82*, 2272-2281.

[261] Venable, J. D., Dong, M. Q., Wohlschlegel, J., Dillin, A., Yates, J. R., Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra. *Nat. Methods* **2004**, *1*, 39-45.
[262] Ghaemmaghami, S., Huh, W. K., Bower, K., Howson, R. W., Belle, A., Dephoure, N., . . . Weissman, J. S., Global analysis of protein expression in yeast. *Nature* **2003**, *425*, 737-741.
[263] Hwang, S. I., Lundgren, D. H., Mayya, V., Rezaul, K., Cowan, A. E., Eng, J. K., Han, D. K., Systematic characterization of nuclear proteome during apoptosis: A quantitative proteomic study by differential extraction and stable isotope labeling. *Mol. Cell. Proteomics* **2006**, *5*, 1131-1145.
[264] Chen, E. I., Cociorva, D., Norris, J. L., Yates, J. R., 3rd, Optimization of mass spectrometry-compatible surfactants for shotgun proteomics. *J. Proteome Res.* **2007**, *6*, 2529-2538.
[265] Liu, K., Zhang, J., Wang, J., Zhao, L., Peng, X., Jia, W., . . . Qian, X., Relationship between sample loading amount and peptide identification and Its effects on quantitative proteomics. *Anal. Chem.* **2009**, *81*, 1307-1314.
[266] Zhang, Y., Wen, Z., Washburn, M. P., Florens, L., Effect of dynamic exclusion duration on spectral count based quantitative proteomics. *Anal. Chem.* **2009**, *81*, 6317-6326.
[267] Lu, P., Vogel, C., Wang, R., Yao, X., Marcotte, E. M., Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nat. Biotechnol.* **2007**, *25*, 117-124.
[268] Turck, C. W., Falick, A. M., Kowalak, J. A., Lane, W. S., Lilley, K. S., Phinney, B. S., . . . Yates, N. A., The Association of Biomolecular Resource Facilities Proteomics Research Group 2006 Study. *Mol. Cell. Proteomics* **2007**, *6*, 1291-1298.
[269] Ryu, S., Gallis, B., Goo, Y. A., Shaffer, S. A., Radulovic, D., Goodlett, D. R., Comparison of a label-free quantitative proteomic method based on peptide ion current area to the isotope coded affinity tag method. *Cancer Inform.* **2008**, *6*, 243-255.
[270] Colinge, J., Chiappe, D., Lagache, S., Moniatte, M., Bougueleret, L., Differential proteomics via probabilistic peptide identification scores. *Anal. Chem.* **2005**, *77*, 596-606.
[271] Chen, J., Ryu, S., Gharib, S. A., Goodlett, D. R., Schnapp, L. M., Exploration of the normal human bronchoalveolar lavage fluid proteome. *Proteomics. Clinical applications* **2008**, *2*, 585-595.
[272] Patel, V. J., Thalassinos, K., Slade, S. E., Connolly, J. B., Crombie, A., Murrell, J. C., Scrivens, J. H., A comparison of labeling and label-free mass spectrometry-based proteomics approaches. *J. Proteome Res.* **2009**, *8*, 3752-3759.
[273] Carvalho, P. C., Han, X., Xu, T., Cociorva, D., Carvalho Mda, G., Barbosa, V. C., Yates, J. R., 3rd, XDIA: Improving on the label-free data-independent analysis. *Bioinformatics* **2010**, *26*, 847-848.
[274] Panchaud, A., Scherl, A., Shaffer, S. A., von Haller, P. D., Kulasekara, H. D., Miller, S. I., Goodlett, D. R., Precursor acquisition independent from ion count: How to dive deeper into the proteomics ocean. *Anal. Chem.* **2009**, *81*, 6481-6488.
[275] Anderson, N. L., Anderson, N. G., The human plasma proteome: History, character, and diagnostic prospects. *Mol. Cell. Proteomics* **2002**, *1*, 845-867.
[276] Jacobs, J. M., Adkins, J. N., Qian, W. J., Liu, T., Shen, Y., Camp, D. G., 2nd, Smith, R. D., Utilizing human blood plasma for proteomic biomarker discovery. *J. Proteome Res.* **2005**, *4*, 1073-1085.
[277] Omenn, G. S., States, D. J., Adamski, M., Blackwell, T. W., Menon, R., Hermjakob, H., . . . Hanash, S. M., 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*, 3226-3245.
[278] Good, D. M., Thongboonkerd, V., Novak, J., Bascands, J.-L., Schanstra, J. P., Coon, J. J., . . . Mischak, H., Body fluid proteomics for biomarker discovery: Lessons from the past hold the key to success in the future. *J. Proteome Res.* **2007**, *6*, 4549-4555.
[279] Celis, J. E., Rasmussen, H. H., Olsen, E., Madsen, P., Leffers, H., Honore, B., . . . et al., The human keratinocyte two-dimensional protein database (update 1994): Towards an integrated approach to the study of cell proliferation, differentiation and skin diseases. *Electrophoresis* **1994**, *15*, 1349-1458.
[280] Berggård, T., Linse, S., James, P., Methods for the detection and analysis of protein–protein interactions. *Proteomics* **2007**, *7*, 2833-2842.
