A quantitative mass spectrometry-based proteomic analysis of mammalian cell-lines, SH-SY5Y, *Homo sapiens* neuroblastoma cell line, and PC-12Adh, *Rattus norvegicus* pheochromocytoma cell line, under hypoxic stress

by

Jinal Patel

A thesis submitted to the Faculty of Graduate and Postdoctoral Affairs in partial fulfillment of the requirements for the degree of

Master of Science

in

Chemistry

Carleton University
Ottawa, Ontario

Your file Votre référence
ISBN: 978-0-494-91542-4

Our file Notre référence
ISBN: 978-0-494-91542-4

Abstract

A quantitative mass spectrometry–based proteomic workflow was applied to characterize the changes in the proteome of human neuroblastoma, SH-SY5Y cell line, rat pheochromocytoma, PC-12Adh cell line, and neuronally differentiated rat pheochromocytoma cell line after treatment to hypoxia for 24 hour. The optimization in the workflow proved to be success at increasing the dataset of identified proteins. A computational quantification algorithm was utilized to compute the relative amounts of labeled proteins in the hypoxia-treated sample and its appropriate control for each biological sample. Approximately 10% of the total identified proteins for each biological sample were differentially regulated. Amongst the differentially regulated proteins, approximately 70% of the population was observed to be up-regulated while the remaining was down-regulated. Analysis of the datasets using bioinformatic approaches indicated the preferentially synthesis of proteins involved in hypoxia adaptation processes.

Acknowledgements

Firstly, I would like to heartily thank my supervisor, Jeffrey Smith, for seeing the potential in me to undertake a mighty project. His guidance, patience and advice throughout the project have been valuable and educative. I would also like to extend my thanks to William Willmore, for culturing the cells used in this project, allowing access to equipment and instrument in his lab and providing valuable insight into the biochemistry aspect of the project.

The key component behind the success of the results in this work was a quantification script, written by Michel Dumontier. I am very grateful for his work and assistance. I would also like to extend my gratitude to James Green (parsing databases) and Dawn Jin (IPA analysis).

Besides, I would like to pass regards to the past and present members of group and the collaborators. It has been a wonderful time working and socializing with all of you. A specific thanks to Karl Wasslen for *obeying* my demands for the instrument time.

A special thanks to Ryan Girgrah for his motivation and relaxing banter. I would like to deeply thank Matthew Huebsch for being the most enlightening sounding board and a great friend. I also owe a deep gratitude to Shivani Shah for being a pillar of encouragement even from miles away.

I would also like to acknowledge the Petlikar family for their generosity and immense support since my arrival in Canada. I am most grateful of my parents for their unconditional love and support for my every choice and demand. The last honourable mention goes to my brother, Vikesh, for all his love, support and motivation.

Table of Contents

List of Tables

This is the List of Tables.

List of Figures

This is the List of Figures.

1 Chapter: Introduction

The beginning of the 21st century marked the completion of the sequencing of the human genome as well as the genomes of several other model organisms. Advances in high-throughput sequencing and microarray-expression technology have enabled the accumulation of vast amounts of genomic knowledge. The next challenge for researchers lies in decoding this knowledge in terms of coding region localization and regulation, as well as the prediction of gene function[1]. The translation of genotypic knowledge into particular phenotypes is not an easy task. Given the complexity of a biological system, the term 'phenotypes' refers to the observable traits of an organism as well as the various aspects of cellular and molecular organization[2].

The central dogma of molecular biology, shown in Figure 1.1, represents the flow of information from DNA (gene) via replication, transcription and translation. It dictates that proteins are the final product of genes; however, there is no strict linear relationship between a gene and its complement protein. Complement proteins undergo post-transcriptional and post-translational modifications which are not apparent from the DNA sequences[3]. Hence, gene-centered studies provide an incomplete view of biological systems.

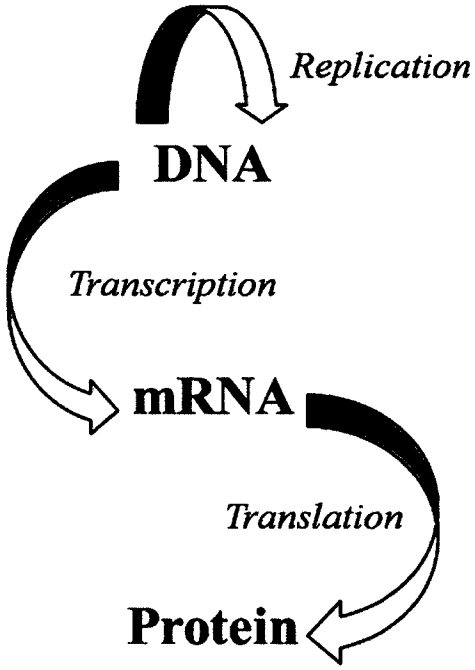


Figure 1-1: The central dogma of molecular biology.

The need for comprehensive knowledge of biological systems is driving attention towards the understanding of protein function and their interaction networks by proteomic analysis. The term, 'proteome', refers to the total complement of PROTEins expressed by the genOME of an organism or tissue and the large-scale study of these proteins is referred as 'proteomics'. Traditionally, proteomic analysis involved separation of proteins from a given cell by 2-dimensional gel electrophoresis (2-GE) and cataloguing the spots to create a database of the expressed proteins[3,4]. In the 1990's, mass spectrometry emerged as a powerful analytical method for analyzing proteins[3]. The invention and development of soft ionization techniques, ESI and MALDI, by 2002 Nobel Laureates[5], John Fenn[6] and Koichi Tanaka[7], were the pioneering basis of analyzing biomolecules using mass spectrometry. Advancements in mass spectrometry, in conjunction with new separation techniques and bioinformatic tools, has enabled higher

throughput analyses of proteins within complex samples. Presently, mass spectrometry-based proteomic workflows allow identification and quantitative profiling of proteomes and thus contribute to the understanding of the biochemistry of proteins, processes and pathways as well as the discovery of potential targets for pharmaceutical companies[3].

Aspects of protein biochemistry, as well as some fundamental aspects of mass spectrometry-based proteomics, are discussed in detail below, followed by a brief description of the biochemistry of oxidative stress (hypoxia) in mammalian cells.

1.1 Protein Biochemistry

Proteins are the functional units of all biological processes. They are biopolymers synthesized within cells consisting of amino acids that are covalently linked together via amide bonds (or peptide bonds). Amino acids, also known as residues, are compounds that contain an amino group, a carboxylic acid group and a functional group (or R-group or side chain) bonded to a central carbon atom, the C_α atom. In biological systems, all proteins are composed of 20 common amino acids, shown and classified in Figure 1.2. At physiological pH, the amino acids are zwitterionic compounds. The amino acids are classified into hydrophobic, uncharged and charged hydrophilic and 'special' based on the polarity of their functional group. The 3 'special' amino acids have unique properties: glycine has 2 H atoms bonded to the C_α atom, proline has a cyclic structure around the C_α atom and thus has a secondary amino group and lastly cysteine contains a thiol group that can form disulphide bonds with other cysteine residues in close proximity. With the exception of glycine, all amino acids have a chiral center around the C_α atom; hence

amino acids can have a D- or L- enantiomeric configuration. However, all amino acids in biological proteins have an L-stereochemical configuration.

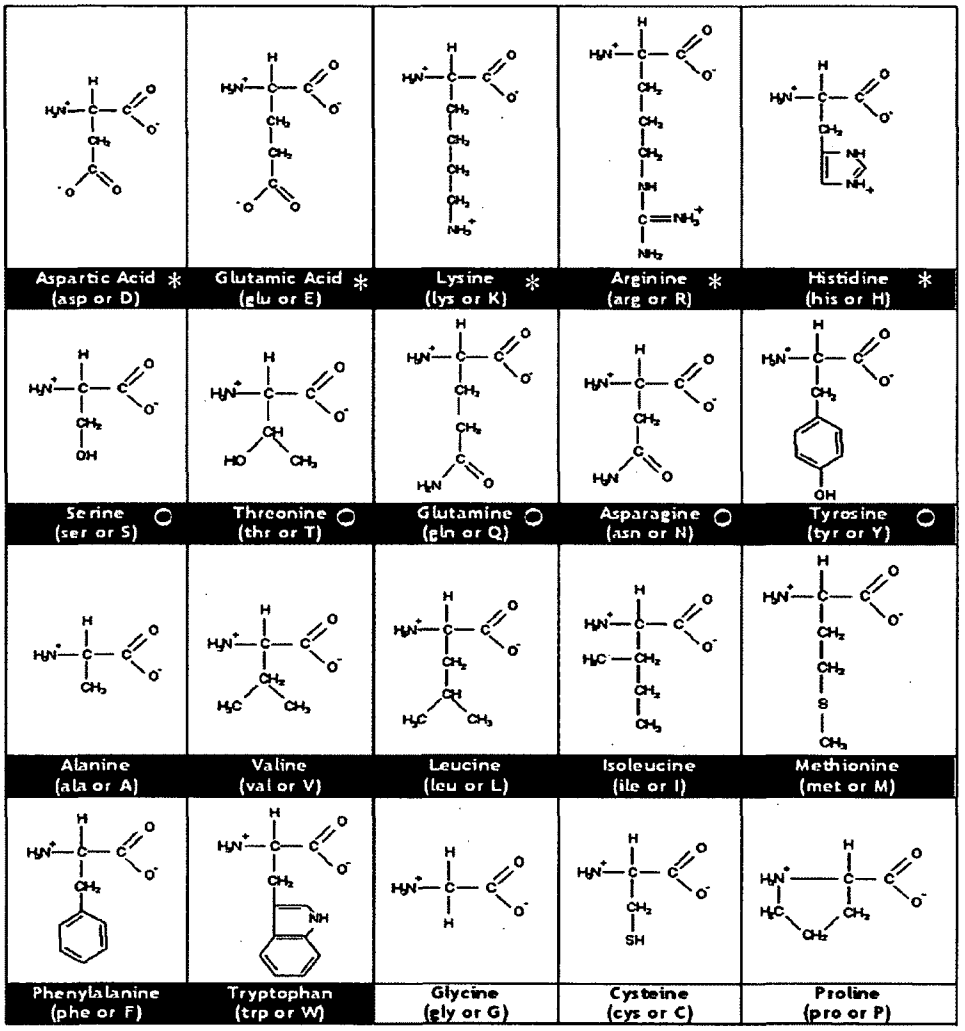


Figure 1-2: The 20 common amino acids found within a biological system. The hydrophilic and hydrophobic characteristics are attributed by the side chain of the amino acids and indicate by blue or red shading, respectively. The charged/uncharged state of hydrophilic residues is indicated by */○, respectively. The unique amino acids are represented by yellow shading. Adapted from Protein-UQDI[8].

The carboxylic group of an amino acid will react with an amino group on another amino acid by a condensation reaction to form a peptide bond between the two amino acid residues. The generated dipeptide (2 covalently-linked amino acids) would have two termini: the N-terminus and the C-terminus. The unbounded amino group and unbounded carboxylic group on a peptide are referred to as the N-terminus and C-terminus, respectively. As more amino acids residues are added to the peptide, a sequence of covalently linked amino acids is formed and referred to as the polypeptide chain (>10 covalently-linked amino acids).

The order of amino acids in a protein sequence is dictated by the order of deoxynucleotide bases in the DNA of a protein-coding gene. In brief, the double stranded- DNA for the gene of interest will unwind and serve as a template for RNA transcriptase. The 3' to 5' DNA segment will be transcribed into a messenger RNA (mRNA). The ribosome will read the mRNA three nucleotides (codon) at a time. Each codon has a corresponding amino acid, which is carried to the ribosome by a transfer RNA (tRNA) molecule. The tRNA-amino acid complexes will align along the mRNA, via base pairing, bringing sequential amino acids in close proximity to form a peptide link between the amino acids. Hence the triplet codes on the mRNA get translated into sequence of amino acids in a polypeptide. The polypeptide chain will fold up to generate a three-dimensional structure of the protein.

The three- dimensional structure of protein largely defines its properties and functions. A variety of human diseases (e.g. Alzheimer's and Parkinson's) have been associated with misfolding of proteins; misfolded proteins form aggregates that prove to

be fatal[9]. There are four basic levels of organization for protein structure, as shown in Figure 1.3.

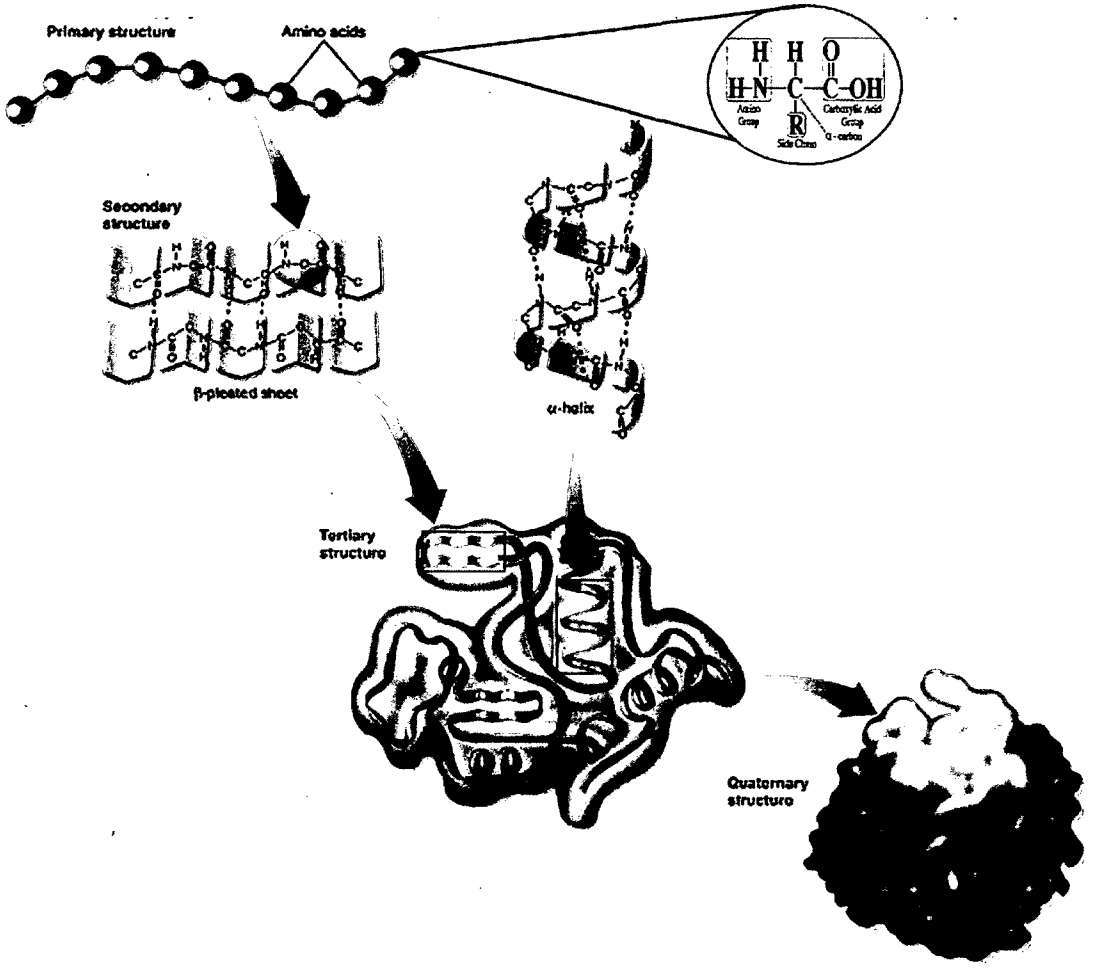


Figure 1-3: Four levels of structural organization observed within a mutlimeric protein. Adapted from Chemistry Images[10].

The primary structure (1°) of a protein is the linear amino acid sequence of its polypeptide chain(s). Conventionally, the polypeptide sequences are listed from N-terminus to C-terminus. The peptide bonds along the C_α atoms of the amino acid residues in the polypeptide chain form the protein backbone and the properties of functional

groups attached to the C_α atoms will dictate the secondary structure (2°) of the protein as well as tertiary (3°) and quaternary (4°). The hydrophobic amino acids will form the hydrophobic regions in the proteins whereas the hydrophilic amino acids will form hydrophilic region in the proteins. Typically, the core of a protein is made of hydrophobic regions and the surface is made of hydrophilic regions to maintain solubility of proteins in an aqueous environment.

The secondary structure of a protein defines the local conformation of the backbone: α-helices, β-sheets (pleated sheets) and turns. The α-helix structure is held together by hydrogen bonds between amide and carbonyl groups that are four residues apart along the polypeptide backbone. Each turn in an α-helix contains 3.6 residues. The functional groups of the amino acid residues are facing outwards in order to avoid steric interference with the backbone. The hydrophilic or hydrophobic properties of the outward facing functional groups will determine whether the α-helix is hydrophilic or hydrophobic. Pleated sheet or β-sheet is a planar rippled structure containing an average of 6 polypeptide strands with 5-8 amino acid residues. The β-sheet is held together by hydrogen bonds between the amide and carbonyl groups of parallel or antiparallel polypeptide strands. The hydrophilicity and hydrophobicity of the β-sheet is similar to α-helices. Lastly, turns or coils are short polypeptide segments that connect the α-helices and β-sheets within the protein as well as alter the direction of the polypeptide backbone (reverse turns/ β- bends). These structures are irregular and commonly consist of glycine and proline residues since they facilitate the required conformation.

Tertiary structure refers to the three-dimensional arrangement of protein secondary structure. This folding structure is based on the hydrogen bonds between the

side chains of the amino acids and their interactions with the surrounding environment. The tertiary structure has additional interactions to that of secondary structures. It contains covalent disulphide bridges, ionic interactions, hydrogen bonding and hydrophobic interactions. The tertiary structure would have distinct regions, called domains, where each domain has a distinct function (for example, a transmembrane protein would have a hydrophobic domain that allows it to be embedded into the membrane) or structure (for example, a β-barrel). Some proteins, once assembled into its 3° structure, will associate with other polypeptide subunits to form the multimeric structure, quaternary structure. A quaternary protein may consist of identical or non-identical subunits; for example, glutamine synthetase has 12 identical subunits (homomeric) whereas haemoglobin has 2α-units and 2β-units (heteromeric). Other groups (such as lipids, carbohydrates, metal ions, etc.) can be incorporated into the 4° structure to assist in its function e.g. heme groups in haemoglobin. In addition to the interactive forces described above, protein structure is also dependent on the physiological pH, temperature and hydration state. *In vivo*, chaperone proteins facilitate the correct folding of proteins at physiological conditions as well as the correction of misfolded proteins to avoid protein aggregation.

It is estimated that the human genomes comprises of 20,000-25,000 genes[11] while the number of proteins in the human proteome is estimated to be over 1 million[12]. This estimation suggests that a single gene is capable of encoding multiple proteins. The increase in complexity from gene to multiple proteins is facilitated by post-transcriptional and post-translational modifications. Transcription initiation at alternate promoters, differential transcriptional termination and alternative splicing of the mRNA generate

different mRNA transcripts[12]. Post-transcriptional modifications of the mRNA are only observed in eukaryotes while post-translational modifications are observed in prokaryotes and eukaryotes. Post-translational modifications (PTMs) are chemical modifications to the proteins which increase the functional diversity of the proteome[13]. Post-translational modifications can occur at any stage in the life cycle of the protein. They regulate the activity, localization and interaction of the protein with other cellular components. There are about 450 PTMs listed in the UniProt database of PTMs[14] and the most common PTMs include phosphorylation, glycosylation, ubiquitination, nitration, methylation and acetylation. The functions of these common PTMs are summarized in the Table 1.1 below:

Table 1-1: Some commonly observed post-translational modifications on biological proteins. The associated mass difference as seen in mass spectral data and the biological function controlled by the PTMs are also indicated. Adapted from Mann, *et al.*[15]

PTM type	ΔMass (Da)	Function
Phosphorylation (pSer, pThr, pTyr)	+80	Activation/inactivation of enzymes, signal transduction, cell apoptosis
Glycosylation (N- or O-linked)	+203, >800	Cell-cell recognition, structural, secretion
Ubiquitination	>1000	Protein destruction signaling
Methylation	+14	Regulation of gene expression
Acetylation	+42	Protein stabilization, regulation of protein-DNA interaction
Disulphide formation	-2	Protein stability

The PTMs indicated in the above table involve enzymatic addition of small moieties to specific amino acid residues or specific groups within the amino acid residues; phosphorylation will be discussed in detail below. However, another method in which proteins are modified is proteolysis, which plays an important role in activation of enzymes and degradation of misfolded and controlled proteins. Proteolysis is an

irreversible PTM and is controlled by proteases, which are highly regulated and compartmentalized protein-digesting enzymes. An example of proteolytic PTM involves cleavage of four arginine-glycine peptide bonds in the central globular region of the zymogen, fibrinogen, to result in fibrin monomers which aggregate in to fibrin arrays that form the network for blood clots[16]. All of the proteases are also controlled and regulated by proteolytic PTMs.

Protein phosphorylation is a reversible modification, which plays a key role in many signal transduction pathways in biological systems. It includes the addition of a phosphate moiety on the hydroxyl group of serine, threonine and tyrosine residues in a protein. Protein kinases catalyze the addition of the phosphate group from cellular adenosine triphosphate (ATP) molecules on to the residues. The removal of phosphate groups by hydrolysis is referred to as dephosphorylation and is catalyzed by protein phosphatases. In humans, it is estimated that approximately 30% of the proteome is phosphorylated at a given time[17]. Over half of all cellular proteins are phosphorylated during their lifetime and phosphoproteins (phosphorylated proteins) may contain multiple phosphorylation sites (a total >100,000 sites)[18]. There are 518 known protein kinases and 65 known protein phosphatases in the human proteome[19] thus suggesting a shared homology at the site of phosphorylation between different phosphoproteins[20]. Abnormal phosphorylation and dephosphorylation have been shown to be cause or consequence of a number of human diseases such as cancer and diabetes[17]. Therefore, attention is currently being driven towards discovering pharmaceutical drugs that inhibit a particular protein kinase, phosphatase or one of their substrates that is behaving aberrantly[17, 20].

1.2 Mass Spectrometry-based Proteomics

As previously mentioned, the invention and discovery of soft ionization techniques was the founding principle of mass spectrometry-based protein analysis. With these soft ionization techniques, intact proteins and peptides could be ionized and subsequently sequenced based on their fragmentation patterns using mass spectrometry (and, more recently, with the help of bioinformatics tools); the principles and instrumentation will be further emphasized below. Prior to these discoveries, Edman degradation was the only option for sequencing proteins, more specifically peptides. In brief, Edman degradation involves isolation of a peptide, followed by the cleavage of one amino acid, extracting the cleaved complex into an organic solvent and analyzing the product using HPLC. Edman degradation is still a preferred and reliable technique to sequence an unknown protein. However, it is time-consuming, has low throughput and requires complete isolation of every peptide from its digested protein. Therefore it is not an ideal choice for protein sequencing or identification from large proteomic samples. As a result of the development of soft-ionization techniques, as well as the availability of genome sequence knowledge to form comprehensive proteome databases, mass spectrometry has become the method of choice for sequencing proteins. The advantages of protein sequencing using mass spectrometry are the rapid speed of analysis, high sensitivity, ability to generate large amounts of information per experiment and ability to characterize post-translational modifications.

There are five key steps to proteomic analysis by mass spectrometry: Sample preparation, separation, ionization, mass spectrometry and bioinformatics. In brief, sample preparation involves generation of multiple copies of the protein source (e.g. cells

or tissue or biological fluids) and extraction of the desired proteins or the entire proteome sample. After obtaining the sample, its complexity could be reduced by various separation techniques e.g. electrophoresis or liquid chromatography. The separated sample is then ionized prior to analysis by a mass spectrometer to measure the mass to charge ratios of individual fragments. The spectral data collected for the fragments is compared to databases using bioinformatic tools to identify the proteins in the sample. Each of these steps will be iterated in depth in the following sub-sections.

The specifics of these steps are dependent on the type of approach to MS-based proteomic sequencing: top-down sequencing or bottom-up sequencing. The top-down sequencing approach involves ionization of intact proteins, introduction of the molecular ions into the mass spectrometer and subjection to gas-phase fragmentation. This approach is claimed to be capable of complete protein sequencing and identifying protein isoforms and PTMs[21]. The favoured instruments for this strategy are FT-ICR[22], hybrid ion trap FT-ICR or hybrid ion trap–orbitrap[23] which are expensive to purchase and operate; the favoured dissociation techniques are ECD and ETD which have low-efficiency and require long ion accumulation and detection times[22]. The major pitfalls for top-down approach are that the high molecular mass proteins (>100kDa) and highly hydrophobic proteins are usually unsolved and the interpretation of the complex spectral data generated by multiply-charged proteins is difficult and limited by lack of availability of protein identification tools[21]. The complexity of the spectral data could be reduced by fractionation, however there is a need for improved methods to separate intact proteins[24].

In bottom-up sequencing approach, the proteins are enzymatically cleaved into peptides, separated and then introduced into the mass spectrometer. One of the prominent

peptide separation techniques used in this approach is multidimensional liquid chromatography which results in high resolution orthogonal peptide separations and thus successful identification of proteins in the complex biological samples[25]. More so, multiple quantitative techniques have been developed for peptide analysis using this approach[26]. The bottom-up approach is a more mature and widely used sequencing approach and hence there are a variety of instruments such as 3D and linear ion traps, hybrid quadrupole-TOF, and TOF-TOF mass spectrometers available to use in this strategy[23]. However, the fundamental limitation of this approach is that only a fraction of the total peptides of a given protein are identified and this may lead to loss in identification of PTMs, particularly for peptides of low abundance or poor ionization efficiency[23].

Both of these approaches are complementary and have their individual strengths and weakness; however, the bottom-up strategy has proven to be a well-optimized strategy in MS-based proteomics and is the preferred approach for the vast majority of MS-based proteomic studies, including the work presented here. Therefore the five key steps for proteomic analysis iterated below: mass spectrometry, sample preparation, separation and peptide and protein identification and quantification using bioinformatic tools, will be based on the bottom-up strategy.

1.2.1 Mass Spectrometry

Mass spectrometry is a powerful analytical tool for identifying unknown compounds, quantifying known compounds and elucidating the structural properties of compounds. Originally, mass spectrometry was prominent among physicists and chemists

but the invention of soft ionization techniques has extended the field into biological science. The specifics and principles discussed in this section will be focused on biological application of mass spectrometry.

Mass spectrometers determine the mass to charge ratio of a molecule based upon the motion of the charged molecule in an electric or magnetic field. By convention, m is the numerical value for the mass in terms of atomic mass unit (amu) or Dalton (Da) and z is the integer equal to the number of electrons lost/protons gained (positive ions) or electrons gained/protons lost (negative ions). The unit for mass-to-charge ratio is Thomson (Th)[27]. The trajectories of ions are dependent on each ion's mass to charge ratio which is tightly regulated by and mathematically related to precise electric or magnetic fields within the instrument. Scanning the m/z ratios in a given experiment and gathering their intensity will create a mass spectrum, shown in Figure 1.4.

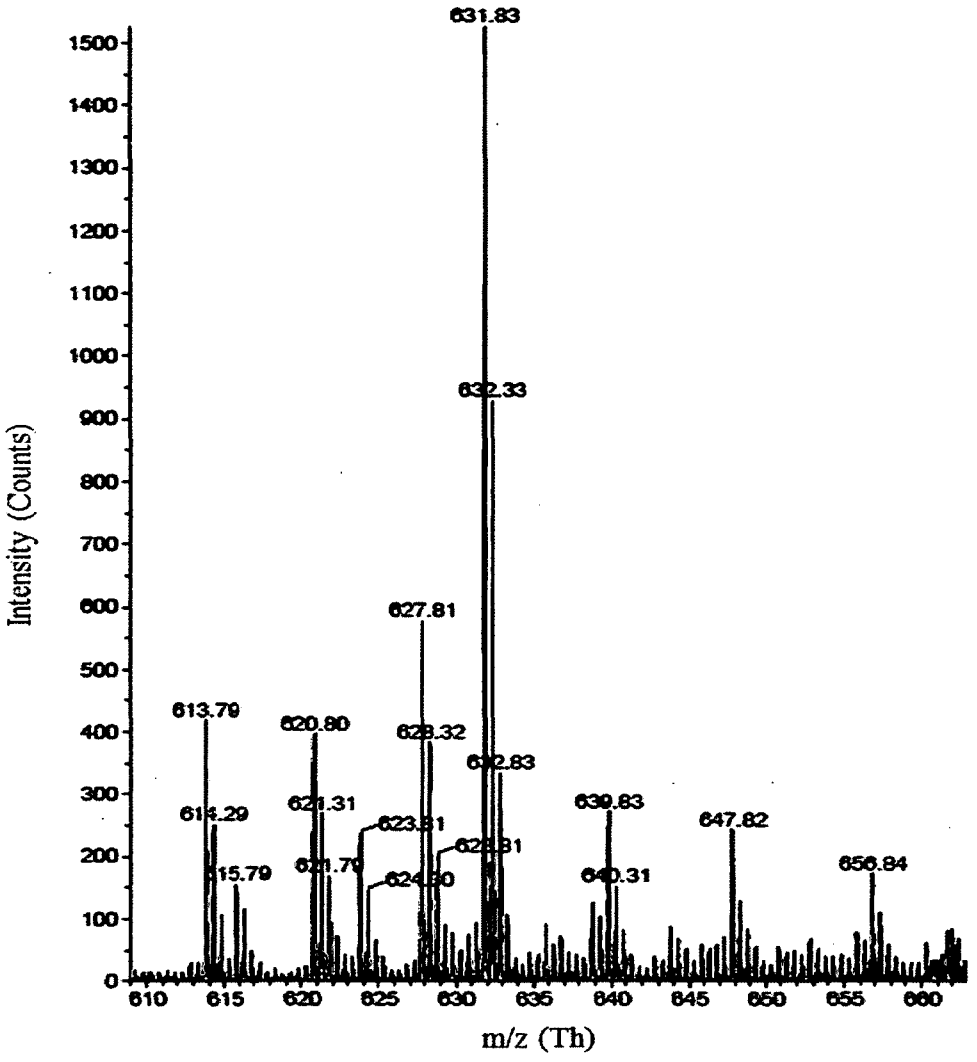


Figure 1-4: A mass spectrum showing the intensity (counts) versus the m/z values of ions.

All mass spectrometers consist of 3 main components: an ionization source, mass analyzer(s) and a detector. Each of these components will be mainly discussed in the context of a specific instrument, an electrospray quadrupole-time-of-flight mass spectrometer. The basis of mass spectrometry is the separation of ions in the gas-phase; however, not all compounds are volatile or ionic and hence an ionization source is necessary. There are a variety of ionization techniques used in mass spectrometry. The choice of the ionization technique is dependent on the class of molecules such as electron

ionization (EI) and chemical ionization (CI) for volatile organic molecules, atmospheric pressure chemical ionization (API) for less-polar molecules and electrospray (ESI) and matrix-assisted laser desorption/ionization (MALDI) for polar non-volatile biomolecules.

Following formation, ions are accelerated into the one or more mass analyzers where they may be separated in space (ion trap), in time (time-of-flight) or as a mass filter (quadrupole) based on their mass-to-charge ratios (m/z). The choice of mass analyzer depends on the resolution, sensitivity, mass range and scan rate required for an experiment. Resolution or resolving power, R, refers to the separation of two ions and is given by the equation

$$R = \frac{m}{\Delta m}$$

where m is the m/z ratio of the ion and Δm is the full width of its peak in the mass spectrum at half of its maximum intensity (FWHM). Therefore, greater resolution is required as the mass difference between ions decreases. The mass range refers to highest m/z ratio transmitted by the mass analyzer and scan rate refers to the rate at which a mass spectrum is scanned. Ideally, a mass spectrometer should have high resolution, broad mass range and a fast scan rate.

After separation, a detector is used to convert the separated ions into an electric signal which is then amplified and displayed in form of a mass spectrum. A detector is chosen based on their response time, dynamic range, gain and geometry. The common detectors currently used are electron multiplier, channeltrons and multichannel plates.

Mass spectrometers are maintained under high vacuum allowing ion transmission without collision or interaction with other ions. Ion-molecule collisions or interactions can lead to spectral ambiguity and decreased sensitivity and resolution. Most instruments

are maintained at 10^{-4} to 10^{-8} torr (10^{-2} to 10^{-5} Pa) and these low pressures are most commonly generated using two types of pumps. At the first stage, a mechanical pump will provide a rough vacuum down to 10^{-3} torr (0.1Pa). And at the second stage, the high vacuum is generated by a diffusion or turbomolecular pump.

The work in this thesis was conducted with a quadrupole time-of-flight mass spectrometer equipped with an electrospray ionization source. The principles and comprehensive view on its functioning is presented in the following sections.

1.2.1.1 Electrospray Ionization

Electrospray ionization (ESI) is a commonly used ionization technique in biochemistry and medical sciences. It was invented by M. Dole[28] and was adapted for biological molecules by J. Fenn[6]. Figure 1.5 shows a schematic of the essential principles of ESI in the formation of positive ions, however negative ESI is possible with reversal of the voltage. A large voltage (+2 to +5kV) is applied to the capillary through which a solution containing the analyte ions is flowing. The positive ions in the solution will experience an electric field towards the tip of the capillary while the negative ions will move away from the tip. The positive ions are expelled from the tip and form a Taylor cone[29]. The cone has 2 characteristics: a jet and a plume. The cone-jet is the beginning of the electrospray and the plume is formed due to Coulombic fissions. As the droplets travel towards the mass spectrometer, the solvent begins to evaporate hence concentrating the charge density at the surface of the drop. The droplet will finally reach the Rayleigh instability limit, the point at which the Coulombic repulsion between the positive charges is greater than the surface tension of the droplet and fission will occur. The droplets will

repeatedly undergo Coulombic fission until bare ions are produced. This represents the charge residue model, proposed by the Dole *et al.*[28] and the primary desolvation mechanism for large molecules like proteins[30]. Another model of desolvation is ion evaporation model, proposed by Iribane *et al.*, where bare ions and small solvated ion clusters are ejected from very small and highly charged droplets (radii <10nm)[31].

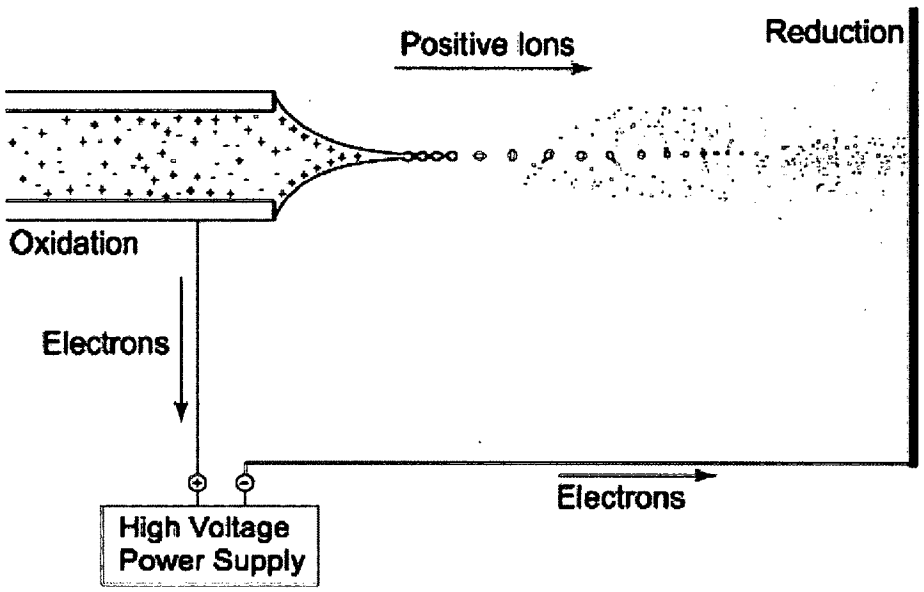


Figure 1-5: A schematic of electrospray ionization. Adapted from Kebarle, et al.[32]

A variation in electrospray is nanospray which has a flow rate through the capillary of ≈20nL/min[33]. This technique was introduced by Wilhm and Mann[33] and has been adapted as standard for biological mass spectrometry due to reduced sample consumption. Conventional electrospray capillary has an internal diameter of 100μm and flow rates of 1-10μL/min and thus results in droplets size in the micron range[33]. On the other hand, nanospray tips are made by pulling the spraying capillary to a fine tip with a 1-2μm orifice diameter[34]. Using these nanospray tips coupled to a nanoHPLC, a low flow

rate of ≈20nL/min can be achieved and which generates droplet size in the range of 200nm[34].

Nanospray ionization sources offer two options for introducing the sample into the mass spectrometer: direct infusion and on-line nanoHPLC. For direct infusion, the nanospray tips are coated with Au/Pd and for on-line nanoHPLC, the tips are uncoated fused silica; both of which are commercially available. With nanospray ionization, much lower voltage is applied, typically in the range of 0.7-1kV and has shown to have a greater tolerance for non-volatile salts within the samples[34,35].

For samples to be ionized by ESI and nanoESI, the analytes need to have a charge and for proteins, this is achieved through protonation for positive ESI/nanoESI and deprotonation for negative ESI/nanoESI. Proteins are zwitterionic compounds and the net overall charge on the proteins is dependent on the number of basic and acidic amino acid residues they contain (the isoelectric point, pI) and the pH of the solution. However, in mass spectrometry, the ions have to be in gas-phase hence the charge of the protein in gas-phase will be dependent on the proton affinities of the basic amino acids in the proteins. Proton affinity is the enthalpy of reaction of deprotonation at 298K and is thus temperature-dependent. The basic residues, arginine, lysine and histidine, have higher proton affinities than some of the common solvents and buffer components used in biological analysis, shown in Table 1.2. For positive ESI, the pH of the solution is normally lowered by the addition of an organic acid e.g. formic or acetic acid; for negative ESI, the pH is increased by addition of ammonia or an amine base.

Table 1-2: Proton affinities of basic residues and some common compounds. Reproduced from Smith[35].

Molecule	Proton Affinity (kJ/mol)/(kcal/mol)
Arginine	1051.0/251.2
Lysine	996.0/238.0
Histidine	988.0/236.0
Ammonia	853.6/204.0
Acetone	812.0/194.1
Acetonitrile	779.2/186.2
Methanol	754.3/180.3
Water	691.0/165.2

1.2.1.2 Mass Analyzers: Quadrupoles and Time-of-flight

1.2.1.2.1 Quadrupoles

A quadrupole mass analyzer consists of four parallel electrodes arranged in a radial array. Ideally, the electrodes should have a hyperbolic cross-section but due to economic constraints, their cross-section is circular where the radius of the rods is 1.148 times the quadrupole field radius. A direct current potential and an alternating radio frequency potential are applied to opposite pairs of rods, as seen in Figure 1.6.

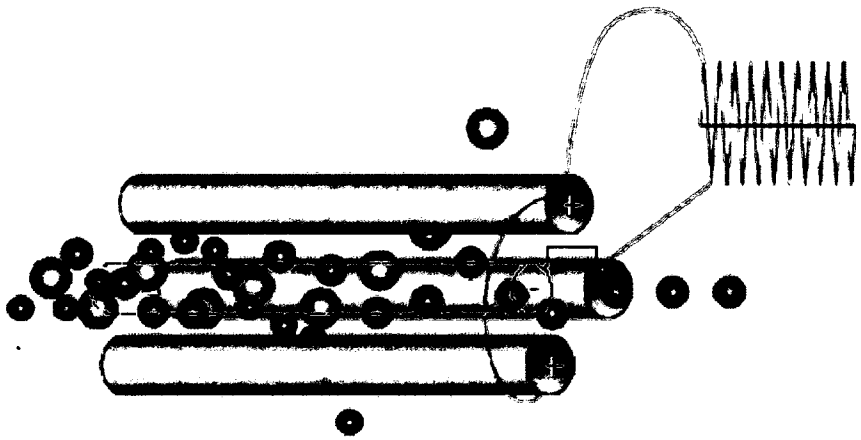


Figure 1-6: Quadrupole mass analyzer with a high frequency voltage generator. Adapted from Glish, et al.[36]

Traditionally, mass spectrometers resolve ions by dispersing ions in space or in time. However the quadrupole work similarly to a tunable variable bandpass filter where ions within a narrow mass region (<1Th) are allowed to pass through them[37]. By electronically sweeping the position of the bandpass region, quadrupole can be used to resolve ions based on m/z ratio.

When an ion beam enters the quadropole, the ions will experience an electric field in the x-z plane and the y-z plane due to the potentials applied to the rods. The potential on the rods in the x-z plane, E_x, and y-z plane, E_y, are represented by the following equations:

$$Ex = -[U + Vcos(\omega t)]$$

and

$$Ey = [U + Vcos(\omega t)]$$

where U is the magnitude of the DC potential, V is the magnitude of the alternating RF waveform and ω is the angular frequency of the RF waveform.

The simultaneous application of two different voltages to the rods results in selective transmission of ions through the quadrupole. The quadrupole will act as a high bandpass mass filter when a positive DC potential is applied on the x-z plane rods and a low bandpass mass filter when a negative DC potential is applied on the y-z plane rods. Under positive DC potential, the overall potential on the x-z plane is positive which focusses the heavy ions onto the center axis of the rods while the light ions will experience a large acceleration which causes them to collide with the rods and discharge. The light ions are pumped out of the instrument as neutral species and only the heavy ions are transmitted to the detector. Whereas, the potential on the y-z plane will be the same magnitude as the x-z plane but opposite sign and under a negative DC potential, the overall potential applied on the y-z plane will be negative. That means, the light ions will be focused onto the center axis of the rods and the heavy ions will eliminated from the ion beam. Collectively, a narrow bandpass mass filter is established; see Figure 1.7. The shaded triangle in Figure 1.7 indicates the range of ions transmitted through the quadrupole. The width of that triangle is determines the mass resolution; the narrower the triangle, the higher the resolution and vice versa. A narrow bandpass can be achieved by increasing the DC potential relative to the RF potential; however this compromises the sensitivity of the quadrupole (decrease in height of the triangle).

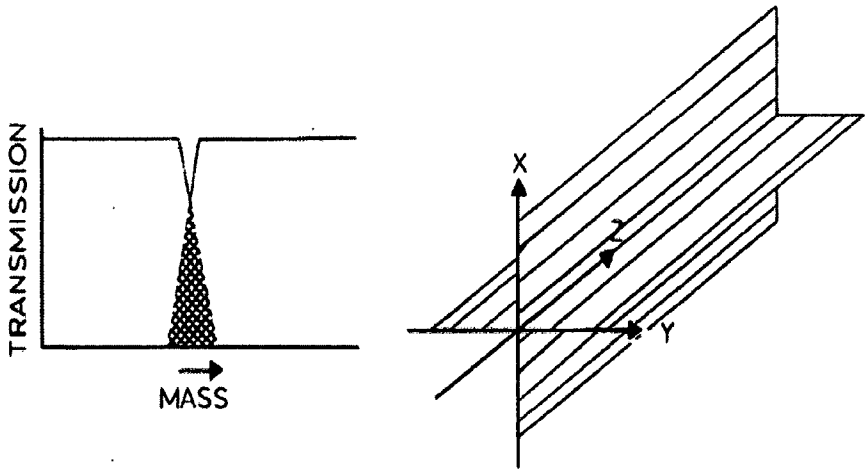


Figure 1-7: Overlap of low pass filtering in the y-z plane with high pass filtering in the x-z plane resulting in a narrow bandpass mass filtering by quadrupole. Adapted from Miller, et al.[37]

The trajectory of the ions through the quadrupole is described by the canonical form of the Mathieu's differential equation:

$$\frac{d^2u}{d^2\xi} + [a_u + 2q_u Cos2\xi]u = 0$$

where u is the transverse displacement from the center of the field in the x- and y-directions and ξ is a parameter derived from the angular frequency (□) and time (t) by the following relation: $\xi = \frac{2\omega}{t}$. The dimensionless parameters, a_u and q_u, are defined by

$$a_u = \frac{4eU}{\omega^2 {r_0}^2 m}$$

and

$$q_u = \frac{4eV}{\omega^2 {r_0}^2 m}$$

where m is the ion mass and r_0 is the field radius of the quadrupole.

The solutions of these parameters can be classified as bounded or unbounded. In case of bounded solution, the solution is finite and thus physically corresponds to a stable trajectory for the ion. In contrast, an unbounded solution corresponds to an unstable trajectory for the ion and filtering out of that ion from the ion beam. The stability diagram, illustrated in Figure 1.8, shows the stable (shaded region) and unstable solutions to the equations of the parameters, a_x and q_x where a_x is related to the DC potential and q_x is related to RF potential.

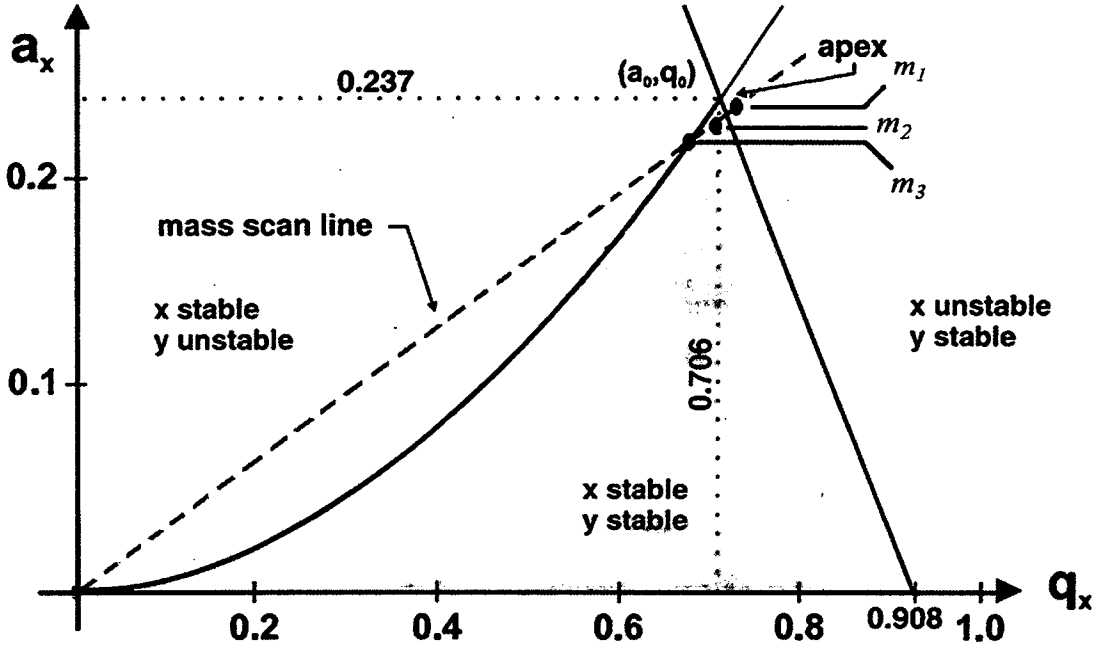


Figure 1-8: The stability diagram where the shaded region represents stable solution to the Mathieu's differential equation. Adapted from Blaum, et al.[38]

The intersection between the scan line and the stable region dictates the mass resolution of the quadrupole. Typically, the *a* and *q* values are chosen such that the scan line intersects with the tip of the stability region and hence a maximum mass resolution is

attained. Once those a and q values are determined, they are maintained at constant while the DC and RF potentials are simultaneously increased to allow ions of different mass to be transmitted through the quadrupole in a series of increasing m/z values. This would result in a plot of ion transmission/abundance against m/z values, called mass spectrum.

Another operational mode for the quadrupole is the RF-only mode, where no DC potential is applied to the rods. In this case, the parameter, a, will be zero and the slope of the scan line in the stability diagram will also be zero. Thus a large number of ions would have stable trajectories and be transmitted through the quadrupole. The RF-only mode quadrupole play an important role in the hybrid quadrupole time-of-flight instrument (*vide infra*).

1.2.1.2.2 Time-of-flight

In a time-of flight (TOF) mass analyzer, the ions are accelerated linearly down a flight tube and resolved based on their travel time through the flight tube. The TOF analyzer has 3 main regions: the ion accelerator, drift tube and ion detector. In the ion accelerator, the ion acquires a specific kinetic energy defined by the following equation:

$$KE = zeEs = \frac{mv^2}{2}$$

where z is the number of charge on the ion, e is the charge of the electron, E is the intensity of the electric field, s is the axial displacement of the ions within the ion accelerator, m is the mass of the ion and v is the velocity of the ion. The equation can be re-arranged to show the relation between velocity and the m/z of the ion. Therefore, the velocity of the ion can be defined by

$$v = \sqrt{\frac{2zeEs}{m}}$$

The equation illustrates that the velocity of the ion is inversely proportional to the square-root of the m/z of the ion. This means, ions with a lighter m/z will have a higher velocity and *vice versa* for heavy m/z ions. The magnitude of the velocity for each ion will determine the time taken by the ion to travel through the drift tube and be detected. The equation below relates the time taken by the ion to traverse the drift tube and velocity:

$$t = \frac{L}{v}$$

where L is the distance travelled by the ion in the drift tube and v is the velocity of the ion. Overall, the ions with a heavy m/z will have a lower velocity and thus take a longer time to travel through the drift tube and be detected. While the ions with a light m/z, will have a higher velocity and will travel through the drift tube in a shorter time. This is illustrated in Figure 1.9.