[281] Nilsen, T. W., Graveley, B. R., Expansion of the eukaryotic proteome by alternative splicing. *Nature* **2010**, *463*, 457-463.
[282] Blakeley, P., Siepen, J. A., Lawless, C., Hubbard, S. J., Investigating protein isoforms via proteomics: A feasibility study. *Proteomics* **2010**, *10*, 1127-1140.
[283] Modrek, B., Lee, C., A genomic view of alternative splicing. *Nat. Genet.* **2002**, *30*, 13-19.
[284] Modrek, B., Lee, C. J., Alternative splicing in the human, mouse and rat genomes is associated with an increased frequency of exon creation and/or loss. *Nat. Genet.* **2003**, *34*, 177-180.
[285] Pan, Q., Shai, O., Lee, L. J., Frey, B. J., Blencowe, B. J., Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.* **2008**, *40*, 1413-1415.
[286] Wang, E. T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., . . . Burge, C. B., Alternative isoform regulation in human tissue transcriptomes. *Nature* **2008**, *456*, 470-476.
[287] de la Grange, P., Dutertre, M., Martin, N., Auboeuf, D., FAST DB: A website resource for the study of the expression regulation of human gene products. *Nucleic Acids Res.*, *33*, 4276-4284.
[288] Hanash, S. M., Pitteri, S. J., Faca, V. M., Mining the plasma proteome for cancer biomarkers. *Nature* **2008**, *452*, 571-579.
[289] Hortin, G. L., Jortani, S. A., Ritchie, J. C., Jr., Valdes, R., Jr., Chan, D. W., Proteomics: A new diagnostic frontier. *Clin. Chem.* **2006**, *52*, 1218-1222.
[290] Tirumalai, R. S., Chan, K. C., Prieto, D. A., Issaq, H. J., Conrads, T. P., Veenstra, T. D., Characterization of the low molecular weight human serum proteome. *Mol. Cell. Proteomics* **2003**, *2*, 1096-1103.
[291] Rifai, N., Gillette, M. A., Carr, S. A., Protein biomarker discovery and validation: The long and uncertain path to clinical utility. *Nat. Biotechnol.* **2006**, *24*, 971-983.
[292] Kasimir-Bauer, S., Circulating tumor cells as markers for cancer risk assessment and treatment monitoring. *Mol. Diagn. Ther.* **2009**, *13*, 209-215.
[293] States, D. J., Omenn, G. S., Blackwell, T. W., Fermin, D., Eng, J., Speicher, D. W., Hanash, S. M., Challenges in deriving high-confidence protein identifications from data gathered by a HUPO plasma proteome collaborative study. *Nat. Biotechnol.* **2006**, *24*, 333-338.
[294] Hortin, G. L., Sviridov, D., The dynamic range problem in the analysis of the plasma proteome. *J. Proteomics* **2010**, *73*, 629-636.
[295] Surinova, S., Schiess, R., HuÃàttenhain, R., Cerciello, F., Wollscheid, B., Aebersold, R., On the development of plasma protein biomarkers. *J. Proteome Res.* **2010**, *10*, 5-16.

[296] FDA, FDA clears a test for ovarian cancer: Test can help identify potential makignancies, guide surgical decisions. (Accessed June 2011. FDA News Release). http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/2009/ucm182057.htm, **2009**.
[297] Fung, E. T., A recipe for proteomics diagnostic test development: The OVA1 test, from biomarker discovery to FDA clearance. *Clin. Chem.* **2010**, *56*, 327-329.
[298] Anderson, N. L., The clinical plasma proteome: A survey of clinical assays for proteins in plasma and serum. *Clin. Chem.* **2010**, *56*, 177-185.
[299] Anderson, N. L., The roles of multiple proteomic platforms in a pipeline for new diagnostics. *Mol. Cell. Proteomics* **2005**, *4*, 1441-1444.
[300] Ransohoff, D. F., Bias as a threat to the validity of cancer molecular-marker research. *Nat. Rev. Cancer* **2005**, *5*, 142-149.
[301] Carr, S. A., Anderson, L., Protein quantitation through targeted mass spectrometry: The way out of biomarker purgatory? *Clin. Chem.* **2008**, *54*, 1749-1752.
[302] Hanash, S., Building a foundation for the human proteome: The role of the Human Proteome Organization. *J. Proteome Res.* **2004**, *3*, 197-199.
[303] Muthusamy, B., Hanumanthu, G., Suresh, S., Rekha, B., Srinivas, D., Karthick, L., . . . Pandey, A., Plasma proteome database as a resource for proteomics research. *Proteomics* **2005**, *5*, 3531-3536.
[304] Hancock, W., Omenn, G., LeGrain, P., Paik, Y.-K., Proteomics, human proteome project, and chromosomes. *J. Proteome Res.* **2011**, *10*, 210-210.
[305] Legrain, P., Aebersold, R., Archakov, A., Bairoch, A., Bala, K., Beretta, L., . . . Omenn, G. S., The human proteome project: Current state and future direction. *Mol. Cell. Proteomics* **2011**.
[306] Ahmed, F. E., Sample preparation and fractionation for proteome analysis and cancer biomarker discovery by mass spectrometry. *J. Sep. Sci.* **2009**, *32*, 771-798.