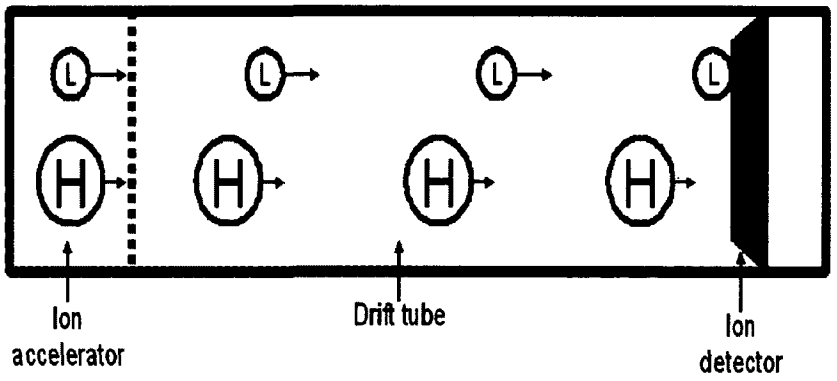


Figure 1-9: General schematic of a drift tube in time-of-flight mass spectrometry. Reproduced from Smith[35].

When the ion arrives on the detector, its arrival time is recorded and correlated to its m/z and thus a mass spectrum is created by recording the arrival time of all ions in the ion beam. The resolution of the TOF is dependent on the arrival times of ions. If ions of identical m/z arrive in a single pulse, the resolution is higher than if the ions arrived over a time span. The arrival times of ions with identical m/z are influenced by the following factors: the initial kinetic energy, the initial direction of the velocity and the spatial distribution of the ions. Prior to acceleration, the ions entering into the TOF will have a different kinetic energies and the magnitude of these energies add on to the acquired kinetic energy by the ions in the accelerator and thus broaden the window of the arrival time. The initial direction of the velocity of the ions will impact the resolution; if the initial direction of the velocity of the ion is opposite to the drift tube, the ions will have a delayed arrival time regardless of having the same kinetic energy magnitude. The spatial distribution of identical ions within the accelerator will dictate the amount of kinetic energy gained by the ions. If an ion is located near to the drift tube, it receives a shorter acceleration and hence travels slower through the drift tube. While an identical ion located further from the drift tube will experience a longer acceleration and hence have a faster travel through the drift tube. The effect of the initial kinetic energy, the initial direction of the velocity and the spatial distribution can be reduced by using a reflectron.

A reflectron is an ion mirror which causes an ion to change the direction of its velocity. As seen in Figure 1.10, the use of a reflectron/ion mirror, increase the distance travelled by the ions in the drift tube and corrects the differences in travel time of identical ions. In the case of varying initial kinetic energy between ions with identical m/z, the ions with the excess energy, U_R, will penetrate further into the reflectron while

those with appropriate energy, U_B, will penetrate a smaller distance into the reflectron. This results in the correction of the ions' drift through the first distance travelled through the drift tube and thus arriving at the detector in unison.

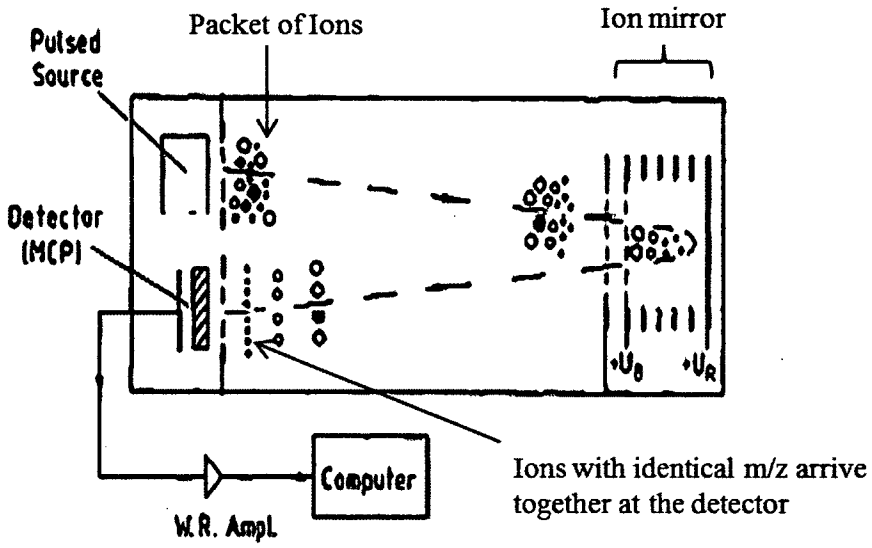


Figure 1-10: Schematic of a reflectron time-of-flight mass spectrometer. Adapted from Mamyrin[39].

In a reflectron TOF, the distance travelled by the ion double the length of the drift tube and hence the time taken for an ion to be detected is also doubled. This increases the difference between the arrival times of ions with identical m/z and thus improving the spectral resolution. In addition to reflectron, orthogonal acceleration has shown to increase the resolution[40] by overcoming the impedance due to the initial direction of the velocities for ions with identical m/z values. Orthogonal acceleration involves acceleration of the ions in the orthogonal direction, which is the direction of the TOF, as seen in Figure 11.1.

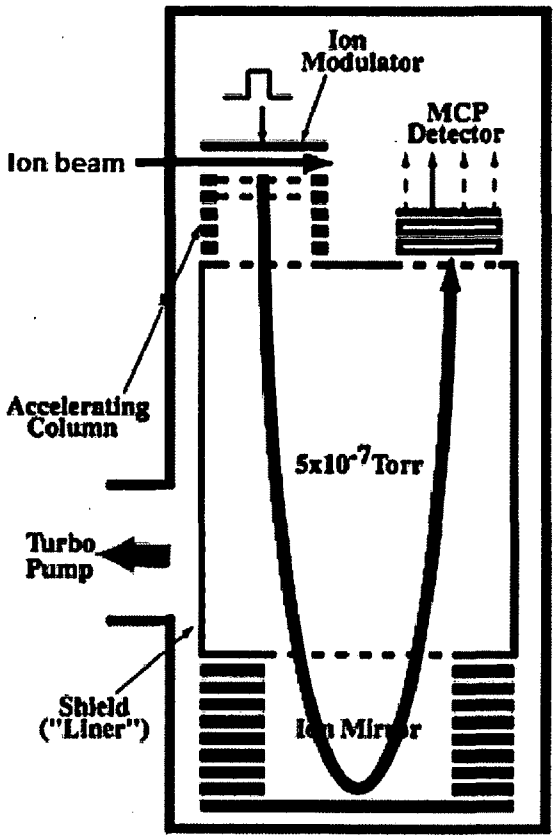


Figure 1-11: Schematic of orthogonal time-of-flight mass spectrometer. Adapted from Chernushevich, et al.[41]

In the accelerating column of the orthogonal TOF, the ion beam is directed into the field-free ion fill-up region, as shown in Figure 1.12. Once in the fill-up region, the ions are rapidly accelerated orthogonally to the relative velocity of the beam. The orthogonal velocity is induced onto the ions by simultaneously applying a positive pulse on the push-out electrode (POE) and a negative pulse on the G1 grid while the ions are accelerated. The consecutive grids, G2 and G3 are of lower potential than G1 and the potential on the G3 grid is identical to the TOF potential. The decreasing potentials of the grids direct the orthogonally-accelerated ions into the TOF. Since the potential on the last

grid, G3, is equal to that of TOF, the ions will experience no applied potential and will travel at their orthogonal velocity. When the ions are pushed into the TOF, the potential on the POE and G1 are reset to zero allowing a new packet of ions into the orthogonal accelerator.

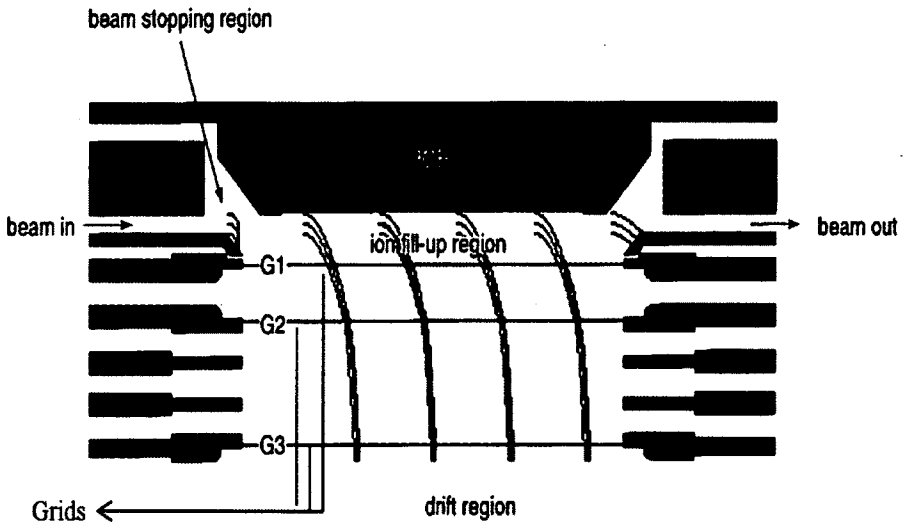


Figure 1-12: Schematic of the orthogonal acceleration region in the orthogonal time-of-flight mass spectrometer. Adapted from Mlynski, et al.[40]

The time taken to accelerate and push out a packet of ions from the orthogonal accelerator is dependent on the ions with the greater m/z in the TOF. These ions will take a longer time to travel through the TOF's drift tube and hence the release of the new packet of ion is timed such that its lighter ions will not surpass those heavier ions from the previous pulse to the detector.

1.2.1.3 Detector

All detectors work on a basic principle of neutralizing the resolved ion on to an electrode and registering the magnitude of the generated electric signal. The simple principle is demonstrated by a Faraday's collector. The limitation of this simple principle is the strength of the signal which is relative to the number of ions that impinge on the detector. The signal strength is generally low for a Faraday's collector hence an external preamplifier is required to process the signal such that it could be registered by an electronic controller (computer). However the detectors in modern instruments are reliant on amplification of the signal by generation of secondary electrons. Some examples of modern instrument detectors are electron multiplier, channeltrons and multichannel plates. These detectors are sequentially arranged electrodes made of emissive material, dynodes. When an ion impinges on the surface of the dynode, an electron is emitted. This electron, under a field, will impact with the dynode generating electrons as it propagates through the channel. Each new electron generated will create more electrons and eventually a cascade of electrons is formed from each ion. The cascade of electrons is pulled towards the anode plate at the end of the dynodes generating an easily detectable electric current.

In a multichannel plate (MCP) detector, the configuration of the dynodes is shown in Figure 1.13. The MCP detector is a dual-plate chevron arrangement (v-shape) of the multiple channels, dynodes. This chevron configuration prevents the positive ions to penetrate to the second MCP by maintaining a large directional change.

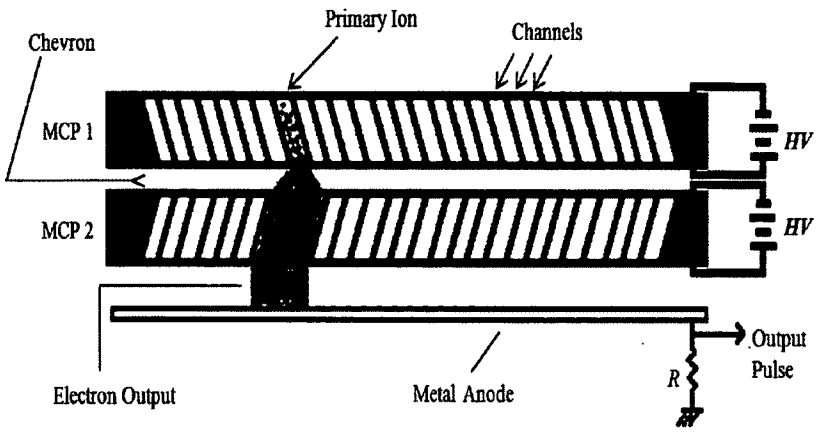


Figure 1-13: Side view of a multichannel plate (MCP) detector. Adapted from Wiza[42].

When the ion strikes the dynode, a cascade of electrons is created down MCP1 and into MCP2. The cascade of electrons will be detected by the anode and an output current is generated. The gain, G, at each dynode is given by

$$G = \left[\frac{AV}{2\alpha V_0^{1/2}}\right]^{\frac{4V_0\alpha^2}{V}}$$

where V_0 is the initial energy of the emitted secondary electron, V is the applied voltage, α is the length to diameter ratio of the microchannels and A is the proportionality constant between the secondary electron yield (δ) and electron collision energy (V_c) given by

$$\delta = AV_c^{1/2}$$

Typically, the dynodes in an MCP are 12μm in diameter and the gap between the MCP1 and MCP2 is 150μm[42]. The overall diameter of an MCP is 25mm with a bias angle of 8□ between the two plates[42]. This bias angle between the two MCP increases the gain due to the excitation of multiple channels in the MCP2 by the cascade of electrons from MCP1.

After the electric signal created by an ion and its secondary emission is registered, the MCP detector will require a 'dead' time to recover and register the events of the next ion. The 'dead' time is on the order of 1ns therefore, ions impacting the MCP detector less than 1ns apart will not be electrically distinguished by the time-to-digital converter (TDC). If the ions are co-impinging the detector, the anode will receive a large flow of electrons and be saturated. In order to overcome the above limitations, the MCP detector used in TOF instruments has four chevrons and each chevron has its own anode. This multi-anode MCP detector is known as a four-anode MCP detector. In this case, the ions arrive at the detector at the same time but their location will be different and therefore the cascade of electrons generated by the impact of the ions will be registered by different anodes. A TDC will record the arrival time upon an electric pulse from the anode and stores the arrival time in memory. The start signal for TDC to record the arrival time comes from the orthogonal accelerator in the TOF when it pushes out a packet of ions. The arrival times of all ions in a given acquisition period are stored in the TDC and used to calculate the m/z values of the ions and generate a mass spectrum; a plot of relative intensity at each arrive time and m/z values of each ion. The presence of four-anode MCP improves the counting efficiency of the TDC by a factor of 2.5[43].

1.2.1.4 Operation of a QqTOF mass spectrometer

In the chosen instrument for this work, there are two mass analyzers incorporated: a quadrupole and a time-of-flight. The abbreviated name, QqTOF, is adopted for both of the incorporated mass analyzing techniques in the instrument; where the Q refers to a mass-resolving quadrupole, q refers to a radio frequency (RF)-only quadrupole collision

cell and TOF refers to time-of-flight mass spectrometer. A schematic diagram of the hybrid QqTOF mass spectrometer is shown in Figure 1.14, with an electrospray ionization source. The instrument is considered to be hybrid since it consists of three quadrupoles, *q0*, *Q1* and *q2* and a TOF mass analyzer. Generally, a mass-resolving quadrupole is designated with the capitalized letter, *Q*, and the RF-only quadrupole is designated with non-capiltalized letter, *q*.

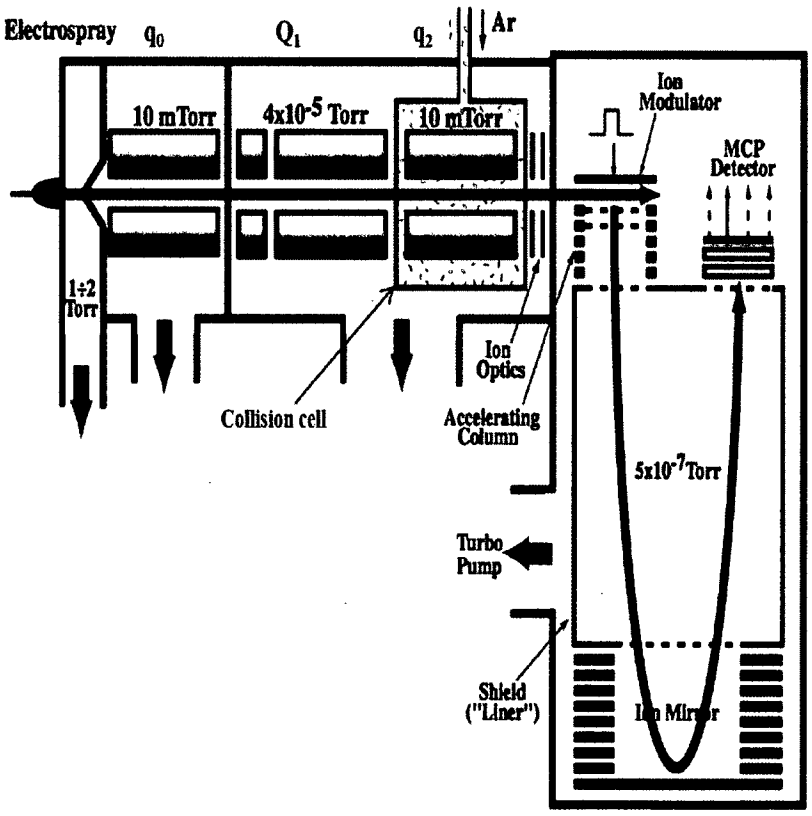


Figure 1-14: Schematic of hybrid quadrupole time-of-flight mass spectrometer. Adapted from Chernushevich, *et al.*[41]

The ions generated by the ESI source pass through the orifice on the front end of the mass spectrometer into sequentially lower regions of pressure. The sample is ionized

at atmospheric pressure, 760Torr, and moves into the mass spectrometer past the orifice, skimmer and into the q_0 quadrupole which is maintained at 10mTorr. The q_0 quadrupole is an RF-only quadrupole used to provide collisional damping and to focus the ion beam into the mass analyzers[44]. This collisional damping at the interface is used to decrease both radial and axial ion velocities of the ions and thus improve ion transmission into the Q1 quadrupole[44].

The Q1 quadrupole is operated at a pressure of $4x10^{-5}$Torr and this low pressure reduces the probability of a collision occurring between the ions and any neutral molecules within the instrument. The Q1 can be operated in the RF-only mode or mass filter mode. Under the RF-only mode, the Q1 will transmit all the ions in a given time through the quadrupole. The ions will pass through the q_2 and TOF and be detected. The m/z values of the ions in the given time are recoded in a single mass spectrum, MS scan. In contrast, when the Q1 is operated in a mass-filter mode, only a single chosen ion, termed the parent ion, will be transmitted into the q_2. Typically a mass window of 1-3Th is chosen for the parent ion in order to transmit the isotopic cluster[41]. The transmitted parent ion would undergo fragmentation in the collision cell, q_2, to generate daughter ions. The daughter ions will be resolved by the TOF followed by detection on the MCP detector. The m/z values of the daughter ions and their intensities are plotted as a mass spectrum, referred to as a product ion scan or MS/MS scan.

The q_2 quadrupole is special quadrupole due to its roles as a collision cell and in ion focusing. It is filled with an inert gas such as argon. The ions are accelerated to energies between 20 and 200eV before entering the q_2 where they undergo collision with the inert neutral gas. When the parent ions collide with the neutral gas molecules, they

will fragment and this is referred to as collision-induced dissociation (*vide infra*). The ion focusing properties of the q_2 will guide the scattered ion fragments back to the central axis of the quadrupole hence preventing the loss of fragment ions by scattering after collision[45]. For MS scan, the energy within the collision cell is kept below 10eV to prevent any dissociation of ions.

The application of same mass resolving principle twice or by combining two different mass resolving principles is referred to as tandem mass spectrometry. Tandem mass spectrometry has become the most powerful tool in peptide sequencing and protein identification.

1.3 Sample Preparation

As previously mentioned, proteins are extracted from the cells and digested prior to mass spectrometric analysis in the bottom-up approach. The proteins can be obtained from cultured cells or tissue samples or biological fluids. Firstly, the proteome needs to be efficiently extracted into a buffer to maintain the proteins in a soluble state in order to be subjected to enzymatic digestion. A combination of chemical and physical cell disruption techniques can be used to extract the proteins from the cells. The common chemical disruption techniques involve the use of detergents such as NP40 and CHAPS or chaotropic agents such as urea and thiourea. Medium to high concentrations of these chemical disruptors are required to solubilize the proteins and/or denature them. The choice of chemical disruption technique is dependent on the protease used to digest the proteins and the downstream analysis of the sample. For instance, using detergents would require a downstream step of removing the detergent since they are known to impede

enzymatic digestion and dominate the mass spectra[46]. Whereas chaotropic agents don't affect the spectral quality, they do impede the enzymatic activity hence require dilutions prior to digestion[47]. The proteins are also subjected to heat denaturation and reduction and alkylation of cysteine residue (break disulphide bridges) prior to digestion.

The most important step in sample preparation for proteomics is enzymatic digestion of proteins to peptides; the most common proteolytic enzyme used in proteomic studies is trypsin. Proteases can be either sequence-specific or less sequence-specific. For peptide sequencing purposes, the sequence-specific proteases are preferred over the less sequence-specific proteases. The less sequence-specific proteases do not cleave at a specific cleavage site on the protein and hence create a complex mixture of proteins which are difficult to interpret[48]. Trypsin is a sequence-specific protease that cleaves peptide bonds on the carboxyl-end of arginine and lysine residues in a protein unless those residues are followed by a proline residue.

Trypsin is a commonly used protease in mass spectrometry since it generates peptides with an ideal length (6-10 residues for yeast proteome)[49] for MS sequence analysis. In additional to ideal length, most tryptic-digested peptides have a basic residue at their C-terminus. This generates an additional site of protonation on the peptide besides its N-teminal amino group. Hence the typical charge on a tryptically-digested peptide would be 2+ during electrospray in positive ion mode. The ability to induce multiple charges on peptides is advantageous during collisional-induced dissociation of the peptides and detection of the fragments.

1.4 Peptide/Protein Identification

When a tryptically-digested peptide is introduced into the mass spectrometer by electrospray ionization, they are usually doubly charged and are designated as $(M+2H)^{2+}$ in which M is the mass of the peptide and H^+ is the mass the proton. Peptides can have a charge state greater than 2+ if they contain histidine groups which can also be protonated. The charge state of a peptide can easily be determined by analyzing its ion cluster peaks in the MS scan. Ion cluster peaks arise due to the 1% probability of each carbon atom being ^{13}C isotope rather than ^{12}C atom. These cluster peaks have a mass difference of 1Da however the scale on a MS scan is m/z values, thus the difference between the ion cluster is peak is represented by the ratio of mass difference between the clusters, 1Da, and the charge on the peptide, z. For example, in Figure 1.15 below, the mass difference between each of the ion cluster peak is approximately 0.3, therefore the charge on the peptide is calculated to be 3.

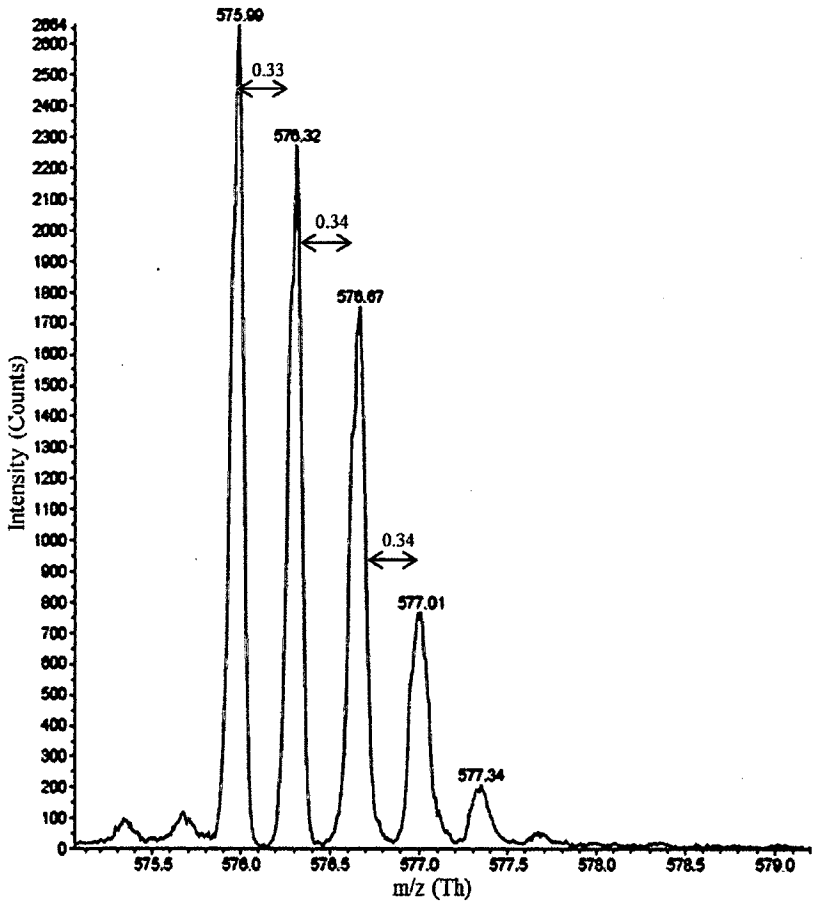


Figure 1-15: Mass spectra showing difference in m/z for ion cluster peaks for a peptide ion with m/z ratio of 575.99Th.