[307] Hoffman, S. A., Joo, W. A., Echan, L. A., Speicher, D. W., Higher dimensional (Hi-D) separation strategies dramatically improve the potential for cancer biomarker detection in serum and plasma. *J. Chromatogr. B Analyt. Technol. Biomed. Life. Sci.* **2007**, *849*, 43-52.
[308] Faca, V., Pitteri, S. J., Newcomb, L., Glukhova, V., Phanstiel, D., Krasnoselsky, A., . . . Hanash, S., Contribution of protein fractionation to depth of analysis of the serum and plasma proteomes. *J. Proteome Res.* **2007**, *6*, 3558-3565.
[309] Lee, H. J., Lee, E. Y., Kwon, M. S., Paik, Y. K., Biomarker discovery from the plasma proteome using multidimensional fractionation proteomics. *Curr. Opin. Chem. Biol.* **2006**, *10*, 42-49.
[310] Pernemalm, M., Orre, L. M., Lengqvist, J., Wikstrom, P., Lewensohn, R., Lehtio, J., Evaluation of three principally different intact protein prefractionation methods for plasma biomarker discovery. *J. Proteome Res.* **2008**, *7*, 2712-2722.
[311] Westermeier, R., Naven, T., Hopker, H.-R., in: Westermeier, R., Naven, T., Hopker, H.-R. (Eds.), *Proteomics in Practice*, Liquid Chromatography Techniques, Wiley-VCH, Weinheim **2008**, pp. 151-211.
[312] Echan, L. A., Tang, H. Y., Ali-Khan, N., Lee, K., Speicher, D. W., Depletion of multiple high-abundance proteins improves protein profiling capacities of human serum and plasma. *Proteomics* **2005**, *5*, 3292-3303.
[313] Pieper, R., Su, Q., Gatlin, C. L., Huang, S. T., Anderson, N. L., Steiner, S., Multi-component immunoaffinity subtraction chromatography: an innovative step towards a comprehensive survey of the human plasma proteome. *Proteomics* **2003**, *3*, 422-432.

[314] Polaskova, V., Kapur, A., Khan, A., Molloy, M. P., Baker, M. S., High-abundance protein depletion: Comparison of methods for human plasma biomarker discovery. *Electrophoresis* **2010**, *31*, 471-482.
[315] Qian, W. J., Kaleta, D. T., Petritis, B. O., Jiang, H., Liu, T., Zhang, X., . . . Smith, R. D., Enhanced detection of low abundance human plasma proteins using a tandem IgY12-SuperMix immunoaffinity separation strategy. *Mol. Cell. Proteomics* **2008**, *7*, 1963-1973.
[316] Shen, Y., Jacobs, J. M., Camp, D. G., 2nd, Fang, R., Moore, R. J., Smith, R. D., . . . Tompkins, R. G., Ultra-high-efficiency strong cation exchange LC/RPLC/MS/MS for high dynamic range characterization of the human plasma proteome. *Anal. Chem.* **2004**, *76*, 1134-1144.
[317] Aye, T. T., Scholten, A., Taouatas, N., Varro, A., Van Veen, T. A., Vos, M. A., Heck, A. J., Proteome-wide protein concentrations in the human heart. *Mol. Biosyst.* **2010**, *6*, 1917-1927.
[318] Tang, H. Y., Beer, L. A., Speicher, D. W., in: Simpson, R. J., Greening, D. W. (Eds.), *Methods in Molecular Biology Methods and Protocols: Serum/Plasma Proteomics Methods and Protocols*, In-Depth Analysis of a Plasma or Serum Proteome Using a 4D Protein Profiling Method, Springer Science **2011**, pp. 47-67.
[319] McCormack, A. L., Schieltz, D. M., Goode, B., Yang, S., Barnes, G., Drubin, D., Yates, J. R., 3rd, Direct analysis and identification of proteins in mixtures by LC/MS/MS and database searching at the low-femtomole level. *Anal. Chem.* **1997**, *69*, 767-776.
[320] Elias, J. E., Haas, W., Faherty, B. K., Gygi, S. P., Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations. *Nat. Methods* **2005**, *2*, 667-675.
[321] Gomez, F. A., Microfluidics in protein chromatography. *Methods Mol. Biol.* **2011**, *681*, 137-150.
[322] Chaurand, P., Stoeckli, M., Caprioli, R. M., Direct profiling of proteins in biological tissue sections by MALDI mass spectrometry. *Anal. Chem.* **1999**, *71*, 5263-5270.
[323] Caldwell, R. L., Caprioli, R. M., Tissue profiling by mass spectrometry. *Mol. Cell. Proteomics* **2005**, *4*, 394-401.
[324] Andersson, M., Groseclose, M. R., Deutch, A. Y., Caprioli, R. M., Imaging mass spectrometry of proteins and peptides: 3D volume reconstruction. *Nat Meth* **2008**, *5*, 101-108.
[325] Wolf-Yadlin, A., Hautaniemi, S., Lauffenburger, D. A., White, F. M., Multiple reaction monitoring for robust quantitative proteomic analysis of cellular signaling networks. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 5860-5865.
[326] Picotti, P., Bodenmiller, B., Mueller, L. N., Domon, B., Aebersold, R., Full dynamic range proteome analysis of S. cerevisiae by targeted proteomics. *Cell* **2009**, *138*, 795-806.
[327] Lange, V., Picotti, P., Domon, B., Aebersold, R., Selected reaction monitoring for quantitative proteomics: a tutorial. *Mol. Syst. Biol.* **2008**, *4*, 222.
[328] Anderson, L., Hunter, C. L., Quantitative mass spectrometric multiple reaction monitoring assays for major plasma proteins. *Mol. Cell. Proteomics* **2006**, *5*, 573-588.
[329] Kuzyk, M. A., Smith, D., Yang, J., Cross, T. J., Jackson, A. M., Hardie, D. B., . . . Borchers, C. H., Multiple reaction monitoring-based, multiplexed, absolute quantitation of 45 proteins in human plasma. *Mol. Cell. Proteomics* **2009**, *8*, 1860-1877.
[330] Pinkse, M. W., Uitto, P. M., Hilhorst, M. J., Ooms, B., Heck, A. J., Selective isolation at the femtomole level of phosphopeptides from proteolytic digests using

2D-NanoLC-ESI-MS/MS and titanium oxide precolumns. *Anal. Chem.* **2004**, *76*, 3935-3943.
[331] Zhang, H., Li, X. J., Martin, D. B., Aebersold, R., Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nat. Biotechnol.* **2003**, *21*, 660-666.
[332] Stahl-Zeng, J., Lange, V., Ossola, R., Eckhardt, K., Krek, W., Aebersold, R., Domon, B., High sensitivity detection of plasma proteins by multiple reaction monitoring of N-glycosites. *Mol. Cell. Proteomics* **2007**, *6*, 1809-1817.
[333] Job, D., Haynes, P. A., Zivy, M., Plant proteomics. *Proteomics* **2011**, *11*, 1557-1558.
[334] Salinger, M., Climate variability and change: Past, present and future, an overview. *Climatic Change* **2005**, *70*, 9-29.
[335] Jorrin-Novo, J. V., Maldonado, A. M., Echevarria-Zomeno, S., Valledor, L., Castillejo, M. A., Curto, M., . . . Redondo, I., Plant proteomics update (2007-2008): Second-generation proteomic techniques, an appropriate experimental design, and data analysis to fulfill MIAPE standards, increase plant proteome coverage and expand biological knowledge. *J. Proteomics* **2009**, *72*, 285-314.
[336] Neilson, K. A., Gammulla, C. G., Mirzaei, M., Imin, N., Haynes, P. A., Proteomic analysis of temperature stress in plants. *Proteomics* **2010**, *10*, 828-845.
[337] Farrell, T. C., Fox, K. M., Williams, R. L., Fukai, S., Genotypic variation for cold tolerance during reproductive development in rice: Screening with cold air and cold water. *Field Crops Research* **2006**, *98*, 178-194.
[338] Imin, N., Kerim, T., Rolfe, B. G., Weinman, J. J., Effect of early cold stress on the maturation of rice anthers. *Proteomics* **2004**, *4*, 1873-1882.
[339] Boyer, J. S., Westgate, M. E., Grain yields with limited water. *J Exp Bot* **2004**, *55*, 2385-2394.
[340] Prohens, J., Fita, A. M., Rodriguez-Burruezo, A. n., Raigon, M. a. D., Plazas, M., Vilanova, S., Breeding for the present and the future: Achievements and constraints of conventional plant breeding and contributions of genomics to a new green revolution. *Bulletin UASVM Horticulture* **2011**, *68*, 26-33.
[341] Hedden, P., Constructing dwarf rice. *Nat Biotech* **2003**, *21*, 873-874.
[342] Evans, L. T., *Feeding the Ten Billion. Plant and Population Growth*, Cambridge Univ. Press, Cambridge **1998**.
[343] Gowda, M., Venu, R. C., Roopalakshmi, K., Sreerekha, M. V., Kulkarni, R. S., Advances in rice breeding, genetics and genomics. *Mol. Breeding* **2003**, *11*, 337-352.
[344] Parry, S., Dupré, J., Introducing nature after the genome. *Sociological Rev.* **2010**, *58*, 3-16.
[345] Weale, A., Ethical arguments relevant to the use of GM crops. *New Biotechnology* **2010**, *27*, 582-587.
[346] Goff, S. A., Ricke, D., Lan, T. H., Presting, G., Wang, R., Dunn, M., . . . Briggs, S., A draft sequence of the rice genome (*Oryza sativa L. ssp. japonica*). *Science* **2002**, *296*, 92-100.
[347] Normile, D., Pennisi, E., Rice: Boiled down to bare essentials. *Science* **2002**, *296*, 32-36.
[348] Ito, Y., Katsura, K., Maruyama, K., Taji, T., Kobayashi, M., Seki, M., . . . Yamaguchi-Shinozaki, K., Functional analysis of rice DREB1/CBF-type transcription factors involved in cold-responsive gene expression in transgenic rice. *Plant Cell Physiol.* **2006**, *47*, 141-153.