From the MS scan, the peak intensity and the m/z values of the peptides in a sample can be determined. Additional information about the peptides can be obtained from an MS/MS scan. When the Q1 is operating as a mass-filter, it would isolate a peptide ion from the mixture in the ion beam and transmit it to the collision cell, q_2, one mass at a time. In the q_2 cell, the peptide will fragment by colliding with argon molecules and the fragmented ions are resolved by the TOF. This fragmentation is referred to as collision-induced dissociation (CID). Under low energy (<100eV) CID[50], the peptide will fragment along the backbone of amide bonds to produce structurally informative

sequence ions and less useful small neutral species such as water, ammonia, etc. The transfer of kinetic energy of the peptide along its backbone upon collision with argon molecule will result in the cleavage of amide bonds. The resulting sequence ions are usually either b- or y-ions depending on the retention of the charge. That means, if the charge is retained by the amino-terminal of the dissociated sequence, the sequence ion is a b-ion and if the charge is retained by the carboxyl-terminal of the dissociated sequence, it is a y-ion. Figure 1.16 illustrates the fragmentation pattern of the peptide and the designation of the ions. The figure also demonstrates the numbering of the fragment ions: the b-ions are numbered sequentially from the N-terminus while the y-ions are number sequentially from the C-terminus.

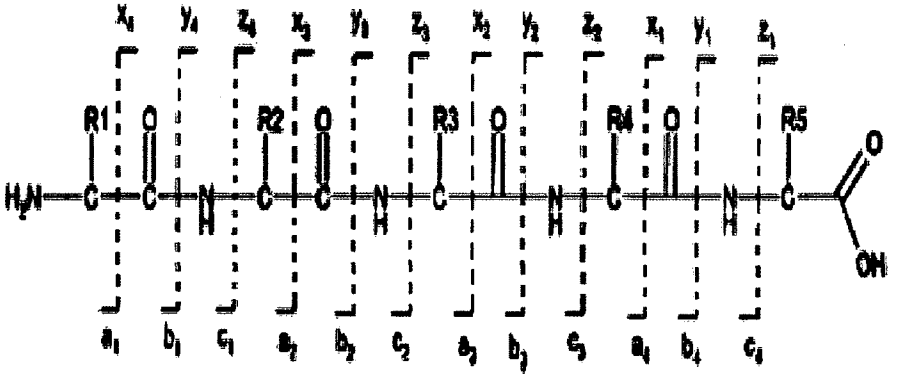


Figure 1-16: Fragmentation pattern of peptide ion along its backbone. Adapted from Paizs, et al.[51]

The x- and z-ions are formed when the charge is retained on the C-terminal fragment and the a- and c-ions are formed when the charge is retained by the N-terminal fragment. The subscripted number on the fragment ions refer to the number of amino acid residues within the daughter ion. The nomenclature for sequence ions was proposed by

Roepstorff and Fohlman[52] and later modified by Biemann[53]. Typically, the y-ions predominate the MS/MS scan for QqTOF instrument[48]. Doubly (or multiply) charged peptide will yield y- and b-ions. A simplified example of a CID spectrum, also known as MS/MS scan is shown in Figure 1.17. The sequence of amino acid residues within the peptide can be determined from the mass difference between successive fragment ions of the same type (e.g. y_n and y_{n-1}).

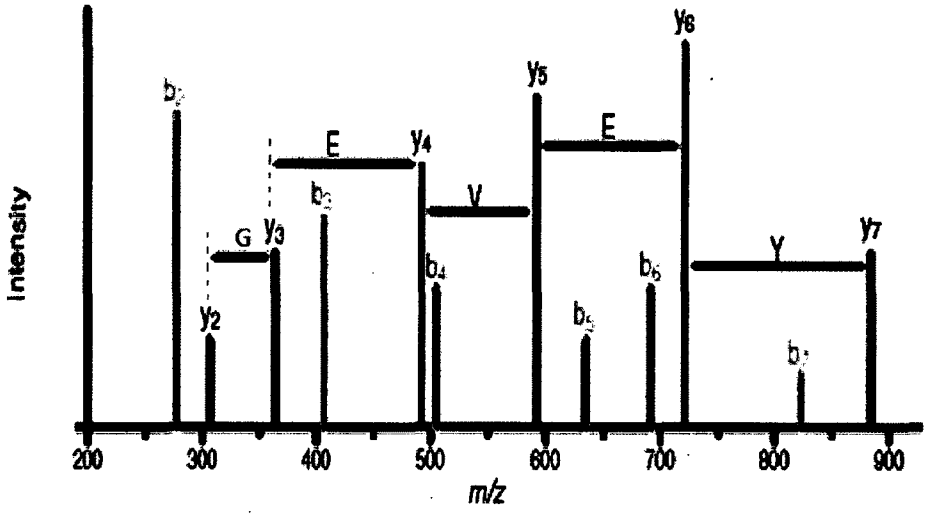


Figure 1-17: Schematic of collision-induced dissociation spectrum. Adapted from Sadygov, et al.[54]

The determination of the amino acid sequence within the peptide by consideration of mass difference between neighbouring peaks in the MS/MS scan is referred to *De Novo* sequencing. The mass difference is compared to the masses of the amino acids to determine the identity of the residue. This sequencing technique is time-consuming and requires extensive knowledge of behavior of gas-phase ions. There are a few drawbacks

associated to manual sequencing of peptides. Firstly, the information in the MS/MS scan is often not complete and there may be presence of peaks (noise/contaminants/peaks of ions that underwent neutral loss) which may not belong to the series of fragment ions. Secondly, it is difficult to assign a specific amino acid(s) to some mass differences. For instance, leucine and isoleucine residues have an identical mass and thus it is difficult to assign an exact identity. The presence of modifications on the amino acid residues could also complicate the sequence interpretation.

To overcome the inherent problem of *De Novo* sequencing, computational algorithms were developed to match the experiment MS/MS spectrum to theoretical MS/MS spectrum in a database. The database contains the theoretical MS/MS spectra of all proteins (known and theoretical) which are compiled from the genome sequence of a species. The theoretical MS/MS spectra are generated by *in silico* digestion of the proteins and fragmentation of the resultant peptides. Hence the identity of the peptide and the protein can be obtained by database searching. There are many algorithm based software packages, such as Mascot, Sequest, ProteinProspector and others, available for peptide/protein identification.

There are four models of matching spectral data to the database: descriptive, interpretative, stochastic and statistical and probability model. For the purposes of this work, only the statistical and probability model will be described further. The statistical and probability model is based on the determining the peptide identification score from the product of the probabilities of fragment matches of that peptide. Mascot is based on the MOWSE algorithm but it applies the probability-based scoring.

MOWSE algorithm was based on matching the fragment molecular weights with database molecular weight (DBMw) entry[55]. The error tolerance on the fragment molecular weight (FMw) and the enzyme used in digestion are user-selectable parameters. Based on the selected parameter, if the fragment molecular weight was within the criterion, DBMw-tolerance-1 < FMw < DBMw+tolerance+1, the fragment is assigned as a 'hit'[55]. The fragments are grouped according to their molecular weights and the protein from which they originate. The frequency within each group is normalized with the maximum observed frequency amongst all the groups to yield a frequency between 0 and 1[55]. The protein score is product of the normalized frequency values of the matched fragments. This product is then inverted and normalized to an average protein molecular weight of 50kDa[55,56]. Thus, the final score is given by

$$Score = \frac{50000}{H\ x\ P_N}$$

where H is the molecular weight of the 'hit' protein and P_N is the product of normalized frequency of the matched peptides.

The principle behind Mascot is the application of a probability-based model on the MOWSE algorithm. The probability, P, of the match between the experimental spectrum and the protein sequence in the database being a random event is calculated[57]. The match with the lowest probability is reported as the best match and *vice versa*. The protein score is reported by

$$S_{Mascot} = -10Log_{10}(P)$$

where S_{Mascot} is the Mascot score and P is the probability. Based on this negative logarithmic relationship between the score and probability, the best match is the one with the highest score.

Many modern mass spectrometers are capable of switching between MS and MS/MS scans by information-dependent acquisition (IDA). For QqTOF instrument used in this work, the IDA experiment is set up to perform a 1s MS scan, select four peaks with highest intensities and perform a 2-3s product ion scan on each of the selected peaks. The MS and MS/MS spectral data is saved and those m/z values selected for product ions are placed on an exclusion list for 90s. In this way, several MS and MS/MS scans can be performed on a sample and the spectral data can be saved under one experiment. The spectral data from each experiment can be searched against a database using Mascot to identify the peptides and proteins in the sample.

1.5 Separation Techniques

In the sample preparation stage, many components such as detergents, chaotropes and salts are introduced into the biological sample to extract the proteome. Besides these external components introduced into the biological sample, the biological sample itself may contain many biological contaminants such as nucleic acids, lipids and metabolites. Both the components and contaminants need to be removed from the analytical proteome sample. These compounds interfere with the analysis of the proteome by mass spectrometry. Many of these compounds have charges which interfere with the ionization of the peptides. They are also capable of suppressing the peptide ion signal and increase the background noise in mass spectrometry. To alleviate such interferences, the contaminants and external components need to be removed from the sample to enable analysis of the proteome within the sample. In addition to removal of the compounds, the complexity of the proteome also needs to be reduced to increase the resolution of the

spectral data and sensitivity of the ion signal. There are two main techniques utilized to remove the contaminants and reduce the complexity of the proteome: electrophoresis and liquid chromatography.

Electrophoresis involves separation of proteins on a polyacrylamide matrix under the influence of an applied electric field. In this technique, the proteins can be separated based on their molecular weight and/or their isoelectric point. Once the proteins are separated, the proteins are digested in the gel and extracted to be analyzed by bottom-up mass spectrometry. Electrophoresis will not be discussed further since it is beyond the scope of this work.

Liquid chromatography involves separation of proteins as well as peptides based on their physiochemical properties such as charge, size, hydrophobicity or ligand specificity. A sample can be subjected to multiple chromatographic separation techniques to reduce the complexity of the sample. In this work, four techniques have been used to separate the peptides: strong cation exchange chromatography, strong anion exchange chromatography, immobilized-metal affinity chromatography and reverse-phase chromatography. Each of these techniques will be briefly discussed below.

The common principle behind the above mentioned chromatographic techniques is the retention of peptides onto a stationary phase (loading) and elution of the peptides by altering the properties (such as pH, salt concentration, hydrophobicity) of the liquid mobile phase. A chromatographic technique can be coupled directly to an electrospray ionization source of a mass spectrometer. This coupling is referred to as online analysis (also, known as shot-gun proteomics) and it requires a solvent pumping system called high performance liquid chromatography (HPLC). Since the primary role

of an HPLC in this work is to pump a gradient of two solvents through an analytical column, the functioning of this instrument will not be discussed. The preferred chromatographic technique for online analysis is reverse-phase chromatography since electrospray is intolerant towards salts which lead to clustering of ions.

Reverse-phase liquid chromatography (RPLC) is separation technique based on the hydrophobicity of an analyte. In this case, an analytical column is packed with fused silica beads functionalized with C_{18} molecules. When an aqueous sample containing peptides is loaded onto the column, the hydrophobic amino acids within the peptides will enable the binding of the peptides to the C_{18} resins in the column. The separation of the peptides from the stationary phase can be achieved by passing a gradient of increasing concentration of organic solvent (such as acetonitrile) through the RPLC column. Peptides containing fewer hydrophobic residues will elute at higher concentrations of aqueous solvent while those containing abundant hydrophobic residues will elute at a higher concentrations of organic solvent.

Molecules such as salts are non-hydrophobic thus do not interact with the C_{18} resin and will be washed out of the column by an aqueous flow. Typically, an off-line (not coupled to MS analysis) RP column, often referred to as pre-column, is used to desalt (removal of non-volatile salts) a sample containing peptides. Once desalted, the loaded pre-column is then connected to an RP analytical column which allows separation of peptides at nanoflow rates. RP columns are also used in the sample preparative stage to remove urea from the lysis buffer.

A second dimension of separation can be applied to the peptide sample prior to RPLC. Peptides can be separated based on their overall charge by ion exchange

chromatography (IEC). The two IEC used in this work are strong cation exchange (SCX) and strong anion exchange (SAX) chromatographies.

In SCX, the resins have sulphonate functionalities ($-SO_3H$) bound to a solid support. These groups will reversibly bind to positively charged peptide ions. At low pH (~2), the side chains of the acidic amino acid residues within a peptide will be protonated and hence the peptide will have an overall positive charge (due to the protonated N-terminal and/or basic residues). Then, the peptides can be sequentially eluted by increasing the pH or salt concentration of the elution buffers. The sequential increase in the pH of the elution buffer will manipulate the overall charge of the peptide. At a given pH, the net charge on the peptide will be zero and thus that peptide will be eluted at that pH. Salt elution is reliant on competition between peptides and salts for the binding sites on the SCX resin. Sequentially increasing the ionic strength of the elution buffer will provide more salt ions to preferably bind to the resin and displace the peptides. SCX has been shown to be conducted in tandem with on-line RPLC as well as off-line. The on-line mode has enable automation of the elution process and requires low sample amount[58]. However, off-line mode allows fractionation of large sample amount and increased resolution[58].

In SAX, the resins have quantenary ammonium groups which give a positive charge to the resin. These positive groups will bind to negatively charged peptides. The pH of the sample containing peptides is made alkaline prior to loading on to an SAX column. The peptides are then eluted by sequentially decreasing the pH of the elution buffer.

Immobilized metal-ion affinity chromatography (IMAC) is a simple enrichment technique used in capturing phosphorylated peptides. The principle behind this affinity-based chromatography is the coordination between electron donor groups on the peptide and chelated metal ions on iminodiacetic acid (IDA) functionalized resin. The phosphate groups on the peptides are rich in electron donors, oxygen, which coordinate with IDA-Fe(III). The phosphopeptides can be displaced from the resin by using phosphate elution buffer. The phosphate groups in the elution buffer will compete with the phosphate groups on the peptides for the coordination site on the resin.

1.6 Quantification

The aspects of mass spectrometry-based proteomics discussed up to now are in relation to qualitative analysis i.e. identification of peptide and protein. However, in the past decade, research in proteomics has been focused on the ability to quantify the protein(s) of interest. Most quantitative proteomic studies are aimed at comparison between various conditions (usually 2/3 states) i.e. healthy and diseased or control and stress treatment. The ion intensity of a peptide's mass spectrometric signal may not be directly correlated to the amount of peptide present in the proteome[59]. The varying physiochemical properties of the peptides will lead to difference in mass spectrometric response[60]; for example, more abundant peptides will suppress the ionization and detection of low abundant peptides and thus influencing the ion intensity of the low abundant peptides. Hence a relative quantitative analysis of proteome, by comparative studies, is preferred over direct peptide quantification. There are two major ways of

quantitative analysis of proteome: by label –free methods or stable-isotope labeling methods.

Briefly, label-free methods involve the direct comparison of peak intensity of each peptide ion or comparison of the number of identified MS/MS spectra for each protein[61]. This quantitative method can be applied to compare the whole proteome of multiple conditions. The workflow of this technique is simple, low-cost and fast, however, the LC-MS/LC-MS/MS analyses need to be carefully controlled[60]. Due to the dynamic properties of the peptides, the run-to-run ionization conditions may differ and thus difficult to maintain run-to-run reproducibility. Also, the bias of information-dependent acquisition towards a more abundant peptide in case of co-eluting peptides will lead to masking of certain proteins[62].

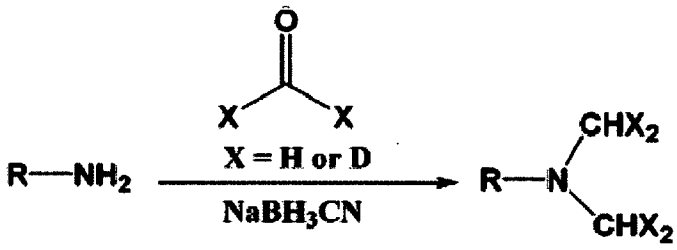
The stable-isotope approaches involve introduction of stable isotopes into the proteins or peptides. When a peptide is labeled with stable isotopes, it will differ from unlabeled peptide by mass of the stable isotopes incorporated. Typically the chromatographic properties (retention times) of the labeled peptides are same as their respective unlabeled peptide and hence they would co-elute during RPLC. Retention time shift is less observable between peptide containing label with ^{13}C, ^{15}N, ^{18}O isotopes[63]. However, peptides containing labels with ^{2}H (also referred as ^{2}D) have a small shift in retention time on RPLC compared to their unlabeled counterpart[64]. This shift will require an additional step in data analysis by integration of the chromatographic time scale. The comparison of the peak intensity between the light and heavy version of the same peptide in MS spectra or tandem MS spectra will achieve a relative quantitation of the peptide.

There are three main methods of introducing stable isotopes such as ^{13}C, ^{15}N, ^{18}O and ^{2}H (also referred as ^{2}D): Chemical, enzymatic and metabolic incorporation. Metabolic incorporation is an *in vivo* approach of incorporating stable isotopes into protein by growing cells in special media containing the stable isotopes. The popular metabolic incorporation technique is stable isotope labeling by amino acid in cell culture, SILAC. It involves the use of growth media containing isotopically labeled amino acids such as ^{2}H-leucine, ^{13}C-lysine, ^{13}C-tyrosine,^{13}C-arginine and/or $^{13}C/^{15}N$-arginine[65]. This *in vivo* strategy is advantageous due to the early stage of incorporation of isotopic labels on the protein which enables combination of labeled and unlabeled samples at cellular level and thus both the sample will be affected the same way during sample preparation and mass spectrometric analyses. This is a costly labeling strategy and is limited to cell cultures only. Tissue and plasma samples cannot be quantitated by this labeling strategy.

Enzymatic and chemical labeling strategies are performed *in vitro* on proteins and peptides. The enzymatic labeling strategy involves incorporation of ^{18}O during proteolytic cleavage of the proteins in the presence of "heavy" water ($H_2{}^{18}O$)[60]. One proteome sample can undergo proteolytic cleavage (digestion) in the presence of "light" water (H_2O) while the other proteome sample in presence of "heavy" water ($H_2{}^{18}O$). This generates a mass shift of 2Da between the labeled and unlabeled peptides. For multiply charged peptides, this mass shift is small and can interfere with the peptide's ion cluster peaks arising from naturally occurring isotopes. A minimum mass difference of 4Da is required between labeled and unlabeled peptides in order to reduce ^{13}C interference (ion cluster peaks)[59].

Chemical labeling approaches have been widely adapted as viable protein quantification strategies. They involve chemical derivatization of the side chains of

amino acid residues such as lysine and cysteine or the N-termini or C-termini of peptides. Many strategies have been developed to chemically label protein/peptides and the reagents for the strategy are readily available and cost-effective. In this work, a dimethyl labeling strategy is used for quantitation. This labeling strategy targets the N-termini and the epsilon-amino group of lysine residues using formaldehyde and subsequent reduction by cyanoborohydride[66]. The reaction involves addition of methyl (non-deuterated/deuterated) moieties on the N-terminus and/or ε-amino group of lysine residue, as shown in the equation below.



This chemical derivatization is applicable to global proteome sample since all tryptic digested proteins would generate peptides with N-terminus. Many tryptic digested peptides would have a C-terminal lysine residue with the exception of those with arginine and the C-terminus of the protein. Peptides from one sample can be labeled with 'light' (CH_3) moieties while another could be labeled with 'heavy' (CHD_2) moieties. The mass difference between the light and heavy labeled peptides would be 4Da for every modified site. This labeling strategy is fast, specific and cheap. However, subtle shifts in chromatographic elution times arising from the incorporation of deuterium complicates subsequent data analysis.

1.7 Hypoxia and Neuronal Cell Lines

Molecular oxygen (O_2) is essential for cellular respiration in living organisms. Under aerobic conditions, molecular oxygen acts as the final electron acceptor in the electron transport chain and helps to generate adenosine triphosphates (ATPs) from adenosine diphosphates (ADPs) during oxidative phosphorylation. ATPs are energy molecules utilized in cellular processes such as cell division, protein synthesis, signal transduction, cell motility and many others. These processes are vital to maintain cell viability. When the oxygen levels are normal in a cell, a high and constant ratio of ATP/ADP is maintained which aids in the functioning and survival of the cell. Earth's atmosphere is made up of approximately 21% O_2, however most mammalian tissues normally exist at 2% - 9% O_2[67]. When the O_2 level is <2%, the cells are said to be under hypoxia. Severely low O_2 levels (<0.02%) is referred to as anoxia. The term hypoxia is not to be confused with hypoxemia which implies reduced oxygenation of the blood. A reduction in normal O_2 level in the cell (hypoxia) will have consequences on the cellular processes and/or its viability. Hypoxia is encountered in a variety of pathological conditions, including stroke, tissue ischaemia, mountain sickness, and solid tumours[67, 68].

In response to hypoxia, cells alter their gene expression by activation of hypoxia-inducible factor (HIF) transcription factors. HIFs are heterodimeric proteins consisting of α and β subunits. There are three members in the cystolic HIF family: HIF-1α, HIF-2α and HIF-3α. The function of HIF-1α and HIF-2α is well characterized; however, little is known about HIF-3α[69]. HIF-2 is expressed in endothelial cells and in the lungs, kidney, heart and small intestine; whereas, HIF-1 is ubiquitously expressed. HIF-1 was the first transcription factor to be discovered in the family. Semenza and Wang originally

identified a protein complex, designated as HIF-1, bound to the hematopoietic growth hormone erythropoietin (EPO) gene under hypoxic stress[70].

Under normoxic conditions, the HIF-1α is ubiquitously transcribed, translated and subsequently degraded. The oxygen-dependent degradation of HIF-1α is primarily controlled by prolyl hydroxylase domain-containing proteins (PHDs) which use O_2 as a cosubstrate to hydroxylate two proline residues in the HIF-1α[71]. This modification on the HIF-1α leads to its ubiquitination (addition of polyubiquitin) by ubiquitin E3-containing ligase von Hippel-Lindau complex (pVHL) and thus targeted for degradation by proteasome[72]. This has been depicted in Figure 1.18. In addition, the asparagine residue on the C-terminal of the HIF-1α is hydroxylated by another oxygen-sensitive enzyme, Factor Inhibiting HIF-1 (FIH-1) to prevent interaction of HIF-1α with transcriptional coactivator CBP/p300 and thereby suppressing the transactivation activity of HIF-1α[73]. These post-translational modifications inactivate the unstable HIF-1α.