[349] Nakashima, K., Tran, L. S., Van Nguyen, D., Fujita, M., Maruyama, K., Todaka, D., . . . Yamaguchi-Shinozaki, K., Functional analysis of a NAC-type transcription factor OsNAC6 involved in abiotic and biotic stress-responsive gene

expression in rice. *The Plant Journal: for cell and molecular biology* **2007**, *51*, 617-630.
[350] Zhao, B., Liang, R., Ge, L., Li, W., Xiao, H., Lin, H., . . . Jin, Y., Identification of drought-induced microRNAs in rice. *Biochem. Biophys. Res. Commun.* **2007**, *354*, 585-590.
[351] Zhou, J., Wang, X., Jiao, Y., Qin, Y., Liu, X., He, K., . . . Deng, X. W., Global genome expression analysis of rice in response to drought and high-salinity stresses in shoot, flag leaf, and panicle. *Plant Mol. Biol.* **2007**, *63*, 591-608.
[352] Lee, D. G., Ahsan, N., Lee, S. H., Kang, K. Y., Bahk, J. D., Lee, I. J., Lee, B. H., A proteomic approach in analyzing heat-responsive proteins in rice leaves. *Proteomics* **2007**, *7*, 3369-3383.
[353] Ali, G. M., Komatsu, S., Proteomic analysis of rice leaf sheath during drought stress. *J. Proteome Res.* **2006**, *5*, 396-403.
[354] Efeoglu, B., Terzioglu, S., Varying patterns of protein synthesis in bread wheat during heat shock. *Acta Biol. Hung.* **2007**, *58*, 93-104.
[355] Hajheidari, M., Eivazi, A., Buchanan, B. B., Wong, J. H., Majidi, I., Salekdeh, G. H., Proteomics uncovers a role for redox in drought tolerance in wheat. *J. Proteome Res.* **2007**, *6*, 1451-1460.
[356] Lin, S. K., Chang, M. C., Tsai, Y. G., Lur, H. S., Proteomic analysis of the expression of proteins related to rice quality during caryopsis development and the effect of high temperature on expression. *Proteomics* **2005**, *5*, 2140-2156.
[357] Salekdeh, G. H., Siopongco, J., Wade, L. J., Ghareyazie, B., Bennett, J., Proteomic analysis of rice leaves during drought stress and recovery. *Proteomics* **2002**, *2*, 1131-1145.
[358] Yan, S. P., Zhang, Q. Y., Tang, Z. C., Su, W. A., Sun, W. N., Comparative proteomic analysis provides new insights into chilling stress responses in rice. *Mol. Cell. Proteomics* **2006**, *5*, 484-496.
[359] Yamaguchi, T., Nakayama, K., Hayashi, T., Yazaki, J., Kishimoto, N., Kikuchi, S., Koike, S., cDNA microarray analysis of rice anther genes under chilling stress at the microsporogenesis stage revealed two genes with DNA transposon castaway in the 5'-flanking region. *Biosci. Biotechnol. Biochem.* **2004**, *68*, 1315-1323.
[360] Yamaguchi, T., Blumwald, E., Developing salt-tolerant crop plants: Challenges and opportunities. *Trends Plant Sci.* **2005**, *10*, 615-620.
[361] Agrawal, G. K., Jwa, N. S., Rakwal, R., Rice proteomics: Ending phase I and the beginning of phase II. *Proteomics* **2009**, *9*, 935-963.
[362] Imin, N., Kerim, T., Weinman, J. J., Rolfe, B. G., Low temperature treatment at the young microspore stage induces protein changes in rice anthers. *Mol. Cell. Proteomics* **2006**, *5*, 274-292.
[363] Kocher, T., Superti-Furga, G., Mass spectrometry-based functional proteomics: from molecular machines to protein networks. *Nat. Methods* **2007**, *4*, 807-815.
[364] Gingras, A. C., Gstaiger, M., Raught, B., Aebersold, R., Analysis of protein complexes using mass spectrometry. *Nat. Rev. Mol. Cell Biol.* **2007**, *8*, 645-654.
[365] Agrawal, G. K., Bourguignon, J., Rolland, N., Ephritikhine, G., Ferro, M., Jaquinod, M., . . . Rakwal, R., Plant organelle proteomics: Collaborating for optimal cell function. *Mass Spectrom. Rev.* **2011**, *30*, 772-853.
[366] Walther, T. C., Mann, M., Mass spectrometry-based proteomics in cell biology. *JCB* **2010**, *190*, 491-500.
[367] Ekström, S., Ericsson, D., Önnerfjord, P., Bengtsson, M., Nilsson, J., Marko-Varga, G., Laurell, T., Signal amplification using “Spot-on-a-Chip” technology for the identification of proteins via MALDI-TOF MS. *Anal. Chem.* **2000**, *73*, 214-219.

[368] Suckau, D., Resemann, A., Schuerenberg, M., Hufnagel, P., Franzen, J., Holle, A., A novel MALDI LIFT-TOF/TOF mass spectrometer for proteomics. *Anal. Bioanal. Chem.* **2003**, *376*, 952-965.