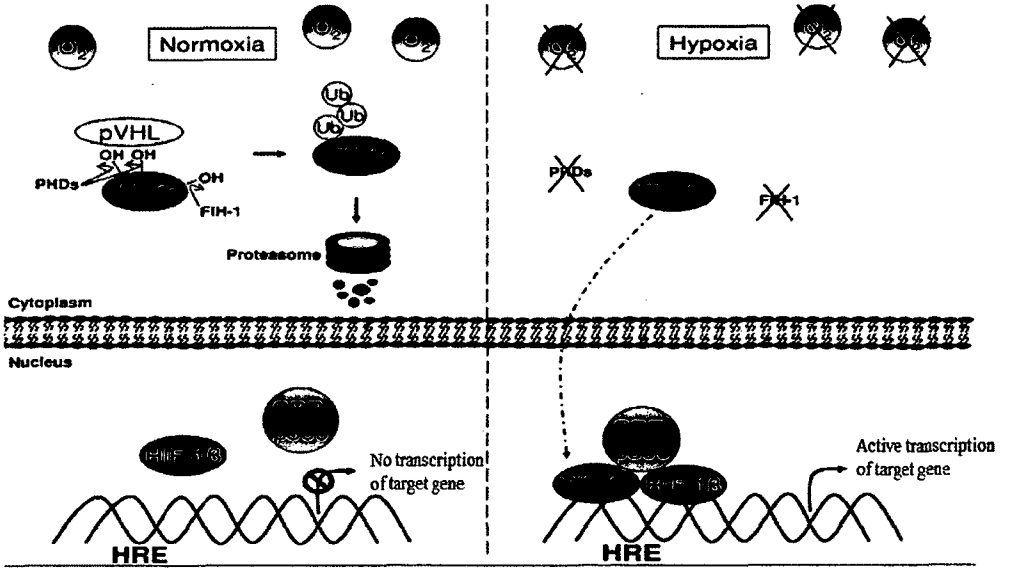


Figure 1-18: HIF-1α regulation by prolyl hydroxylases (PHDs) and factor inhibiting HIF-1 (FIH-1) under normoxia (normal O2 level) and hypoxia (deficient O2 level). Adapted from Gilany, *et al.*[77]

Under hypoxic conditions, the PHDs and the FIH-1 are inactivated and thus the lack of hydroxylation stabilizes the HIF-1α. The HIF-1α then translocates into the nucleus where it will dimerize with its β sub-unit, HIF-1β (also known as the aryl hydrocarbon nuclear translocator). The dimer will bind at the hypoxia response elements (HREs) in the promoter region of the target gene and recruit CBP/p300 to activate transcription. This is summarized in Figure 1.18[73].

The HIF-1 binds to the consensus sequence, 5'-RCGTG-3' (where R is A or G), in the HREs of many oxygen-regulated genes[74]. Besides EPO genes, there have been many genes which are directly or indirectly regulated by HIFs. The HIF-targeted genes include those encoding for vascular endothelial growth factor (VEGF), glycolytic enzymes such as phosphofructokinase L, and phosphoglycerate kinase 1, heat-stress proteins (HSPs) and many others. Each of these proteins influence metabolic adaptation, erythropoiesis, angiogenesis and vascular tone, cell growth and differentiation, survival and apoptosis, and thus are critical factors in mammalian development, physiology and several pathological conditions/disease[75,76].

Hypoxia has been observed in brain tissue after an episode of stroke (sudden occlusion of brain circulation). During a stroke, there is reduced delivery of oxygen and glucose to the brain which is the most metabolically active organ in the body. In human, if the O_2 supply to the brain is not restored within 5-8minutes, the tissue will suffer irreversible damage or death. Usually, patients are routinely supplemented with O_2 gas after an episode of stroke to reduce/reverse the hypoxic condition. However, a study by Ronning and Guldvog showed that oxygen administration worsened the prognosis for patients with moderate or minor strokes[77]. In the light of this adverse effect of oxygen

treatment, oxygen treatment cannot be recommended for all strokes[78]. However, the general consensus still remains that hypoxia should be treated with oxygen supplementation after an episode of stroke[79].

Tumours have a unique microenvironment characterized by heterogeneities in pH, nutrient and oxygen supply. These features are a result of poor vascular supply to the tumourogenic region. The activation of oncogenes and inactivation of tumor suppressor genes results in uncontrolled proliferation of cells (tumor formation). Significant regions of the tumour are at a greater distance from blood-vessels[80]. This causes a decreasing gradient of oxygen from the regions near the blood vessel to those further from the blood vessel. Hence, the interior regions of a solid tumour are under hypoxic stress. For tumours to grow beyond several mm^3, angiogenesis (formation of new blood vessels) must be induced in order to supply oxygen and nutrients to the solid tumour[81]. However, hypoxia will still prevail in tumours in part due to the high rate of cell proliferation (thus unable to meet O_2 demands) and the newly formed blood vessels are commonly structurally and functionally abnormal[81]. Hence, tumors have adapted to the hypoxic environment. The shift from oxidative phosphorylation to glycolysis for majority of the cellular energy within the tumour was described initially by Warburg in 1920s[82]. This shift reduces the dependence of the cell on mitochondrial respiration and therefore indicates the adaptation of cells to hypoxic conditions. Enhanced glucose metabolism and angiogenesis are classical processes occurring in tumours and these processes are controlled by genes that are normally induced in hypoxia. Analysis of human cancer biopsy samples and experimental animal models has shown the crucial role of HIFs in tumorgenesis[83-85].

Although the understanding of HIF-1α, at a molecular level, has increased remarkably over the past years, the understanding of the physiological implications of the various pathways involved, under hypoxic conditions, has lagged behind. Therefore, to understand the complicated network of pathways regulated during hypoxia, detection of many genes as well as proteins is important. The cDNA microarray technique is one of the most powerful tools to elucidate the mechanism of this network. However, a poor correlation between mRNA expression and protein abundance has been reported[86]. A large number of studies have been focused on single proteins or a subset of proteins associated with a pathway by antibody-based techniques. However, these techniques are limited to identification of known proteins for which specific antibodies exists as well as only a small number of proteins can be simultaneously identified. A few studies of global protein expression under hypoxic stress have been conducted by MS-based proteomics, more specifically, two-dimensional electrophoresis coupled with mass spectrometry. Therefore global proteomic investigation on the effects of hypoxia would provide an insight to the mechanisms of various cellular responses to hypoxia and thus of great interest. In this work, the effects of hypoxia on the proteomes of mammalian tumor cell lines, human neuroblastoma (SH-SY5Y) and rat pheochromocytoma (PC-12Adh), were investigated using shot-gun MS-based proteomics.

Neuroblastoma is a malignant (cancerous) tumour that forms in the nervous tissue of infants and young children. The cell line, SH-SY5Y, is thrice-cloned from SH-N-SK, a human-derived uncloned neuroblastoma line obtained from a four-year-old girl in 1970[87, 88]. Pheochromocytoma is rare tumour that usually forms in the adrenal glands. The PC-12Adh cell line is the adherent form of cell line derived from the pheochromocytoma in the medulla of rat adrenal glands. Upon treatment with nerve growth factor (NGF), PC-

12 cells have been observed to readily differentiate into neurons and hence can acts as a model system for studies on neuronal systems[89].

In this work, both of these cell lines were used to test the proteomic dynamics induced by hypoxia using a multidimensional separation, quantitative shot-gun MS-based proteomic approach. Furthermore, the effects of hypoxia on neuronally differentiated PC-12Adh cells were studied in the same manner.

1.8 References

1. Abbott, A. A post-genomic challenge: learning to read patterns of protein synthesis. Nature 402, 715-720 (1999).
2. Gstagier, M. & Aebersold, R. Applying mass spectrometry-based proteomics to genetics, genomics and network biology. Nature Rev. Genet. 10, 617-627 (2009).
3. Pandey, A. & Mann, M. Proteomics to study genes and genomes. Nature 405, 837-846 (2000).
4. O'Farrell. P. H. High resolution two-dimensional electrophoresis of proteins. J. Biol. Chem. 250, 4007-7021 (1975).
5. "The Nobel Prize in Chemistry 2002". Nobelprize.org. 24 Mar 2012 http://www.nobelprize.org/nobel_prizes/chemistry/laureates/2002/.
6. Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F. & Whitehouse, C. M. Electrospray ionization for mass spectrometry of large biomolecules. Science 246, 64-71 (1989).
7. Tanaka, K. et al. Protein and polymer analyses up to m/z 100 000 by laser ionization time-of-flight mass spectrometry. Rapid Commun. Mass Spectrom. 2, 151-153 (1988).
8. "The University of Queensland" di.uq.edu.au/sparqproteins. 14 Apr 2012 http://www.di.uq.edu.au/sparq/images/aastructures.jpg.
9. Chiti, F. & Dobson, C. Protein misfolding, functional amyloid and human disease. Annu. Rev. Biochem. 75, 333-336 (2006).
10. "Chemistry Images" yellowtang.org. 14 Apr 2012 http://www.yellowtang.org/chemistry.php.
11. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of human genome. Nature 431, 931-945 (2004).
12. Ayoubi, T. A. & Van De W.J. Regulation of gene expression by alternative promoters. FASEB J. 10, 453-460 (1996).
13. Jensen, O. N. Modification-specific proteomics: Charaacterization of post-translational modification by mass spectrometry. Curr. Opin. Chem. Biol. 8, 33-41 (2004).
14. UniProt. UniProt Knowledgebase: Controlled vocabulary of posttranslational modifications.http://www.uniprot.org/docs/ptmlist.
15. Mann, M. & Jensen, O. N. Proteomic analysis of post-translational modifications. Nature Biotechnol. 21, 255-261 (2003).
16. Gorkun, O. V., Veklich, Y. I., Weisel, J. W. & Lord, S. T. The conversion of fibrinogen to fibrin: Recombinant fibrinogen typifies plasma firbinogen. Blood 89, 4407-4414 (1997).

17. Cohen, P. The role of protein phosphorylation in human health and disease. Eur. J. Biochem. 268, 5001-5010 (2001).

18. Blom, N., Sicheritz-Pontén, T., Gupta, R., Gammeltoft, S. & Brunak, S. Prediction of post-translational glycosation and phosphorylation of proteins from the amino acid sequence. Proteomics 4, 1633-1649 (2004).

19. Steen, H., Jebanathirajah, J. A., Rush, J., Morrice, N. & Kirschner, M. W. Phoshorylation analysis by mass spectrometry: myths, facts, and the consequences for qualitative and quantitative measurements. Mol. Cell. Proteomics 5, 172-181 (2006).

20. Tarrant, M. K. & Cole, P. A. The chemical biology of Protein phosphorylation. Annu. Rev. Biochem. 78, 797-825 (2009).

21. Wu, S. et al. An integrated top-downa nd bottome-up strategy for broadly charaterizing protein isofroms and modifications. J. Proteome. Res 8, 1347-1357 (2009).

22. Whitelegge J. P. et al. Protein-sequence polymorphisms and post-translational modifications in proteins from human saliva using top-down fourier-transform ion cyclotron resonance mass spectrometry. Int. J. Mass Spectrom. 268, 190-197 (2007).

23. Wehr. T. Top-down versus Bottom-up approaches in Proteomics. LCGC 24, 497 (2006).

24. Roth, M. J., Parks, B. A., Ferguson, J. T., Boyne, M. T. & Kelleher, N. L. "Proteotyping": Population proteomics of human leukocytes using top down mass spectrometry. Anal. Chem. 80, 2857-2866 (2008).

25. Link, A. J. et al. Direct analysi of protein complexes using mass spectrometry. Nat. Biotechnol. 17, 676-782 (1999).

26. Gygi, S. P. et al. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nat. Biotechnol. 17, 994-999 (1999).

27. Cooks, R. G. & Rockwood, A. L. The "Thomson": suggested unit for mass spectroscopists. . Rapid Commum, Mass Spectrom. 5, 93 (1991).

28. Dole, M. et al. Molecular beams of macroions. J. Chem. Phys. 49, 2240-2249 (1968).

29. Taylor, G. Disintegration of water drops in an electric field. Proc. Roy. Soc. London. Ser. A 280, 383-397 (1964).

30. Winger, B. E., Light-Wahl, K. J., Ogorzalek Loo, R. R., Udseth, H. R. & Smith, R. D. Observation and implications of high mass-to-charge ratio ions from electrospray ionization mass spectrometry. J. Am. Soc. Mass Spectrom. 4, 536-545 (1993).

31. Iribane, J. V. & Thomson, B. A. On the evaporation of small ions from charged droplets. J. Chem. Phys. 64, 2287-2294 (1976).

32. Kebarle, P. & Verkerk, U. H. Electrospray: from ions in solution to ions in the gas phase, what we know now. Mass Spectrom. Rev. 28, 898-917 (2009).

33. Wilm, M. S. & Mann, M. Electrospray and Taylor-cone theory, Dole's beam of macromolecules at last? Int. J. Mass Spectrom. 136, 167-180 (1994).

34. Wilm, M. S. & Mann, M. Analytical properties of the nanoelectrospray ion source. Anal. Chem. 68, 1-8 (1996).

35. Smith, J. C. Mass spectrometry-based proteomics: Non-convalent interactions and protein identification. , 1-243 (2005).

36. Glish, G. L. & Vachet, R. W. The basics of mass spectrometry in the twenty-first century. Nat. Rev. Drug Discovery 2, 140-150 (2003).

37. Miller, P. E. & Denton, M. B. The quadrupole mass filter: Basic operating concepts. J. Chem. Educ. 63, 617-623 (1986).

38. Blaum, K. et al. Peak shape for a quadrupole mass spectrometer: comparison of computer simulation and experiment. Int. J. Mass Spectrom. 202, 81 <last_page> 89-89 (2000).

39. Mamyrin, B. A. Laser assisted reflectron time-of-flight mass spectrometry. Int. J. Mass Spectrom. Ion Processes 131, 1-19 (1994).

40. Mlynski, V. & Guilhaus, M. Orthogonal acceleration time-of-flight mass spectrometry. Mass Spectrom. Rev. 19, 65-107 (2000).

41. Chernushevich, I. G., Loboda, A. V. & Thomson, B. A. An introduction to quadrupole-time-of-flight mass spectrometry. J. Mass Spectrom. 36, 849-865 (2001).

42. Wiza, J. L. Microchannel plate detectors. Nucl. Instrum. Methods Phys. Res. , Sect. A 162, 587-601 (1979).

43. Barbacci, D. et al. Multi-anode detection in electrospray time-of-flight mass spectrometry. J. Am. Soc. Mass Spectrom. 9, 2-1333 (1998).

44. Krutchinsky, A. N., Chernushevich, I. V., Spicer, V. L., Ens, W. & Standing, K. G. Collisional damping interface for an electrospray ionization time-of-flight mass spectrometer. J. Am. Soc. Mass Spectrom. 9, 569-579 (1998).

45. Hoffman, E. Tandem mass spectrometry: a primer. J. Mass Spectrom. 31, 129-137 (1996).

46. Wisniewski, J. R., Zougman, A., Nagaraj, N. & Mann, M. Universal sample preparation method for proteome analysis. Nat. Methods 6, 359-363 (2009).

47. Lu, B., McClatchy, D. M., Kim, J. Y. & Yates, J. R. Strategies for shotgun identifiation of integral memebrane proteins by tandem mass spectrometry. Proteomics 8, 3947-3955 (2008).

48. Steen, H. & Mann, M. The ABC's (and XYZ's) of peptide sequencing. Nat. Rev. Mol. Cell Biol. 5, 699-711 (2004).

49. Swaney, D. L., Wenger, C. D. & Coon, J. J. Value of using multiple proteases for large-scale mass spectrometry-based proteomics. J. Proteome Res. 9, 1323-1329 (2010).

50. Papayannopoulos, I. A. The interpreatation of collision-induced dissociation tnadem mass spectra of peptitides. Mass Spectrom. Rev. 14, 49-73 (1995).

51. Paizs, B. & Suhai, S. Fragmentation pathways of protonated peptides. Mass Spectrom. Rev. 24, 508-548 (2005).

52. Roepstorff, P. & Fohlman, J. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. Biomed. Mass Spectrom. 11, 601 (1984).

53. Biemann, K. Contributions of mass spectrometry to peptide and protein structure. Biomed. Environ. Mass Spectrom. 16, 99-111 (1988).

54. Sadygov, R. G., Cociorva, D. & Yates, J. R. Large-scale datase searching using tandem mass spectra: Looking up the answer in the back of the book. Nature Methods 1, 195-202 (2004).

55. Pappin, D. J. C., Hojrup, P. & Bleasby, A. J. Rapid identification of proteins by peptide-mass fingerprinting. Curr. Biol. 3, 327-332 (1993).

56. Song, Z. et al. Development and assessment of scoring functions for protein identification using PMF data. Electrophoresis 28, 864-870 (2007).

57. Perkins, D. N., Pappin, D. J. C., Creasy, D. M. & Cottrell, J. S. Probability-based protein identification by searching databases using mass spectrometry data. Electrophoresis 20, 3551-3567 (1999).

58. Ye, M., Jiang, X., Feng, S., Tian, R. & Zou, H. Advances in chromatographic techniques and methods in shotgun proteome analysis. Trends Anal. Chem. 26, 80-84 (2007).

59. Lill, J. Proteomic tools for quantitation by mass spectrometry. Mass Spectrom. Rev. 22, 182-194 (2003).

60. Bantscheff, M., Schirle, M., Sweetman, G., Rick, J. & Kuster, B. Quantitative mass spectrometry in proteomics: a critical review. Anal. Bioanal. Chem. 389, 1017-1031 (2007).

61. Zhu, W., Smith, J. W. & Huang, C. Mass spectrometry-based label-free quantitative proteomics. J. Biomed. Biotechnol. 2010, 1-6 (2010).

62. Ghaemmaghami, S. et al. Global analysis of protein expression in yeast. Nature 425, 737-741 (2003).

63. Snijders, A. P. L., de Vos, M. G. J. & Wright, P. C. Novel approach for peptide quantitation and sequencing based on 15N and 13C metabolic labeling. J. Proteome Res. 4, 578-585 (2005).

64. Zhang, R., Sioma, C. S., Thompson, R. A., Xiong, L. & Regnier, F. E. Controlling deuterium isotope effects in comparative proteomics. Anal. Chem. 74, 3662-3669 (2002).

65. Yan, W. & Chen, S. S. Mass spectrometry-based quantitative proteomic profiling. Brief Funct. Genomic Proteomic 4, 1-12 (2005).

66. Hsu, J., Huang, S., Chow, N. & Chen, S. Stable-isotope dimethyl labeling for quantitative proteomics. Anal. Chem. 75, 6843-6852 (2003).

67. Bertout, J. A., Patel, S. A. & Simon, M. C. The impact of O2 availability on human cancer Nat. Rev. Cancer. 8, 967-975 (2008).

68. Tune, G. S. Psychological effects of hypoxia: Review of certain literature from the period 1950 to 1963. Percept. Mot. Skills 19, 551-562 (1964).

69. Li, Q. F., Wang, X. R., Yang, Y. W. & Lin, H. Hypoxia regulates hypoxia inducible factor (HIF)-3α expression in lung epithelial cells: characterization and comparison with HIF-1α. Cell Res. 16, 548-558 (2006).

70. Semenza, G. L. & Wang, G. L. A nuclear factor induced by hypoxia via de novo protein synthesis binds to the human erythropoietin gene enhancer at a site required for transcriptional activation. Mol 12, 5447-5454 (1992).

71. Bruick, R. K. & McKnight, S. L. A conserved family of prolyl-4-hydroxylase that modify HIF. Science 294, 1-1340 (2001).

72. Jaakkola, P. et al. Targeting of HIF-α to the von Hippel-Lindau ubiquitylation complex by O2 – regulated prolyl hydroxylation. Science 292, 468-472 (2001).

73. Gilany, K. & Vafakhah, M. Hypoxia: A review. JPS 1, 43-60 (2010).

74. Wenger, R. H. Mammalian oxygen sensing, signalling and gene regulation. J. Exp. Biol. 203, 1253-1263 (2000).

75. Semenza, G. L. et al. Hypoxia, HIF-1, adn the pathophysiology of common human diseases. Adv. Exp. Med. Biol. 475, 123-130 (2000).

76. Weidemann, A. & Johnson, R. S. Biology of HIF-1α. Cell Death Differ. 15, 621-627 (2008).

77. Ronning, O. M. & Guldvog, B. Should stroke victims routinely receive supplemental oxygen?: A quasi-randomized controlled trial. Stroke 30, 2033-2037 (1999).

78. Roffe, C. Hypoxia and stroke. Age Ageing 31, 10-12 (2002).

79. Adams, H. P. et al. Guidelines for the management of patients with acute ischemic stroke. Stroke 25, 1901-1914 (1994).

80. Dasu, A., Toma-Dasu, I. & Karlsson, M. Theoretical simulation of tumour oxygenation and results from acute and chronic hypoxia. Phys. Med. Biol. 48, 2829-2842 (2003).

81. Hanahan, D. & Folkman, J. Patterns and emerging mechanisms of the angiogenic switch during tumorigenesis. Cell 86, 353-364 (1996).

82. Warburg, O. On respiratory impairment in cancer cells. Science 124, 269-270 (1956).

83. Maxwell, P. H., Pugh, C. W. & Ratcliffe, P. J. Activation of the HIF pathway in cancer. Curr. Opin. Genet. Dev. 11, 293-299 (2001).

84. Semenza, G. L. Targeting HIF-1 for cancer therapy. Nature Rev. Cancer 3, 721-732 (2003).

85. Giaccia, A., Siim, B. G. & Johnson, R. S. HIF-1 as a target for drug development. Nature Rev. Drug Discov. 2, 803-811 (2003).

86. Yoo, G. H. et al. Docetaxel induced gene expression patterns in head and neck squamous cell carcinoma using cDNA microarray and PowerBlot. Clin Cancer Res. 8, 3910-3921 (2002).

87. Bielder, J. L., Roffler-Tarlov, S., Schachner, M. & Freedman, L. S. Multiple neurotransmitter synthesis by human nueroblastoma cell lines and clones. Cancer Res. 38, 3751-3757 (1978).

88. Biedler, J. L., Helson, L. & Spengler, B. A. Morphology and growth, tumorigenicity and cytogenetics of human neuroblastoma cells in continuous culture. Cancer Res. 33, 2643-2652 (1973).

89. Greene, L. A. & Tischler, A. S. Establishment of noradrenergic clonal line of rat adrenal pheochromocytoma cells which respond to nerve growth factor. Proc. Natl. Acad. Sci. 73, 2424-2428 (1976).

2 Chapter: Experimental Protocol

2.1 Materials and Equipment

2.1.1 Materials

Protein standards (α- casein from bovine milk, albumin from bovine serum, β-casein from bovine milk, concanavalin A from jack beans, haemoglobin from porcine, myoglobin from horse heart and ovalbumin from chicken egg), reagents (sodium cyanoborohydride, iodoacetamide, formalydehyde, urea phosphoric acid and formic acid) and Bicinchoninic Acid Assay Protein Determination Kit were purchased from Sigma-Aldrich (St. Louis, Missouri). The reagents, ammonium bicarbonate, potassium phosphate, potassium chloride, sodium acetate, sodium chloride and Tris-(hydroxymethyl)aminomethane were purchased from BioShop Canada Inc. (Burlington, Ontario). The HPLC grade solvents, acetonitrile, acetonitrile with 0.1% formic acid and water with 0.1% formic acid were purchased from Fischer Scientific. The D_2-formaldehyde and dithiothreitol were purchased from Cambridge Isotope Laboratories Inc. (Andover, Massachusetts) and Thermo Scientific (Waltham, Massachusetts), respectively. The strong cation exchange resin (bulk), polymeric strong anion exchange resin (bulk) and Bond Elut tC18, 500mg Sep-pak Cartridges, Magic C18AQ (bulk) and Phos-Select Affinity Gel were purchased from Canadian Life Sciences (Peterborough, Ontario), Varian Inc. (Mississauga, Ontario), Michrom Bioresource Inc. (Auburn, California) and Sigma-Aldrich (St. Louis, Missouri), respectively. Fused silica capillary column with inner diameter of 200μm and 700μm, fritted unions and PicoFrit™ emitters were purchased from Polymicro Technologies (Phoenix, Arizona), Idex Health & Science

(Oak Harbor, Washington) and New Objectives (Woburn, Massachusetts), respectively. The formamide and potassium silicate (Kasil 1 and Kasil 1264) were purchased from Promega Corporations (Madisson, Wisconsin) and National Silicates Ltd. (Etobicoke, Ontario; Kasil 1264 was a generous donation).

Cell lines (SHY-SY5Y and PC12Adh) and growth media (ATCC-formulated EMEM) were purchased from America Type Culture Collection (Manassas, Virginia). The F-12 nutrient mixture (HAM, 1X, containing L-glutamine), RPMI-1640 medium, heat-inactivated horse serum, newborn calf serum, fetal bovine serum 'gold', phosphate-buffered saline, pH 7.4 and nerve growth factor 7S were purchased from Life Technologies (Grand Island, New York) and PAA Laboratories GmbH (Pasching, Austria). The Complete Mini protease inhibitor tablets and PhosSTOP phosphatase inhibitor tablets were purchased from Roche Applied Science (Basel, Switzerland). The sequencing grade modified/proteomics grade trypsin (20μg vial) was purchased from either Sigma-Aldrich or Promega Corporations (Madisson, Wisconsin).