[369] Zhang, X., Shi, L., Shu, S., Wang, Y., Zhao, K., Xu, N., . . . Roepstorff, P., An improved method of sample preparation on AnchorChip™ targets for MALDI-MS and MS/MS and its application in the liver proteome project. *Proteomics* **2007**, *7*, 2340-2349.
[370] Tholey, A., Heinzle, E., Ionic (liquid) matrices for matrix-assisted laser desorption/ionization mass spectrometry, applications and perspectives. *Anal. Bioanal. Chem.* **2006**, *386*, 24-37.
[371] Sjödahl, J., Kempka, M., Hermansson, K., Thorsén, A., Roeraade, J., Chip with twin anchors for reduced ion suppression and improved mass accuracy in MALDI-TOF mass spectrometry. *Anal. Chem.* **2005**, *77*, 827-832.
[372] Su, J., Mrksich, M., Using MALDI-TOF mass spectrometry to characterize interfacial reactions on self-assembled monolayers. *Langmuir* **2003**, *19*, 4867-4870.
[373] Brash, J. L., Horbett, T. A., in: Horbett, T. A., Brash, j. L. (Eds.), *Proteins at Interfaces II*, Proteins at Interfaces: An Overview, American Chemical Society, Washinton DC **1995**, pp. 1-23.
[374] Huang, T. T., Mosier, N. S., Ladisch, M. R., Surface engineering of microchannel walls for protein separation and directed microfluidic flow. *J. Sep. Sci.* **2006**, *29*, 1733-1742.
[375] Wolde, P. R. t., Frenkel, D., Enhancement of protein crystal nucleation by critical density fluctuations. *Science* **1997**, *277*, 1975-1978.
[376] Demtröder, W., *Atoms, Molecules and Photons: An Introduction to Atomic-, Molecular- and Quantum Physics*, Springer-Verlag, Berlin Heidelberg **2010**.
[377] Cottrell, J. S., Protein identification using MS/MS data. *J. Proteomics* **2011**, *74*, 1842-1851.
[378] Lam, H., Deutsch, E. W., Eddes, J. S., Eng, J. K., Stein, S. E., Aebersold, R., Building consensus spectral libraries for peptide identification in proteomics. *Nat Meth* **2008**, *5*, 873-875.
[379] Zhang, B., Chambers, M. C., Tabb, D. L., Proteomic parsimony through bipartite graph analysis improves accuracy and transparency. *J. Proteome Res.* **2007**, *6*, 3549-3557.
[380] Alves, P., Arnold, R. J., Novotny, M. V., Radivojac, P., Reilly, J. P., Tang, H., Advancement in protein inference from shotgun proteomics using peptide detectability. *Pac. Symp. Biocomput.* **2007**, 409-420.
[381] Hilario, M., Kalousis, A., Approaches to dimensionality reduction in proteomic biomarker studies. *Brief. Bioinform.* **2008**, *9*, 102-118.
[382] Apweiler, R., Aslanidis, C., Deufel, T., Gerstner, A., Hansen, J., Hochstrasser, D., . . . Schmitz, G., Approaching clinical proteomics: Current state and future fields of application in fluid proteomics. *Clin. Chem. Lab. Med.* **2009**, *47*, 724-744.
[383] Maurya, P., Meleady, P., Dowling, P., Clynes, M., Proteomic approaches for serum biomarker discovery in cancer. *Anticancer Res.* **2007**, *27*, 1247-1255.
[384] Wolters, D. A., Washburn, M. P., Yates, J. R., 3rd, An automated multidimensional protein identification technology for shotgun proteomics. *Anal. Chem.* **2001**, *73*, 5683-5690.
[385] Schaller, J., Gerber, S., Kampfer, U., Lejon, S., Trachsel, C., *Human Blood Plasma Proteins Structure and Function*, John Wiley and Sons, West Sussex, England **2008**.
[386] Peng, J., Elias, J. E., Thoreen, C. C., Licklider, L. J., Gygi, S. P., Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-

MS/MS) for large-scale protein analysis: the yeast proteome. *J. Proteome Res.* **2003**, *2*, 43-50.
[387] Oliveros, J. C., VENNY: An interactive tool for comparing lists with venn diagrams. http://bioinfogp.cnb.csic.es/tools/venny/index.html, **2007**.
[388] Greenough, C., Jenkins, R. E., Kitteringham, N. R., Pirmohamed, M., Park, B. K., Pennington, S. R., A method for the rapid depletion of albumin and immunoglobulin from human plasma. *Proteomics* **2004**, *4*, 3107-3111.
[389] Doucette, A. A., Tran, J. C., Wall, M. J., Fitzsimmons, S., Intact proteome fractionation strategies compatible with mass spectrometry. *Expert Rev. Proteomics* **2011**, *8*, 787-800.
[390] Choi, H., Nesvizhskii, A. I., False discovery rates and related statistical concepts in mass spectrometry-based proteomics. *J. Proteome Res.* **2007**, *7*, 47-50.
[391] Gammulla, C. G., Pascovici, D., Atwell, B. J., Haynes, P. A., Differential proteomic response of rice (*Oryza sativa*) leaves exposed to high- and low-temperature stress. *Proteomics* **2011**, *11*, 2839-2850.