2.1.2 Equipment

- Sonicator - Vibra-Cell VCX130-Watt sonicator (Sonics, Newtown, Connecticut)
- HPLC – HP1090 Series HPLC system (Agilent Technologies, Santa Clara, California)
- Mass Spectrometer – QStar XL mass spectrometer (AB Sciex, Concord, Ontario)

2.2 Cell Culturing

Human, neuroblastoma (SH-SY5Y) cells were cultured as per the product information sheet[1] for ATCC® CRL-2266™. The frozen stock of SH-SY5Y cells was rapidly thawed in a water bath at 37°C, suspended in 9ml of complete culture medium (1:1 mixture of ATCC-formulated EMEM and F12 medium, with fetal bovine serum to a final concentration of 10%) and centrifuged at 125*g* for 5 minutes. The cell pellet was re-suspended in 9ml of complete culture medium in a 150mm cell culture plate and incubated at 37°C in an incubator with 5% CO_2. The medium was renewed every four days with fresh complete medium; the floating cells were recovered by centrifugation and combined with adherent cells. Once the cell confluence was approximately 50%, the cells were split and plated in fresh complete medium; the adherent cells were detached by treatment to fresh 0.25% trypsin, 0.53mM EDTA solution. Two different batches of the cells were used in the experiment: normoxia (21% O_2) and hypoxia (1% O_2). At a confluence of approximate 70-80%, one batch was treated to hypoxic stress, with 5% CO_2 at 37°C in a humidified incubator, for 24 hours. The control batch was under normoxic condition (21% O_2, 5% CO_2) at 37°C in a humidified incubator, for the same time period. After treatment, the cells were rapidly harvested and floating cells recovered by centrifugation and adherent cells detached by cell spatula into ice-cold phosphate-buffered saline (PBS) solution, split into four aliquots, frozen in liquid nitrogen and stored at -80°C until further processing.

Adherent rat pheochromocytoma (PC-12 Adh) cells were cultured as per the product information sheet[2] for ATCC® CRL-1721™(the catalog number of PC-12Adh is CRL-1721.1™; however, CRL-1721 is uncloned PC-12 and thus its culturing protocol

was applied to PC-12Adh). The frozen stock of PC-12 Adh cells was rapidly thawed in a 37°C water bath, suspended in to 9ml of complete growth medium (RPMI-1640 Medium with heat-inactivated horse serum to a final concentration of 10% and fetal bovine serum to a final concentration of 5%) and centrifuged at 180*g* for 15 minutes. The cell pellet was re-suspended in 9ml of fresh complete growth medium in a 150mm cell culture plate and incubated at 37°C in an incubator with 5% CO_2. The medium was renewed every two days with fresh complete growth medium. At a 50% cell confluence, the cells were split and plated in fresh complete medium; the adherent cells were detached by treatment to fresh 0.25% trypsin, 0.53mM EDTA solution and aspiration to split the clusters. The cells were sub-cultured until four batches of cells were obtained at approximate 80% confluence. The four batches of cells underwent different treatments at 37°C in a humidified incubator for a period of 24 hours in this experiment: normoxia undifferentiated, hypoxia undifferentiated, normoxia differentiated (differentiated with nerve growth factor, final concentration of 0.1%) and hypoxia differentiated. After the treatment, the cells from each batch were rapidly harvested, using a cell spatula, into ice-cold PBS solution, split into four aliquots, frozen in liquid nitrogen and stored at -80°C until further processing.

2.3 Cell Lysis and Tryptic-digestion

Each aliquot of cells were suspended in 250μL of ice-cold lysis buffer (8M urea, 75mM NaCl, one tablet of protease inhibitor cocktail per 10ml of buffer and one tablet of phosphatase inhibitor per 10ml) and lysed on ice by sonication for 3 x 60 seconds at 50%

power with 2 minute rest intervals in between[3]. The lysates were centrifuged at 2500*g* at 4^{0}C for 10 minutes to remove the cell debris[3].

The protein concentrations of the lysates were determined using Bicinchoninic Acid Assay Kit and a 1:10 dilution for each sample. The SH-SY5Y cell lysates had a protein concentration of 12.06mg/mL and 8.55mg/mL for normoxia-treatment and hypoxia-treatment, respectively. The PC-12Adh cell lysates had protein concentrations of 8.03mg/mL, 5.56mg/mL, 3.92mg/mL and 3.48mg/mL for undifferentiated normoxia, undifferentiated hypoxia, differentiated normoxia and differentiated hypoxia treatments, respectively.

Aliquots containing 2mg of total protein from normoxia- and hypoxia-treated SH-SY5Y cell lysates, 1mg of protein from normoxia- and hypoxia-treated PC-12Adh cell lysates and 0.8mg of protein from normoxia- and hypoxia-treated differentiated PC-12Adh cell lysates were subjected to digestion with trypsin separately. Initially, the proteins were reduced using dithiothreitol (DTT) at a final concentration of 5mM and incubated at 56^{0}C for 25 minutes. The samples were cooled to room temperature before alkylation of the thiol groups on the cysteine residues by addition of iodoacetamide to a final concentration of 14mM and incubation at room temperature and in the dark for 30 minutes. Any remaining iodoacetamide was quenched by addition of DTT to a final concentration of 5mM and incubation at room temperature and in darkness for 15 minutes. The reduced and alkylated samples were then diluted with a 1:5 (v/v; sample:Tris buffer) ratio using 25mM Tris buffer, pH8.2 and calcium chloride was added to a final concentration of 1mM. The sequencing grade modified/proteomics grade trypsin was re-constituted into 50mM ammonium bicarbonate solution and added to each

lysate aliquot at a 1:100 (w/w; protease:protein) ratio. The samples were gently mixed and incubated overnight at 37°C. The samples were cooled to room temperature and the digestion was halted using 1M phosphoric acid to a final concentration of 50mM. The digested samples were spun at 2500*g* for 10minutes at 4°C and any undigested contents were discarded.

2.4 Dimethyl Labeling

Each hypoxia-treated cell sample was paired with the appropriate control (normoxia-treated cells), where the control peptide sample was labeled with $-CH_3$ moieties on the amino group in the lysine residues and N-termini (light dimethyl labeling) and the peptides from hypoxia-treated sample were labeled with $-CHD_2$ moieties on the amino group in the lysine residues and N-termini (heavy dimethyl labeling).

Peptides from the normoxia-treated SH-SY5Y cells were passed through a conditioned and equilibrated Bond Elut 500mg tC_{18} cartridge and desalted with 9ml of 0.1% formic acid in water. 1mL of light dimethyl labeling solution containing 100mM sodium acetate pH5.5 buffer, 0.3% formaldehyde and 50mM sodium cyanoborohydride was passed through the cartridge followed by a desalting with 9ml of 0.1% formic acid in water. Peptides from the hypoxia-treated SH-SY5Y cells were then bound to the same cartridge and desalted. These peptides were then labeled with heavy dimethyl labeling solution which contains 100mM sodium acetate pH5.5 buffer, 0.3% D_2-formaldehyde and 50mM sodium cyanoborohydride and desalted with 9ml of 0.1% formic acid in water. The combined labeled peptides were washed with 900μL of 0.5% acetic acid to remove

formic acid, eluted with 5mL of 50%acetonitrile/0.5% acetic acid and evaporated to dryness.

The PC-12Adh cell samples were labeled in the same manner as the SH-SY5Y cells, with normoxia-treated PC-12Adh cell sample and differentiated normoxia-treated PC-12Adh cell sample acting as control for hypoxia-treated PC-12Adh cell sample and differentiated hypoxia-treated PC-12Adh cell sample, respectively.

2.5 Immobilized Metal Ion Affinity Chromatography (IMAC)

Phosphopeptides in the labeled samples were enriched using the Phos-Select Iron Affinity Gel[3]. The IMAC resin was washed three times with IMAC binding buffer, 40% acetonitrile, 25mM formic acid in water. The SH-SY5Y, PC-12Adh and differentiated PC-12Adh labeled peptide samples were re-suspended in 1200μL, 600μL and 480μL IMAC binding buffer, respectively and incubated with 50μL, 25μL and 25μL of IMAC resin, respectively, for 1 hour with vigorous shaking at room temperature.

The samples were spun at 15000*g* for 5 minutes and the supernatant was collected as the non-phosphorylated peptide samples. The IMAC resin was rinsed thrice with IMAC binding buffer; the rinse flow-through were collated with their respective non-phosphorylated peptide samples and frozen at -20^{0}C until further processing. The phosphopeptides were eluted by 3 x 5 minutes incubation of resin with 400μL IMAC elution buffer, 50mM K_2HPO_4/NH_4OH pH10 and collating the supernatant into 400μL of 10% formic acid. The phosphoenriched samples were stored at -20^{0}C until LC-MS/MS analysis.

2.6 Strong-Cation-Exchange (SCX)/Strong-Anion-Exchange (SAX) Fractionation

Fused-silica columns, inner diameter of 700μm, were packed with 12μm silica-based SCX or 300Å polymer-based SAX resin to a length of 26cm using a fritted union and a pressure vessel (see Figure 2.1).

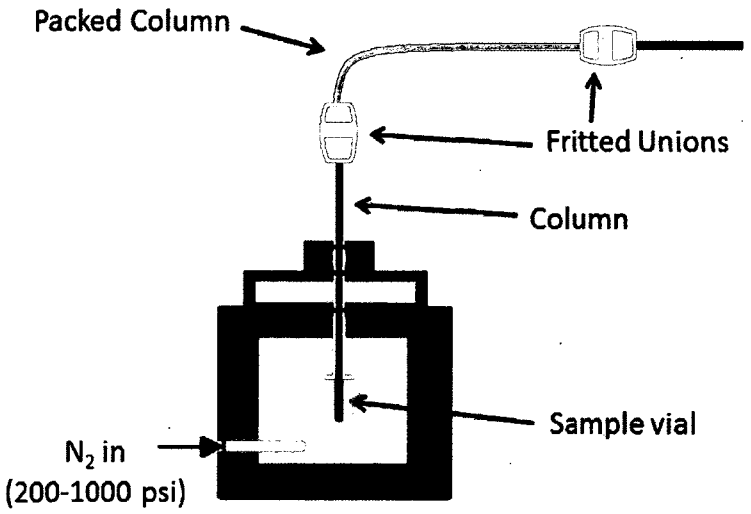


Figure 2-1: Typical setup of pressure vessel for flowing liquids (sample, washes and elution buffers) through the column using pressurized nitrogen gas.

The non-phosphorylated peptide samples were evaporated to dryness and re-suspended in 'ultrapure' water. For SCX fractionation, 1mg of SH-SY5Y non-phosphorylated peptide sample was acidified with H_3PO_4 to a final concentration of 50mM and loaded onto the column using the pressure vessel. The column was then washed with 500μL of 20% acetonitrile/8mM potassium phosphate buffer (KPB, pH 3) using the pressure vessel. The flow-through from the sample loading and washing steps were retained. The sample was then fractionated by sequentially eluting with 400μL of pH 2.5, 3.0, 3.5, 4.0, 4.5, 5.0, 5.5, 6.0, 7.0, 8.0 and finally 800mM ammonium

bicarbonate solution. SAX fractionation was performed in a similar fashion with the following exceptions: the 1mg sample was made basic by addition of ammonium hydroxide to a final concentration of 115mM prior to binding onto the SAX column, the wash buffer was 50mM NH_4OH/20% acetonitrile and the order of sequential pH elution was 8.0, 7.0 6.0, 5.5, 5.0, 4.5, 4.0, 3.5, 3.0 and 2.5.

The PC-12Adh samples were fractionated by SCX and SAX chromatography in similar manner described above with a few exceptions: 0.8mg of differentiated PC-12Adh sample was fractionated by SCX and SAX, for SAX fractionation, samples were made basic by addition of ammonium hydroxide to a final concentration of 130mM and due to damage in fritted union, the undifferentiated PC-12AdH could not be fractionated by SCX.

2.7 LC-MS/MS Analysis

For LC-MS/MS analysis, the reverse-phase chromatography pre-columns and analytical column were packed in-house with Magic C18AQ resin using the pressure vessel. Fused silica columns, with inner diameter of 200μm, were fritted in-house with a mixture of 88μL Kasil 1 and 16μL formamide or 30μL Kasil 1264 and 10μL formamide, and then packed with Magic C18AQ resin to a length of 5cm or 36cm. A 75μm x 4cm analytical column was packed in-house using PicoFrit™ Emitter and Magic C18AQ resin.

For SH-SY5Y samples, 100μL of each fraction (phosphorylated and non-phoshphorylated pH fractions) were loaded on to 200μm x 5cm pre-columns separately and analyzed online using the HPLC to generate a 90-minute gradient of 0.1% formic

acid/water and 0.1% formic acid/acetonitrile. The phosphorylated fractions of PC-12Adh samples were analyzed online in the same manner using a 90-minute gradient. The remaining samples, 400μL of PC-12Adh non-phosphorylated fractions and 300μL of non-phosphorylated fractions of SH-SY5Y samples were loaded onto 200μm x 36cm pre-columns separately and analyzed online using a 180-minute gradient of 0.1% formic acid/water and 0.1% formic acid/acetonitrile. Additionally, 100μL of flow-through from peptide loading and washing for SCX and SAX fractionation for SH-SY5Y and PC-12Adh samples were loaded onto 200μm x 5cm pre-columns and analyzed online using a 30-minute gradient of 0.1% formic acid/water and 0.1% formic acid/acetonitrile with the exception of flow-through from the wash step during SCX fractionation of SH-SY5Y sample. 400μL of that wash flow-through was loaded onto a 200μm x 36cm pre-column and separated online using 180-minute gradient of 0.1% formic acid/water and 0.1% formic acid/acetonitrile. The profiles of the 30-minute, 90-minute and 180-miunte gradient are shown in Figure 2.2.

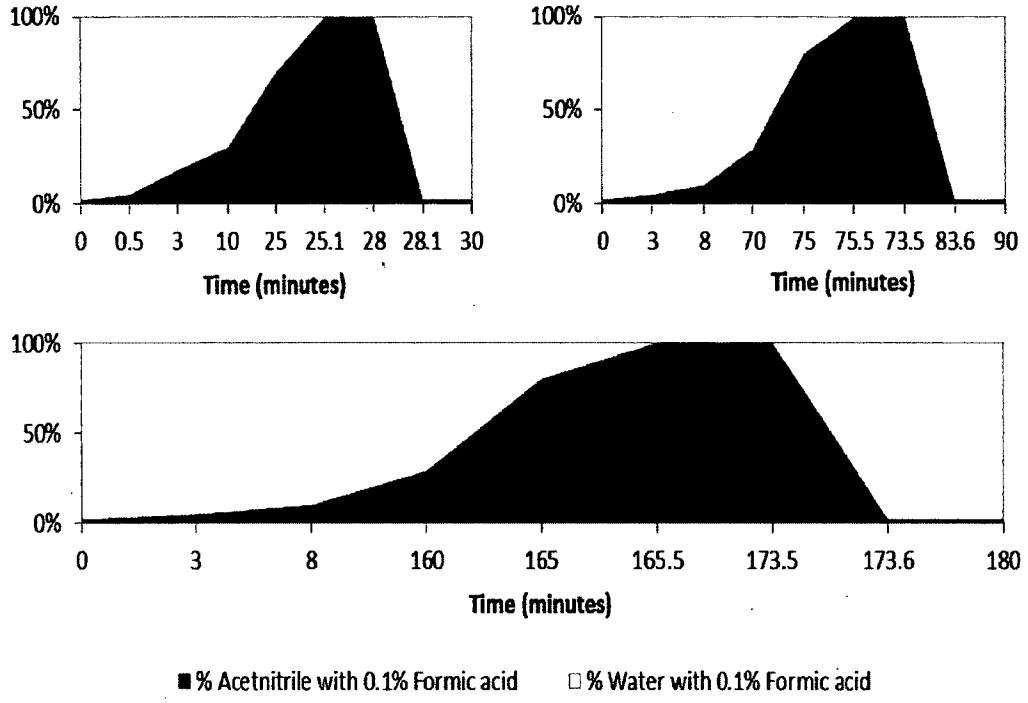


Figure 2-2: Linear gradient conditions of percent acetonitrile (with 0.1% formic acid) and HPLC grade water (with 0.1%Formic acid) used for 30 minutes (top left), 90 minutes (top right) and 180 minutes (bottom) LC-MS analysis.

The peptides, separated by reverse-phase chromatography, were electrosprayed at a voltage of +3,500V via the analytical column and the data was collected on the mass spectrometer using Analyst QS 2.0 software. The analyst software was used to set up an information-dependent acquisition experiment which would survey the MS scan every 1 second between m/z 400-1500Th followed by 3-second product ion scans of the 4 ions with the highest intensity and charge >+1. The experiment was set to place precursor ions in an exclusion list for 90s after being selected for product ion scan.

2.8 Peptide Identification and Quantification

The spectral data for each LC-MS/MS analysis were collated in to an Analyst data file format (.wiff) which was converted to mascot generic format (.mgf) using Mascot.dll Analyst plugin (v1.6b25). Using the search engine, Mascot 2.3[4], protein and peptide identifications were achieved by searching against the IPI Human and IPI Rat protein databases for SHY-SY5Y and PC-12Adh samples, respectively, with the following search parameters: trypsin as the proteolytic enzyme, up to 2 missed cleavages allowed, carbamidomethylation on cysteine residue as fixed modification, light and heavy dimethylation on lysine residue and peptide N-terminus, oxidation on methionine and phosphorylation on serine, threonine and tyrosine residues as variable modifications, mass error tolerance of 100ppm for mass spectra and 0.2Da for MS/MS spectra and only peptides with charge state of 2+, 3+ and 4+ were considered for MS/MS analysis. The resultant identification data was formatted to reject any peptide if its ions score was <15, its significance threshold was <1000, or its MS/MS spectra lacked a series of ≥ 3 consecutive y- or b-ions using an in-house filtering script.

An in-house computational script was developed to quantify the ratio of light to heavy dimethyl-labeled peptides. The following information about each identified peptide was obtained from the Mascot search: its modification i.e. light dimethyl-label or heavy-dimethyl label, the number of modification, its m/z and charge state and the elution time. The analyst data file (.wiff file), containing the spectral information about the peptides, was converted to an open-source format (mzML) using msconvert application of ProteoWizard. Using the quantification script, the maximum intensity of each m/z peak was extracted from the mzML file and the total intensity of the maximum peak (from the

Extracted Ion Chromatogram, XIC) was computed within a 0.2Da interval and 1 minute elution time for each m/z value in the Mascot file. A light-to-heavy intensity ratio was generated for all m/z peaks (light and heavy peaks) with an ion count of at least 20. The resultant data, peptide information and the ratios, was exported into a tab-delimited format.

The protein ratios were determined from the peptide ratios and normalized using the average ratio of total identified proteins in the sample. A 99% confidence interval of $0.68 < x < 1.47$, was observed with a 1:1 ratio of light- and heavy-labeled sample of 7 digested proteins standards (these standard proteins were digested in urea lysis buffer and labeled in the same manner as the biological samples). The identified proteins in the biological samples were considered to be up-regulated or down-regulated if their respective normalized ratio lied outside the 99% confidence interval.

2.9 Bioinformatics

Gene symbols for the identified proteins in the SH-SY5Y and PC-12Adh samples were extracted from their respective IPI database. The gene symbols and the ratios for the identified proteins were collated into a tab-delimited format. The files were sent to Dr. Dawn Jin at Health Canada to be processed for biochemical associations. The .tab file for each dataset was uploaded and analyzed by IPA® software (Ingenuity Systems, California). The datasets were analyzed for functional annotation and association of identified proteins in canonical pathways.

2.10 References

1. "CRL-2266" atcc.org 15 Apr 2012 www.atcc.org/attachments/17406.pdf.
2. "CRL-1721" atcc.org 15 Apr 2012 www.atcc.org/attachments/17593.pdf.
3. Villen, J. & Gygi, S. P. The SCX/IMAC enrichment approach for global phosphorylation analysis by mass spectrometry. *Nat. Protoc.* **3**, 1630-1638 (2008).
4. Perkins, D. N., Pappin, D. J. C., Creasy, D. M. & Cottrell, J. S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**, 3551-3567 (1999).

3 Chapter: Results and Discussion

3.1 Protocol Overview and Optimization

In this work, the effects of 24-hour hypoxia treatment on the proteomes of 3 biological samples, undifferentiated and differentiated rat pheochromocytoma (PC-12Adh) and human neuroblastoma (SH-SY5Y) were quantitatively analyzed by a mass spectrometry-based shot-gun proteomic approach. The controls for each of the biological samples were maintained under normal oxygen level (21%) during the treatment time. The workflow of the mass spectrometry-based proteomic analysis has been summarized in Figure 3.1. The cells were lysed in urea-containing lysis buffer by sonication and the total protein content was determined by BCA assay. Equal amounts of protein from the control and treated biological samples were separately digested in-solution with trypsin. The digested samples were sequentially desalted and labeled with light (control) or heavy (hypoxia-treated) dimethyl-labeling reagents. The combined labeled sample was then subjected to IMAC for phosphoenrichment. The phosphorylated fraction was directly analyzed by RPLC-ESI-MS/MS on QqTOF mass spectrometer. The non-phosphorylated fraction was split in to 2 for further fractionation: one was fractionated into 9 fractions by SCX chromatography while the other was fractionated into 8 fractions by SAX chromatography. Each SCX and SAX fraction was analyzed by RPLC-ESI-MS/MS on QqTOF mass spectrometer. The spectral data was used to identify (using MS/MS spectra) and quantify (using MS spectra) the peptides within the fractions.

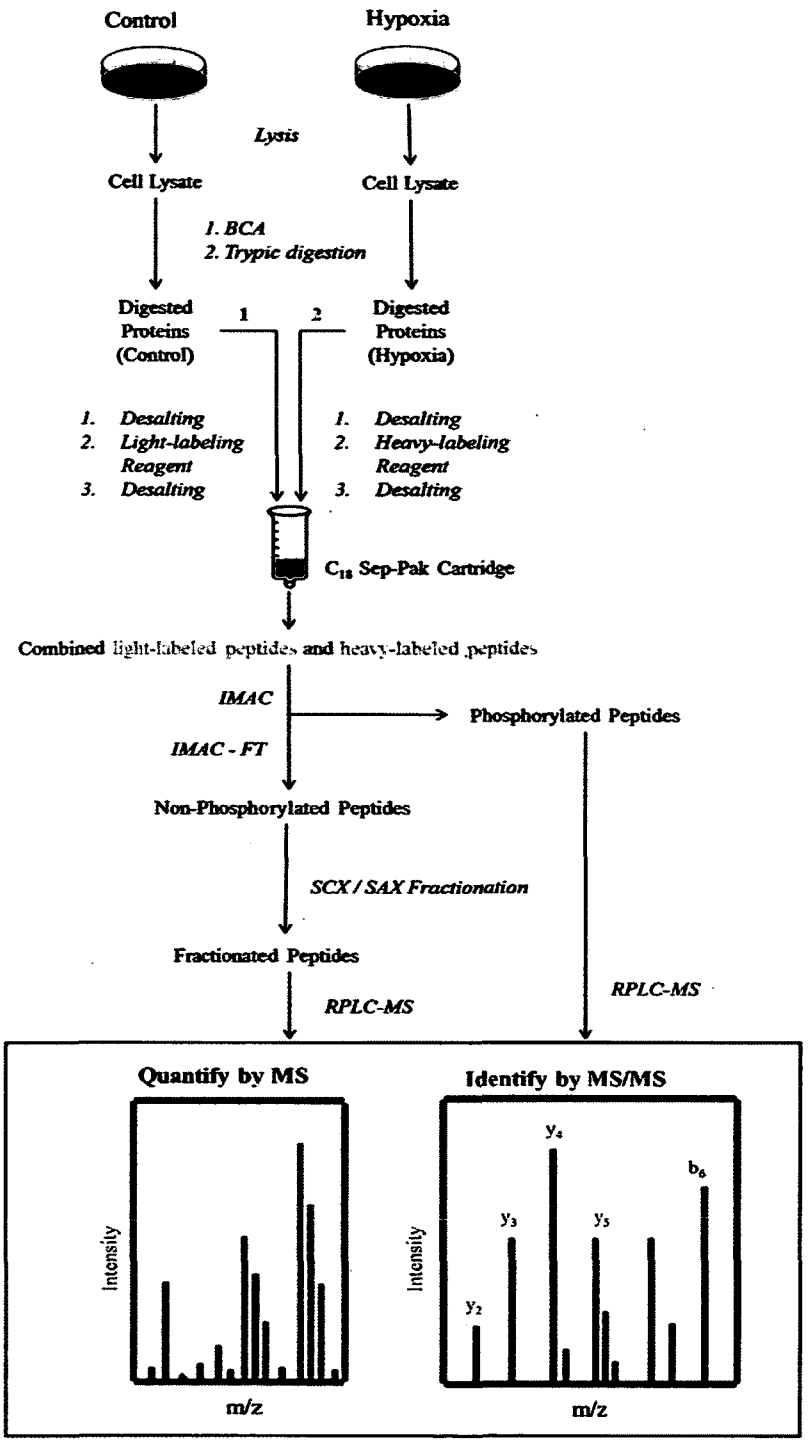


Figure 3-1: Schematic of proteomic approach to the analysis of SH-SY5Y, Undifferentiated PC-12Adh and Differentiated PC-12Adh samples.