[392] Mirzaei, M., Pascovici, D., Keighley, T., George, I., Voelckel, C., Heenan, P. B., Haynes, P. A., Shotgun proteomic profiling of five species of New Zealand *Pachycladon*. *Proteomics* **2011**, *11*, 166-171.
[393] Aebersold, R., Goodlett, D. R., Mass Spectrometry in Proteomics. *Chem. Rev.* **2001**, *101*, 269-296.
[394] Nesvizhskii, A. I., Vitek, O., Aebersold, R., Analysis and validation of proteomic data generated by tandem mass spectrometry. *Nat Meth* **2007**, *4*, 787-797.
[395] Domon, B., Aebersold, R., Options and considerations when selecting a quantitative proteomics strategy. *Nat Biotech* **2010**, *28*, 710-721.
[396] Olinares, P. D., Ponnala, L., van Wijk, K. J., Megadalton complexes in the chloroplast stroma of Arabidopsis thaliana characterized by size exclusion chromatography, mass spectrometry, and hierarchical clustering. *Mol. Cell. Proteomics* **2010**, *9*, 1594-1615.
[397] Lowenthal, M. S., Mehta, A. I., Frogale, K., Bandle, R. W., Araujo, R. P., Hood, B. L., . . . Liotta, L. A., Analysis of albumin-associated peptides and proteins from ovarian cancer patients. *Clin. Chem.* **2005**, *51*, 1933-1945.
[398] Lasker, K., Phillips, J. L., Russel, D., Velazquez-Muriel, J., Schneidman-Duhovny, D., Tjioe, E., . . . Sali, A., Integrative structure modeling of macromolecular assemblies from proteomics data. *Mol. Cell. Proteomics* **2010**, *9*, 1689-1702.
[399] Etzioni, R., Urban, N., Ramsey, S., McIntosh, M., Schwartz, S., Reid, B., . . . Hartwell, L., The case for early detection. *Nat. Rev. Cancer* **2003**, *3*, 243-252.
[400] O'Reilly, M. S., Pirie-Shepherd, S., Lane, W. S., Folkman, J., Antiangiogenic activity of the cleaved conformation of the serpin antithrombin. *Science* **1999**, *285*, 1926-1928.
[401] Folkman, J., Angiogenesis and apoptosis. *Semin. Cancer Biol.* **2003**, *13*, 159-167.
[402] Naumov, G. N., Akslen, L. A., Folkman, J., Role of angiogenesis in human tumor dormancy: animal models of the angiogenic switch. *Cell Cycle* **2006**, *5*, 1779-1787.
[403] Roth, M. J., Parks, B. A., Ferguson, J. T., Boyne, M. T., 2nd, Kelleher, N. L., "Proteotyping": population proteomics of human leukocytes using top down mass spectrometry. *Anal. Chem.* **2008**, *80*, 2857-2866.
[404] Jones, A. M., Bennett, M. H., Mansfield, J. W., Grant, M., Analysis of the defence phosphoproteome of Arabidopsis thaliana using differential mass tagging. *Proteomics* **2006**, *6*, 4155-4165.

[405] Rudella, A., Friso, G., Alonso, J. M., Ecker, J. R., van Wijk, K. J., Downregulation of ClpR2 leads to reduced accumulation of the ClpPRS protease complex and defects in chloroplast biogenesis in Arabidopsis. *The Plant Cell* **2006**, *18*, 1704-1721.
[406] Jorrin, J. V., Maldonado, A. M., Castillejo, M. A., Plant proteome analysis: a 2004-2006 update. *Proteomics* **2006**, *6*, 5529-5548.
[407] Chen, X., Sun, L., Yu, Y., Xue, Y., Yang, P., Amino acid-coded tagging approaches in quantitative proteomics. *Expert Rev. Proteomics* **2007**, *4*, 25-37.
[408] Xie, B., Patton, W. F., Parman, C. E., in: Wang, F. (Ed.), *Biomarker Methods in Drug Discovery and Development*, Top-Down quantitative proteomic analysis using a highly multiplexed isobaric mass tagging strategy, Humana Press **2008**, pp. 231-249.
[409] Breci, L., Haynes, P. A., in: Thiellement, H., Zivy, M., Damerval, C., Mechin, V. (Eds.), *Methods in Molecular Biology - Plant Proteomics Methods and Protocols*, Two-Dimensional nanoflow liquid chromatography-tandem mass spectrometry of proteins extracted from rice leaves and roots, Human Press, Totowa, New Jersey **2006**, pp. 249-266.
[410] Barsnes, H., Vizcaino, J. A., Eidhammer, I., Martens, L., PRIDE Converter: making proteomics data-sharing easy. *Nat. Biotechnol.* **2009**, *27*, 598-599.
[411] R Development Core Team, R: A language and environment for statistical computing. *R Foundation for Statistical Computing* **2008**.
[412] Shilov, I. V., Seymour, S. L., Patel, A. A., Loboda, A., Tang, W. H., Keating, S. P., . . . Schaeffer, D. A., The Paragon Algorithm, a next generation search engine that uses sequence temperature values and feature probabilities to identify peptides from tandem mass spectra. *Mol. Cell. Proteomics* **2007**, *6*, 1638-1655.