Approximately 0.8-1mg of combined peptide sample was fractionated either by SCX or SAX chromatography. Initially, 100μL of each SCX and SAX fraction for the SH-SY5Y sample were analyzed using a 90-minute gradient of 0.1% formic acid/water and 0.1% formic acid/acetonitrile. A total of 2652 peptides were identified using the spectral matching software Mascot from all the SCX and SAX fractions combined into one results file and searched simultaneously. The 200μm x 5cm pre-column has a binding capacity of only 15μg. Due to the inability to load a known amount of fractionated peptides(without incurring sample loss) onto the column, the use of longer (200μm x 36cm) pre-columns, with a binding capacity of 100μg, was proposed. This prompted the need for longer RPLC gradient in order to maintain the chromatographic resolution and prevent bias towards abundant ions and/or ions that protonate preferentially. McCormack, *et al.* illustrated that increasing the length of gradient when analyzing complex mixtures, improves the molecular separation and thus increases the number of MS/MS acquired during the analysis[1]. Therefore, a 180-minute gradient was designed and implemented. A 15% increase in total peptide identification was observed, shown in Figure 3.2.

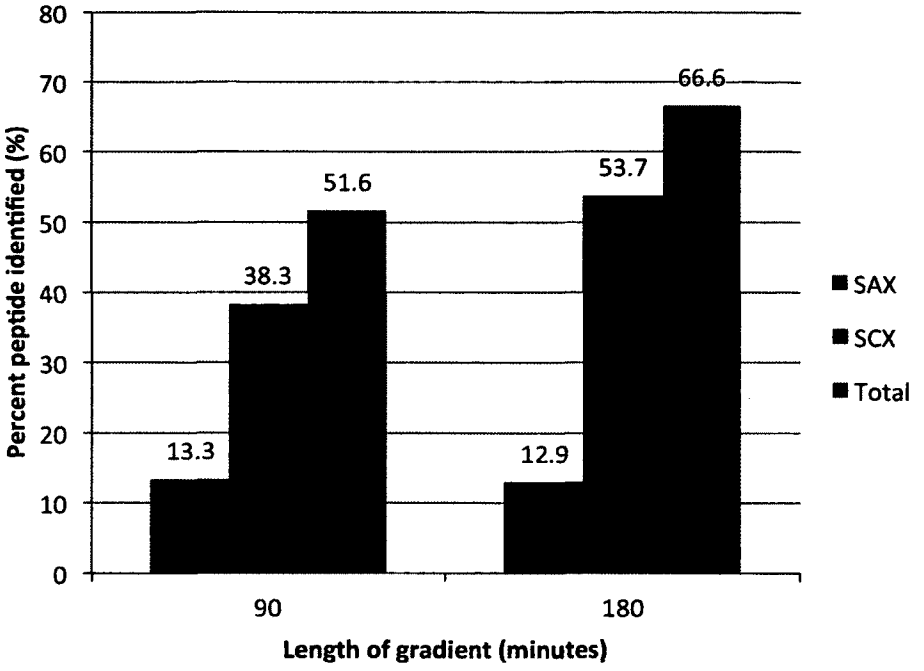


Figure 3-2: Percent of identified peptides by Mascot after RPLC-MS/MS analyses using 90 and 180 minute gradients.

Analysis of the flow-through (FT) from the peptide loading step for SY-SY5Y SAX fractionation showed the presence of peptides (data not shown) which indicates that a portion of the peptides do not bind to the SAX resin. In order for peptides to interact with the quaternary ammonium ions on the SAX resin, the peptides should have a net negative charge, which is conferred by the acidic residues under strong basic environment. Addition of ammonium hydroxide to the peptide samples prior to loading onto the column will increase the pH of the sample. It is critical to ensure that the pH is highly basic, more specifically for tryptically-digested peptides. Trypsin digestion of proteins renders all the resultant peptides with at least one basic residue (lysine or arginine at the C-terminus) with the exception of the peptides resulting from the cleavage

of protein C-terminus. As shown in Figure 3.3, an increase in the final concentration of ammonium hydroxide yields an increase in the total peptides identified by Mascot.

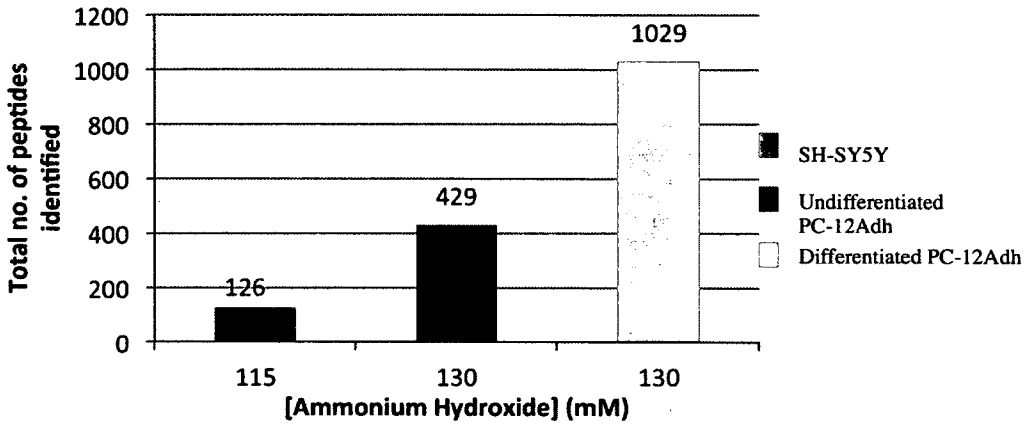


Figure 3-3: Effect of ammonium hydroxide concentration during SAX peptide loading step on the total number of peptides identified by LC-MS/MS and Mascot.

3.2 SCX versus SAX Fractionation

The SCX and SAX datasets for SH-SY5Y and differentiated PC-12Adh samples were evaluated to determine the peptide redundancy between the two separation techniques. Each dataset was individually assessed to remove duplicate peptide sequences (within and between fractions). Then, the SCX dataset was compared to the SAX dataset to evaluate the number of peptides identified by both fractionation techniques. Figure 3.4 shows the overlap between the SCX and SAX separation techniques for SH-SY5Y (top) and differentiated PC-12Adh (bottom) samples. For both the samples, the overlap between the SCX and SAX is less than 2%, suggesting that the two techniques are complementary and analyze different sections of the proteome.

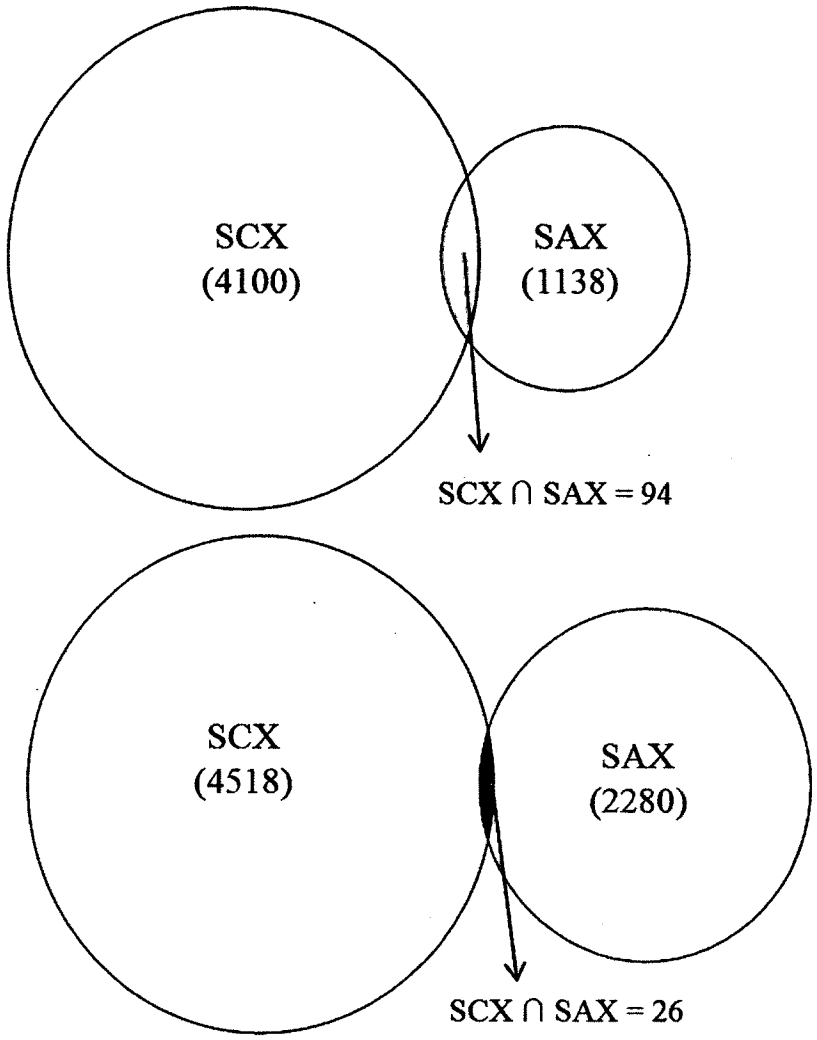


Figure 3-4: Peptide redundancy between SCX and SAX fractions of SH-SY5Y (top) and differentiated PC-12Adh (bottom) samples.

In SCX chromatography, peptides are fractionated according to the number of basic moieties (free N-termini, Arg, Lys and His residues) while SAX chromatography fractionates peptides according to the number of acidic moieties (free C-termini, Asp and Glu residues and acidic PTMs such as phosphorylation and sulphonation). Digestion of peptides with trypsin ensures that majority of the peptides have at least two basic moieties: N-terminus and Arg/Lys residue on the C-termini. At a low pH (<2), most

tryptic-digested peptides have at least a 2+ charge (since the carboxyl group have been protonated) and thus can interact with the sulphonyl groups on the SCX resin. The basic nature of majority of tryptic-digested peptides confers affinity towards the SCX resin and thus the number of peptides identified via SCX fractionation for SH-SY5Y and PC-12Adh samples were higher than SAX fractions.

Trypsin digestion does not restrict the number of acidic residues within the peptides. Therefore, at high pH (>11), the peptides will be negatively charged and the integer value of the charge is dependent on the number of acidic residues and acidic PTMs in/on the peptides. However, Arg residues have a positive charge at pH 11[2], and thus impede the induction of a net negative charge on peptides containing a single carboxyl-site (e.g. peptides with no acidic residues and PTMS). Such peptides would not be able to interact with the SAX resin and thus will be lost to the flow-through. This phenomenon, however, cannot solely account for the lower number of identified peptides by SAX fractionation. Another possible explanation to account for the decrease is the potential difference between the binding capacities of the SAX and SCX resin. Identical amounts of peptides were subjected to SCX and SAX fractionation, however, if the binding capacity of the SCX and SAX resin varied, the amount of peptides fractionated and identified would differ between the separation techniques.

3.3 IMAC

The efficiency of the IMAC was evaluated by comparing the number of phoshphopeptides identified to the number of non-phosphorylated peptides identified by LC-MS/MS and Mascot analysis of the phosphorylated peptide fraction. At an ions score

cut-off of 15, the numbers of phosphopeptides identified were 10, 5 and 2 for SH-SY5Y, differentiated PC-12Adh and undifferentiated PC-12Adh, respectively. After manually inspecting the MS/MS spectra, only 5 (SH-SY5Y), 2 (differentiated PC-12Adh) and 1 (undifferentiated PC-12Adh) phosphopeptides were considered valid. In Figure 3.5, the number of identified phosphopeptides is compared to the non-phosphorylated peptides identified within the phosphoenriched sample of SH-SY5Y, differentiated PC-12Adh and undifferentiated PC-12Adh. The non-phosphorylated peptides are separated into 2 categories: acidic residue-containing peptides and non-acidic residue-containing peptides. The acidic residue-containing peptides compose more than 50% of the total identified peptides for each sample. This indicates that there are non-specific interactions between the Fe^{3+} ions on the IMAC resin and the acidic residue(s) in the peptides. IMAC-Fe^{3+} phosphoenrichment is most popular technique for affinity-based enrichment of phosphorylation peptides. However, IMAC-Fe^{3+} has been known to retain non-phosphorylated peptides[3] and thus further fine-tuning of the IMAC protocol is required to achieve efficient enrichment and reproducibility[4].

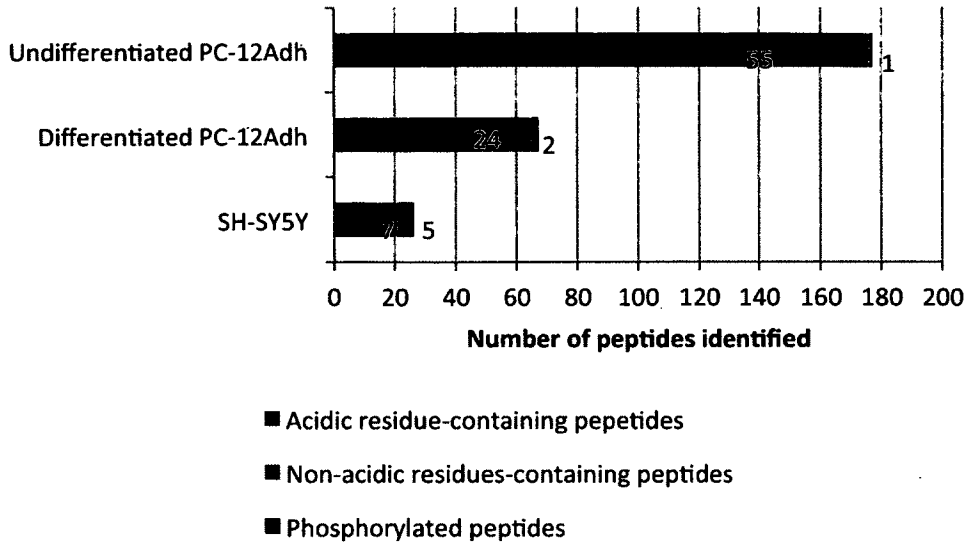


Figure 3-5: Comparison of the number of phosphorylated (black) peptides versus non-phosphorylated peptides (dark grey - acidic residue-containing peptides; light grey - non-acidic residue-containing peptides).

3.4 Quantitative analysis

Typically, the peak intensities of a light-labeled peptides and its heavy counterpart are determined from the MS spectrum and compared to attain an expression ratio of the peptide under 2 different conditions. However, for dimethyl-labeling strategy, the XIC must be used to find the maximum intensity of the light-labeled and heavy-labeled peak. The use of the XIC intensities is necessitated due the chromatographic shift commonly observed between hydrogen and deuterium labeled samples. The chromatographic shift is associated to the hydrophobic differences between hydrogen and deuterium. This isotopic effect is theoretically interpreted from the Born-Oppenheimer approximation to be due to the difference in the vibrational states of the atoms which is dependent on the mass of the

nuclei[5].The evaluation of the maximum peak intensities by using XIC will account for the variation in the isotope ratio across the elution time of the peptide.

Manual quantitation of the peptides is feasible for very small datasets however it is time-consuming and labour-intensive under-taking for large datasets. The manual quantitation process involves evaluation of the ratio of light to heavy-labeled peptides by determining the maximum intensity of the m/z peak and the total intensity of the XIC peak for both the light dimethyl-labeled and heavy dimethyl-labeled peptides from the spectral data. Mascot is used to identify the peptides within the spectral data and their modification (light or heavy dimethyl-label), elution time and charge. In the case of only one of the isotopic pair is being identified by Mascot, the m/z of the other isotopic peak can be calculated by addition (for heavy isotopic peak) or subtraction (light isotopic peak) of the difference in m/z peak to the m/z of the identified isotopic pair. The difference in the m/z between the isotopic pairs is given by

$$\Delta\left(\frac{m}{z}\right) = \frac{n\,x\,4Da}{z}$$

where n is the number of modification sites in the peptides (lysine residue and N-terminus) and z is the charge of the peptide.

An in-house quantification script was written in PHP to automate the quantitation process by computationally following the same steps of manual quantification. Comparing the average ratios calculated by automated XIC method and the maximum peak intensity method revealed that XIC method has a slight bias towards overestimation

of ratios. Therefore the ratios calculated by the maximum peak intensity method were used for further analysis.

A 99% confidence interval of $0.68 < x < 1.47$, was observed with a 1:1 ratio of light- and heavy-labeled sample of 7 digested proteins standards (these standard proteins were digested in urea lysis buffer and labeled in the same manner as the biological samples). This confidence interval accounts for the variation in the ratios arising due to biases within the experimental protocol of digestion and labeling. However, using known amounts of standard proteins to generate that 99% confidence interval cannot account for potential variation arising during cell lysis and total protein quantitation using BCA assay and the complexity of a biological sample. Therefore, a 95% confidence interval was established using the standard deviation of the ratios within each dataset. The proteins with ratios lying outside the confidence interval were considered to be differential regulated; proteins were said to be up-regulated or down-regulated if their ratios were less than the lower boundary or greater than the upper boundary, respectively, of the confidence interval. For the SH-SY5Y samples, a total of 1083 proteins were quantified. The protein ratio was evaluated as the average of the ratios for the number of quantified peptides identified for the protein. The total proteins quantified in the differentiated and undifferentiated PC-12Adh samples were 263 and 1427, respectively. As seen in Figure 3.6, <10% of total proteins are considered to be differentially regulated in each data set.

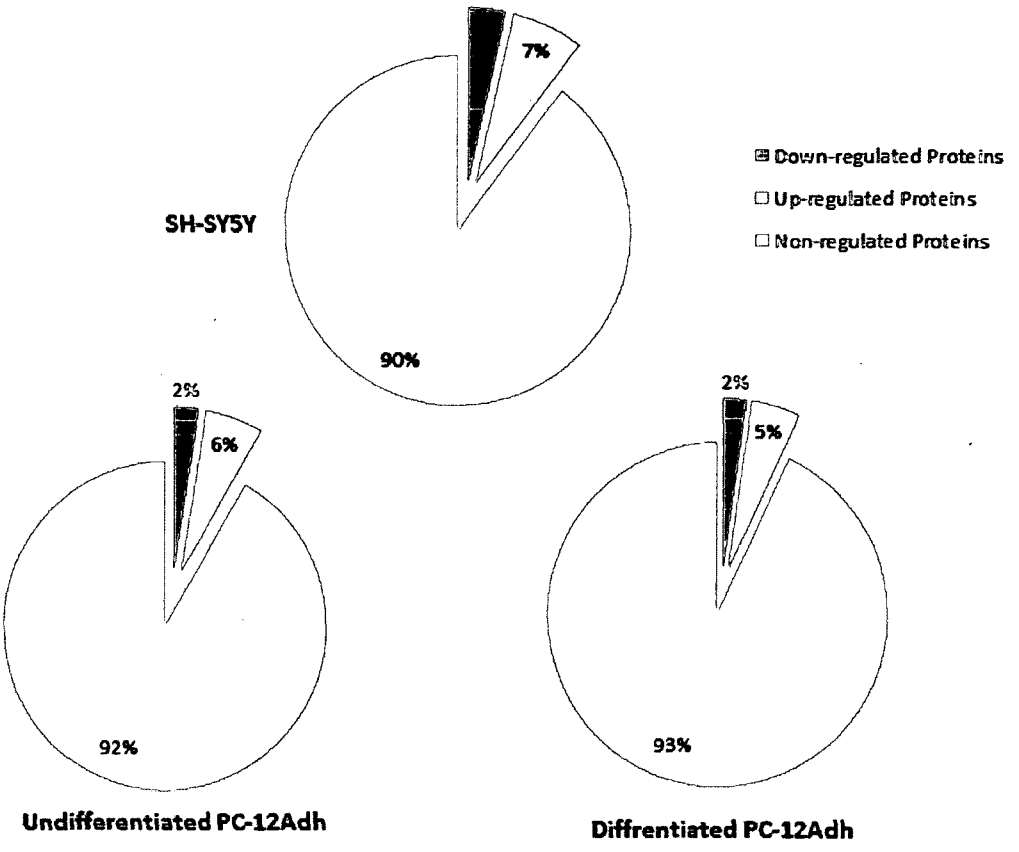


Figure 3-6: Distribution of differentially regulated (up- and down-regulated) proteins within the total quantified SH-SY5Y (top), undifferentiated PC-12Adh (bottom, left) and differentiated PC-12Adh (bottom, right) samples.

3.5 Biochemical interpretation of quantified data

Each dataset was analyzed using IPA software to determine the overlap between the proteins identified in this work and proteins associated with various cellular functions and diseases. Within the SH-SY5Y dataset, the IPA associated 304 out of 1083 proteins to cancer and 124 out of 1083 to neurological diseases. The top five molecular and cellular functions associated with the proteins in the SH-SY5Y are listed in Table3.1. The

number of proteins in the dataset involved in each function and the p-values are also indicated in the table.

Table 3-1: Summary of top molecular and cellular functional annotations within the gene list obtained from LC-MS/MS analysis of SH-SY5Y proteome.

Molecular and Cellular Function	# of molecules	p-value
RNA Post-Transcriptional Modification	49	2.54E-14
Cellular Assembly and Organization	76	9.43E-07
Cellular Function and Maintenance	75	9.43E-07
DNA Replication, Recombination and Repair	35	1.22E-06
Energy Production	19	1.22E-06

For the PC-12Adh (differentiated and undifferentiated) datasets, a poor association between the identified proteins and proteins involved in diseases was observed. In the undifferentiated PC-12Adh dataset, the top disease annotation was endocrine system disorder and it involved 5 out of 288 of the identified proteins. Whereas for the differentiated PC-12Adh dataset, the top annotation involved 13 out of 1473 identified proteins and were associated with cancer. A poor association between the identified proteins in the PC-12Adh datasets and molecular and cellular function annotations was also observed. Table 3.2 shows the top five molecular and cellular functions associated to the identified proteins within the undifferentiated and differentiated PC-12Adh dataset.

Table 3-2:Summary of top molecular and cellular functional annotations within the gene list obtained from LC-MS/MS analysis of undifferentiated and differentiated PC-12Adh proteome.

Dataset	Molecular and Cellular Function	# of molecules	p-value
Differentiated PC-12Adh	Nucleic acid Metabolism	34	4.47E-06
	Small Molecule Biochemistry	75	4.47E-06
	Carbohydrate Metabolism	32	1.21E-04
	Cellular Assembly and Organization	41	3.62E-04
	DNA Replication, Recombination and Repair	11	4.39E-04
Undifferentiated PC-12Adh	Post-Translational Modification	6	4.88E-06
	Protein Folding	5	4.88E-06
	Carbohydrates Metabolism	12	2.37E-04
	Energy Production	4	2.47E-04
	Nucleic Acid Metabolism	12	2.47E-04

The association of proteins in the three different datasets to similar molecular and cellular function annotations implies that the effect of hypoxia on the three biological samples is similar. Summary of the canonical pathway analysis (data not shown) conducted with IPA software show the involvement of the proteins identified within each dataset in various biological pathways. Most of the pathways were similar between the three datasets indicating once again the similar effect of hypoxia on the different biological samples.

To further characterize the functionally enriched annotations within the three datasets, the gene list for each dataset was analyzed with GeneCodis2.0 using biological process (BP), cellular component (CC) and molecular function (MF) gene ontology categories[6, 7]. For each dataset, only genes which showed annotations in the selected gene ontology (GO) categories were analyzed; the others were excluded. More than 70% of the genes within each datasets showed single or multiple annotations (data

not shown). The functionally enriched cellular component GO annotations were mostly cytoplasmic (>34%), cytosolic (>17%) and nuclear (>29%) cellular components, indicating the preference of soluble protein identification with the proteomic workflow in this study. Very few proteins had plasma membrane cellular component annotation. Membrane proteins are difficult to resolve due to their strong hydrophobic nature[8]. The solubilization strategies for membrane proteins involve detergents[9], organic solvents[10, 11] and/or organic acids[12]. In this work, the lysis detergent was chaotrope-based only without any detergents hence the proteomes analyzed were deficient in membrane proteins.