[413] Ye, J., Fang, L., Zheng, H., Zhang, Y., Chen, J., Zhang, Z., . . . Bolund, L., WEGO: a web tool for plotting GO annotations. *Nucleic Acids Res.* **2006**, *34*, W293-297.
[414] Keshamouni, V. G., Michailidis, G., Grasso, C. S., Anthwal, S., Strahler, J. R., Walker, A., . . . Omenn, G. S., Differential protein expression profiling by iTRAQ-2DLC-MS/MS of lung cancer cells undergoing epithelial-mesenchymal transition reveals a migratory/invasive phenotype. *J. Proteome Res.* **2006**, *5*, 1143-1154.
[415] Ali, N. A., McKay, M. J., Molloy, M. P., Proteomics of Smad4 regulated transforming growth factor-beta signalling in colon cancer cells. *Mol. Biosyst.* **2010**, *6*, 2332-2338.
[416] Cui, S., Huang, F., Wang, J., Ma, X., Cheng, Y., Liu, J., A proteomic analysis of cold stress responses in rice seedlings. *Proteomics* **2005**, *5*, 3162-3172.
[417] Hashimoto, M., Komatsu, S., Proteomic analysis of rice seedlings during cold stress. *Proteomics* **2007**, *7*, 1293-1302.
[418] Lee, D. G., Ahsan, N., Lee, S. H., Kang, K. Y., Lee, J. J., Lee, B. H., An approach to identify cold-induced low-abundant proteins in rice leaf. *C. R. Biol.* **2007**, *330*, 215-225.
[419] Komatsu, S., Yamada, E., Furukawa, K., Cold stress changes the concanavalin A-positive glycosylation pattern of proteins expressed in the basal parts of rice leaf sheaths. *Amino Acids* **2009**, *36*, 115-123.
[420] Komatsu, S., Yang, G., Khan, M., Onodera, H., Toki, S., Yamaguchi, M., Over-expression of calcium-dependent protein kinase 13 and calreticulin interacting protein 1 confers cold tolerance on rice plants. *Mol. Genet. Genomics* **2007**, *277*, 713-723.
[421] Kreps, J. A., Wu, Y., Chang, H. S., Zhu, T., Wang, X., Harper, J. F., Transcriptome changes for Arabidopsis in response to salt, osmotic, and cold stress. *Plant Physiol.* **2002**, *130*, 2129-2141.

[422] Chen, L. T., Luo, M., Wang, Y. Y., Wu, K., Involvement of Arabidopsis histone deacetylase HDA6 in ABA and salt stress response. *J Exp Bot* **2010**, *61*, 3345-3353.
[423] Kim, J. M., To, T. K., Nishioka, T., Seki, M., Chromatin regulation functions in plant abiotic stress responses. *Plant Cell Environ.* **2010**, *33*, 604-611.
[424] Wu, T., Yuan, T., Tsai, S. N., Wang, C., Sun, S. M., Lam, H. M., Ngai, S. M., Mass spectrometry analysis of the variants of histone H3 and H4 of soybean and their post-translational modifications. *BMC Plant Biol.* **2009**, *9*, 98.
[425] Zhu, J., Jeong, J. C., Zhu, Y., Sokolchik, I., Miyazaki, S., Zhu, J. K., . . . Bressan, R. A., Involvement of Arabidopsis HOS15 in histone deacetylation and cold tolerance. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 4945-4950.
[426] Titiz, O., Tambasco-Studart, M., Warzych, E., Apel, K., Amrhein, N., Laloi, C., Fitzpatrick, T. B., PDX1 is essential for vitamin B6 biosynthesis, development and stress tolerance in Arabidopsis. *The Plant Journal : for Cell and Molecular Biology* **2006**, *48*, 933-946.
[427] Tunc-Ozdemir, M., Miller, G., Song, L., Kim, J., Sodek, A., Koussevitzky, S., . . . Shintani, D., Thiamin confers enhanced tolerance to oxidative stress in Arabidopsis. *Plant Physiol.* **2009**, *151*, 421-432.
[428] Scafaro, A. P., Haynes, P. A., Atwell, B. J., Physiological and molecular changes in *Oryza meridionalis Ng*., a heat-tolerant species of wild rice. *J Exp Bot* **2010**, *61*, 191-202.
[429] Chen, H., Xiong, L., Pyridoxine is required for post-embryonic root development and tolerance to osmotic and oxidative stresses. *The Plant Journal : for Cell and Molecular Biology* **2005**, *44*, 396-408.
[430] Denslow, S. A., Rueschhoff, E. E., Daub, M. E., Regulation of the Arabidopsis thaliana vitamin B6 biosynthesis genes by abiotic stress. *Plant Physiol Biochem* **2007**, *45*, 152-161.
[431] Sancho Solis, R., Ge, Y., Walker, J., Single amino acid sequence polymorphisms in rat cardiac troponin revealed by top-down tandem mass spectrometry. *J. Muscle Res. Cell Motil.* **2008**, *29*, 203-212.

APPENDICES

APPENDIX I PUBLICATION 1 ARISING FROM THIS THESIS

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

APPENDIX II PUBLICATION 2 ARISING FROM THIS THESIS

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