The molecular functions that are enriched amongst the three dataset were protein binding, nucleotide binding and ATP binding while the biological process annotations included glycolysis, RNA splicing and translational elongation. To further decipher the implications of hypoxia on mammalian cells, the differentially regulated proteins within each dataset were functionally enriched. The differentially regulated proteins within the undifferentiated PC-12Adh dataset had no significantly enriched annotation(s). However, the differentially regulated proteins in the differentiated PC-12 Adh dataset were mostly associated with binding molecular function and cell differentiation biological process, as seen in Figure 3.7. This dataset was resulting from neuronally differentiated PC-12Adh cells, hence the representation of cell differentiation biological process annotation.

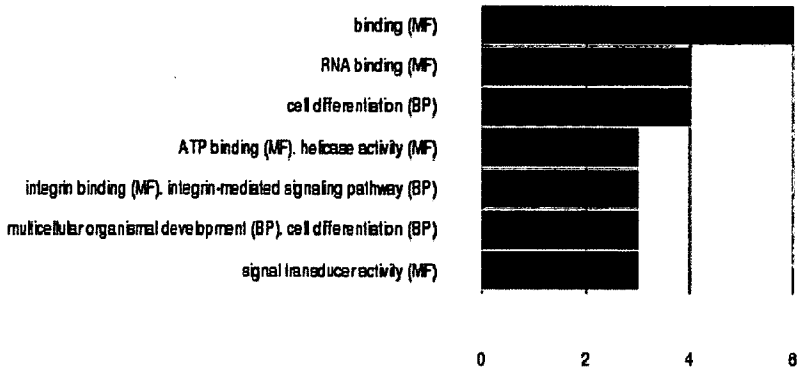


Figure 3-7: Number of genes functionally enriched within the differentially regulated proteins of differentiated PC-12Adh sample. Figure generated by GeneCodis2.0.

The top 6 functionally enriched annotations for SH-SY5Y dataset were associated to binding molecular function, as shown in Figure 3.8. The top 2 functionally enriched biological process annotations are signal transduction and cell proliferation, also shown in the figure. The complete list of gene annotations is represented graphically in Appendix D.

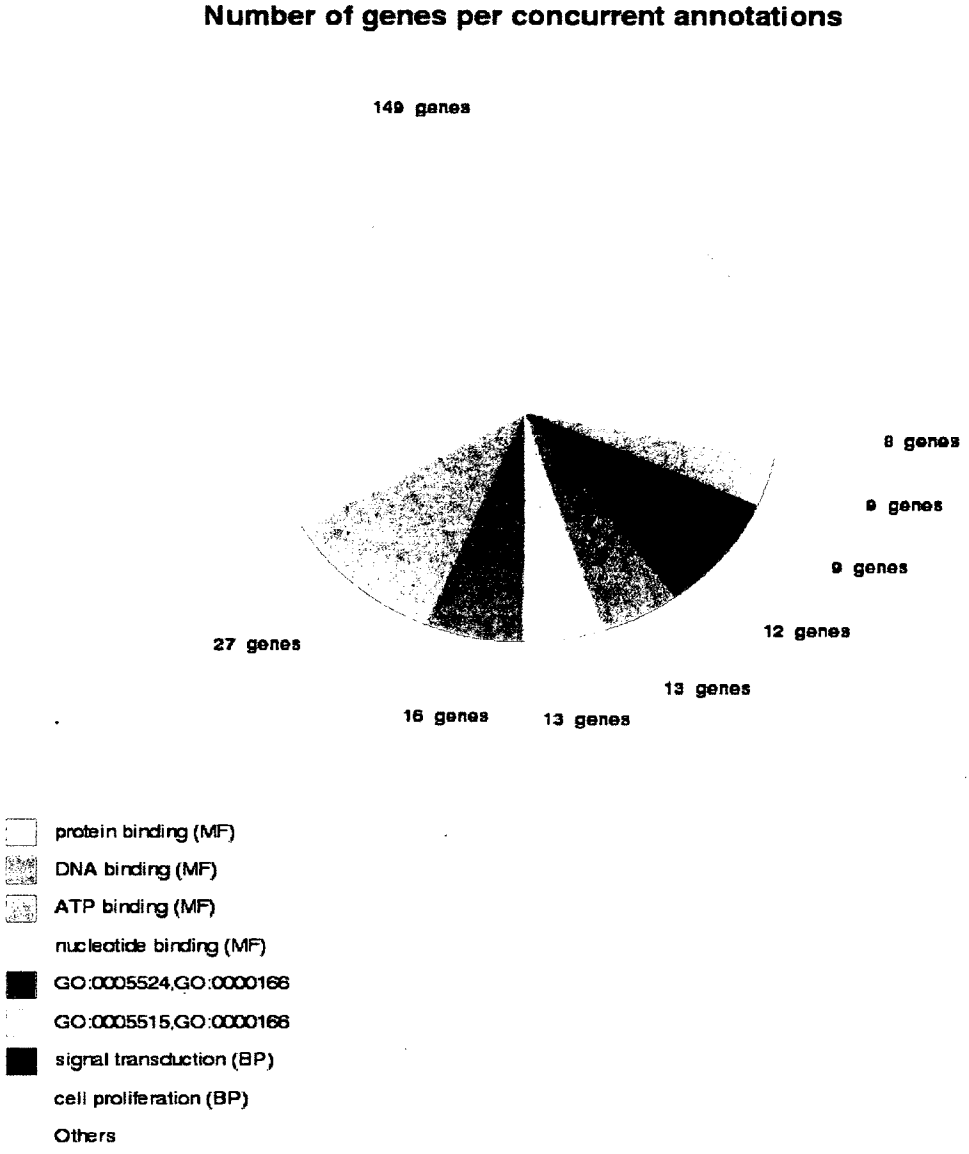


Figure 3-8: Functionally enriched annotations within the differentially regulated proteins of SH-SY5Y dataset. The gene ontology annotations for ATP binding, nucleotide binding and protein binding are G):0005524, GO:0000166 and GO:0005515.

It is noted that there is functional enrichment of signal transduction annotation within the differentially regulated SH-SY5Y and differentiated PC-12Adh dataset. Typically, signal transduction in a cell involves post-translational modifications and

protein-protein interaction in order to generate an appropriate response[13]. To elucidate the potential interactions within the differentially regulated dataset of SH-SY5Y sample and neuronal PC-12Adh sample, the gene list of those proteins were analyzed by GeneMania[14]. The differentially regulated proteins in the neuronal PC-12Adh sample did not show any protein-protein interaction, hence not shown. The up- and down- regulated proteins within the SH-SY5Y were analyzed by GeneMania, separately. The down-regulated protein did not show significant interactions. However, the up-regulated proteins showed interactions amongst certain proteins as shown in Figure 3.9.

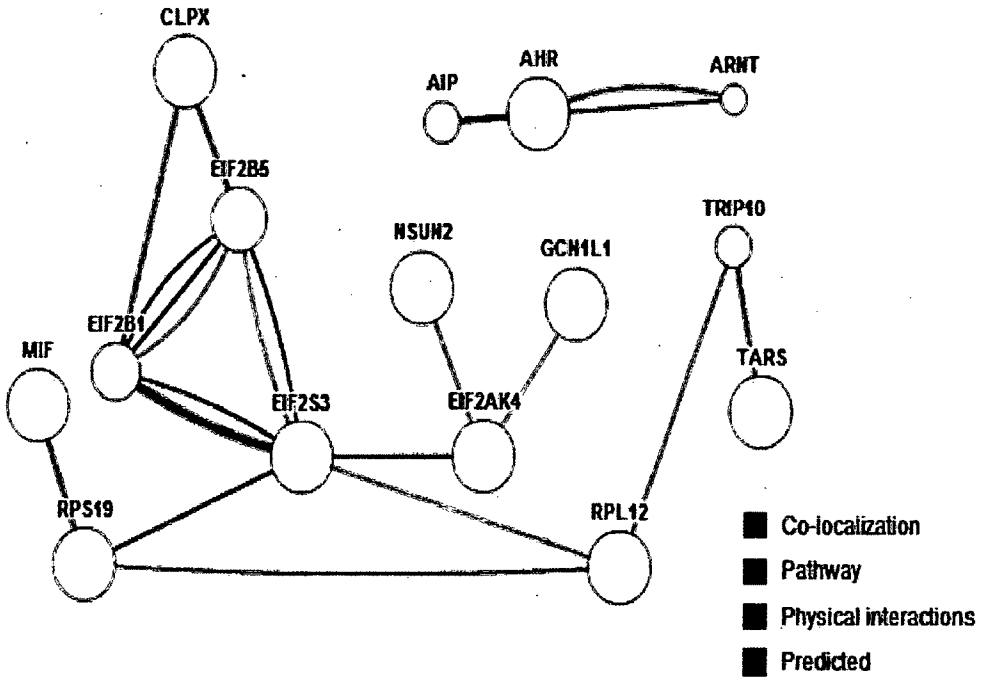


Figure 3-9: Interaction network for up-regulated proteins within the SH-SY5Y dataset. The grey circles represent genes of identified proteins within the dataset and white circles are predicted genes. Figure generated by GeneMania.

In Figure 3.9, two genes are connected by pathway links if the gene products participate in the same reaction in a pathway. The eukaryotic translation initiation factor 2, subunit3 (EIF2S3) is shown to be involved in the same pathway as eukaryotic translation initiation factor 2α kinase 4 (EIF2AK4) which in turn is predicted to be associated with general control of amino-acid synthesis 1-like (GCN1L1, also known as GCN1) and NOP2/Sun domain family member 2 (NSUN2, also known as Misu). The EIF2S3 is predicted to interact with ribosomal protein L12 (RPL12). To know the nature of the predicted interaction amongst the proteins, it is vital to know the role of the proteins and the pathway they interact in.

The EIF2S3 and EIF2AK4 (also known as GCN2) are associated with the translational initiation during protein synthesis[15-17]. Under hypoxia, the lack of ATP directly affects the global translational machinery mainly by modification of translation initiation factors[15, 18]. There are three main steps to translation: initiation, elongation and termination. Each of these steps involves translation factors which associate with ribosomes, transiently, during the translational process of mRNA into protein. There are three class of translation factors: eukaryotic translation initiation, elongation and release/terminaton factors (eIFs, eEFs and eRFs, respectively)[19]. The elongation process consumes almost 99% of the energy needed for translation[19], hence the major site of mRNA translation control is thought to be at the translational initiation step (48, 49). Under hypoxia, the control of mRNA translation has been attributed to phosphorylation of eIF2α[17].

Role of eIF2

eIF2 is a heteromeric protein consisting of 3 subunits termed α (also known as subunit 1, S1), β (also known as subunit 2, S2) and γ (also known as subunit 3,S3) [20]. On the α-subunit, there is a single phosphorylation site, Ser51, where 3 protein kinases act on [20]. The protein kinases include heme-regulated inihibitor (HRI), double-stranded RNA-activated protein kinase (PKR) and the nutrient-regulated protein kinase (GCN2).

The β-subunit has multiple phosphorylation sites for casein kinase II (CK-II), protein kinase C (PKC) and cAMP-dependent protein kinase (PKA)[21]. Additionally, it has 3 poly-lysine domains and 2 consensus guanine nucleotide binding domains on its N-terminal[22, 23]. The second poly-lysine domain is the binding site for initiation factor, eIF5 (5). Even though eIF2β has guanine nucleotide binding domains, they are not involved in guanine nucleotide binding[23]. However, the eIF2γ contains the primary guanine nucleotide binding site and contains all three consensus guanine nucleotide-binding domains[23]. The N-terminus of the γ subunit will bind to methionyl-tRNA$_i$ (Met-tRNA$_i$)[24].

The eIF2 will form a ternary complex with GTP and Met-tRNA$_i$ which then binds to the 40S ribosomal subunit[25]. Other initiation factors, eIF1, eIF1A and eIF3 will be recruited at the 40S subunit to form 42S pre-initiation complex[25]. The pre-initiation complex is then recruited at the 5'-cap of the mRNA by the mRNA 5'-cap-binding complex eIF4F which consists of eIF4G, eIF4E and eIF4A. The eIF4E binds to the 5'-cap and the eIF4G, which is a scaffold protein that binds to eIF4A and eIF3. Once the AUG start codon has been localized, the 60S ribosomal subunit will be recruited at the 43S pre-initiation complex [25]. The eIF5 will hydrolyze the GTP from the GTP-bound eIF2 and

thus releasing all the initiation complexes from the 40S subunit and propagating the translational process into the elongation step [25].

In order for GDP-bound eIF2 to bind to Met-tRNAi for the next cycle of initiation, the GDP has to be replaced with GTP[26]. Guanine nucleotide exchange factor, eIF2B, will catalyze the GDP-GTP exchange on the eIF2[26]. However, when the Ser51 residue on the eIF2α is phosphorylated, eIF2 acts as a competitive inhibitor of eIF2B and thus inhibiting translational initiation[26].

Under stressful conditions, phosphorylation of eIF2 allows the cells to adapt to the conditions by selective translation of transcription factors[27]. The transcription factors, activating transcription factor 4 (ATF4) [28] and activating transcription factor 5 (ATF5) [29] induce expression of genes that facilitate adaptation.

Interpretation of regulated proteins

In this work, the γ subunit of eIF2 has been observed to be up-regulated 3.4 fold while the GCN2 is up-regulated 12.8 fold. The translational activator, GCN1 is shown to be up-regulated 7.6 fold. GCN1 is part of a protein complex which activates the kinase activity of GCN2 in amino acid-starved cells by phophorylation[30]. It has been shown that hypoxia inhibits transporters of selective amino acids, for instance, suppressed L-arginine uptake in pulmonary artery endothelial cells[31] and L-proline in lung firboblasts[32]. Therefore it could be postulated that under hypoxia, the transport of amino acid(s) was impeded and hence an up-regulation of GCN1 and GCN2 is observed within the data. The up-regulation for GCN2 could be in form of over-activation (increased phosphorylation) or over-expression (increased number of copies). It is

difficult to assign the type of up-regulation in this work do to the poor outcome of phosphoenrichment. In spite of the type of up-regulation, an up-regulation of GCN1 could imply there is increase activation of GCN2 and thus increased phosphorylation of eIF2.

3.6 Summary

The quantitative proteomic workflow was applied to characterize the changes induced in the proteome of human neuroblastoma cells (SH-SY5Y), rat pheochromocytoma cell (PC-12Adh) and neuronal rat pheochromocytoma cells (PC-12Adh) after treatment to hypoxia for 24 hours. The workflow incorporated cell lysis, tryptic-digestion and dimethyl labeling of peptides, followed by enrichment of phosphopeptides by IMAC and separation of non-phosphorylated peptides by SCX/SAX and finally, analysis by shotgun mass spectrometry on a hybrid quadrupole-time-of-flight instrument. The optimization of the workflow proved to be successful at increasing the dataset of identified proteins. The SCX and SAX separation techniques showed their complementarity with <2% peptide redundancy between them. An in-house computational quantification script was utilized to compute the relative amounts of labeled proteins in the hypoxia-treated sample and its appropriate control for each biological sample. Less than 10% of the total identified proteins were observed to be differentially regulated within each biological sample. Amongst, the differentially regulated proteins, approximately 70% of the population was observed to be up-regulated while the remaining was down-regulated. Analysis of the datasets using bioinformatic

approaches indicated the preferential synthesis of proteins involved in hypoxia adaptation processes. The limitations and future work is discussed in the following section.

3.7 Limitations and Future work

Mass spectrometry-based shotgun proteomics generates large amounts of data which needs to be collaborated with high-throughput bioinformatic tools to decipher protein-protein interaction networks and signaling pathways in order to generate hypotheses for biological/biochemical studies. However for this work, the major impedance was the limited choice of bioinformatic tools. Only a small set of proteins within a dataset could be analyzed at a time with GeneMania and hence limiting the potential interactions observed within the data. Better bioinformatic tools would be capable of generating hypotheses from as large datasets.

A computational quantification script was utilized to compute the ratios of the labeled peptides for each biological sample. After computing the ratios within each fraction analyzed, the data was validated with an in-house filter to eliminate a poor identified peptide/protein. The validated quantified data was then pooled together to compute the protein ratios. A confidence interval was calculated based on the standard deviation observed within the datasets. Such a confidence interval didn't account for variations/biases arising from the experimental protocol. To overcome this, a confidence interval should be computed using ratios of labeled peptides from identical amounts of identical cell sample processed with the workflow used in this study. Such a stringent confidence level would improve the assignation of up- and down-regulated proteins within the dataset.

Typically, spectral data from various fractions are collated together in from of text file (.mgf) and searched against a database using Mascot. This allows for better assignment of peptides to proteins. However, the overall Mascot search needs to be filtered to validate the identified sequences. However, due to the large datasets, the filter was incapable of processing the data. Improvements in the computational software's would allow confident identification of proteins and better correlation to theoretical knowledge.

Replicates of experiments would not only increase the dataset but it will also improve the confidence in the identified proteins, more specifically for differentially regulated proteins. Repeat analysis of the undifferentiated PC-12Adh cells would generate a complete dataset and enable lateral comparison with SH-SY5Y dataset and differentiated PC-12Adh dataset to elucidate the effects of hypoxia on undifferentiated cells and neuronal cells, respectively.

IMAC is the most widely used phosphoenrichment strategy; however it is reputed for non-specific binding to non-phosphorylated peptides such as acidic peptides. Another option to attempt analysis of phosphoproteome could be using titanium dioxide beads which are highly specific[4, 33]. Phosphoproteome analysis on a hypoxia-treated cellular proteome would reveal potential signaling and regulatory proteins involved in cellular adaptation to low oxygen.

3.8 References

1. McCormack, A. L. et al. Direct analysis and identification of proteins in mixtures by LC/MS/MS and database searching at the low-femtomole level Anal. Chem. 69, 767-776 (1997).
2. Hattan, S. J., Marchese, J., Khainovski, N., Martin, S. & Juhasz, P. Comparative study of [three] LC-MALDI workflows for the analysis of complex proteomic samples. J. Proteome Res. 4, 1931-1941 (2005).
3. Han, G. et al. Large-scale phosphoproteome analysis of human liver tissue by enrichment and fractionation of phosphopeptides with strong anion exchange chromatography. Proteomics 8, 1346-1361 (2008).
4. Nuhse, T. S., Stensballe, A., Jensen, O. N. & Peck, S. C. Large-scale analysis of in vivo phosphorylated membrane proteins by immobilized metal ion affinity chromatography and mass spectrometry. Mol. Cell. Proteomics 2, 1234-1243 (2003).
5. Turowski, M. et al. Deuterium isotope effects on hydrophobic interactions:The importance of dispersion interactions in the hydrophobic phase. J. Am. Chem. Soc. 125, 13836 (2003).
6. Carmona-Saez, P., Chagoyen, M., Tirado, F., Carazo, J. M. & Pascual-Montano, A. GENECODIS: a web-based tool for finding significant concurrent annotations in gene lists. Genome Biol. 8, R3 (2007).
7. Nogales-Cadenas, R. et al. GeneCodis: interpreting gene lists through enrichment analysis and integration of diverse biological information. Nucleic Acids Res. 37, 317-322 (2009).
8. Chen, E. I., Hewel, J., Felding-Habermann, B. & Yates, J. R. Large scale protein profiling by combination of protein fraction and multidimensional protein identification technology (MudPIT). Mol. Cell. Proteomics 5, 53-56 (2005).
9. Han, D. K., Eng, J., Zhou, H. & Aebersold, R. Quantitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry. Nat. Biotechnol. 19, 946-951 (2002).
10. Blonder, J. et al. Enrichment of integral membrane proteins for proteomic analysis using liquid chromatography-tandem mass spectrometry. J. Prot. Res. 1, 351-360 (2002).
11. Goshe, M. B., Blonder, J. & Smith, R. D. Affinity labeling of highly hydrophobic integral membrane proteins for proteome-wide analysis. J. Prot. Res. 2, 153-161 (2003).
12. Washburn, M. P., Wolters, D. & Yates, J. R. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nat. Biotechnol. 19, 242-247 (2001).

13. Tang, W. et al. Proteomics studies of brassinosteroid signal transduction using prefractionation and two-dimensional DIGE. Mol. Cell. Proteomics 7, 728-738 (2008).
14. Warde-Farley, D. et al. The GeneMANIA prediction server: biological network intergration for gene prioritization and predicting gene function. Nucleic Acids Res. 38, 214-220 (2010).
15. Liu, L. et al. Hypoxia-induced energy stress regulates mRNA translation and cell growth. Mol. Cell. 21, 521-531 (2006).
16. Koritzinsky, M. et al. Gene expression during acute and prolonged hypoxia is regulated by distinct mechanisms of translational control. EMBO J. 25, 1114-1125 (2006).
17. Blais J. D. et al. Activating transcription factor 4 is translationally regulated by hypoxic stress. Mol. Cell Biol. 24, 7469-7482 (2004).
18. Thomas, J. D. & Johannes, G. J. Identification of mRNAs that continue to associate with polysomes during hypoxia. RNA. 13, 1116-1131 (2007).
19. Proud, C. G. Signalling to translation: how signal transduction pathways control the protein synthetic machinery. Biochem J. 403, 217-234 (2007).
20. Kimball, S. R. Eukaryotic initiation factor eIF2 Int. J. Biochem. Cell Biol. 31, 25 <last_page> 29-29 (1999).
21. Welsh, G. I., Price, N. T., Bladergroen, B. A., Bloomberg, G. & Proud, C. G. Identification of novel phosphorylation sites in the b-subunit of translation initiation factor eIF-2. Biochem. Biophys. Res. Commun. 201, 1279-1288 (1994).
22. Das, S., Maiti, T., Das, K. & Maitra, U. Specific interaction of eukaryotic translation initiation factor 5 (eIF5) with the β-subunit of eIF2. J. Biol. Chem. 272, 31712-31728 (1997).
23. Naranda, T., Sirangelo, I., Fabbri, B. J. & Hershey, J. W. B. Mutations in the NKXD consensus element indicate that GTP binds to the ?-subunit of translation initiation factor eIF2. FEBS Lett. 372, 249 (1995).
24. Gaspar, N. J. et al. Translational initiation factor eIF-2. Cloning and expression of the human cDNA encoding the g-subunit. J. Biol. Chem. . 269, 3415-3422 (1994).
25. Preiss, T. & W Hentze, M. Starting the protein synthesis machine: eukaryotic translation initiation. Bioessays 25, 1201-1211 (2003).
26. Kleijn, M. & Proud, C. G. The activation of eukaryotic initiation factor (eIF)2B by growth factors in PC12 cells requires MEK/ERK signalling. FEBS Lett. 476, 262-265 (2000).
27. Wek, R. C., Jiang, H. Y. & Anthony, T. G. Coping with stress: eIF2 kinases and translational control. Biochem. Soc. Trans. 34, 7-11 (2006).

28. Vattem, K. M. & Wek, R. C. Reinitiation involving upstream ORFs regulates ATF4 mRNA translation in mammalian cells. Proc. Natl. Acad. Sci. U. S. A. 101, 11269-11274 (2004).

29. Zhou, D. et al. Phosphorylation of eIF2 directs ATF5 translational control in response to diverse stress conditions.. J. Biol. Chem. 283, 7064-7073 (2008).

30. Vazquez de Aldana, C. R., Marton, M. J. & Hinnebusch, A. G. GCN20, a novel ATP binding cassette protein, and GCN1 reside in a complex that mediates activation of the eIF2-alpha kinase GCN2 in amino acid-strved cells. EMBO J. 14, 3194-3199 (1995).

31. Block, E. R., Herrera, H. & Couch, M. Hypoxia inhibits L-arginine uptake by pulmonary artery endothelial cells. Am. J. Physiol. 269, L574-L580 (1995).

32. Berk, J. L., Hatch, C. A. & Goldstein, R. H. Hypoxia inhibits amino acid uptake in human lung fibroblasts. J. Appl. Physiol. 89, 1425-1431 (2000).

33. Larsen, M. R., Thingholm, T. E., Jensen, O. N., Roepstorff, P. & Jorgensen, T. J. Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns Mol. Cell. Proteomics 4, 873-886 (2005).

4 Chapter: Appendices

Appendix A

Table 4.1 and 4.2 list of differentially up- and down- regulated proteins observed within undifferentiated PC-12Adh neuronal cells upon treatment of hypoxia.

Table 4-1: Up-regulated proteins in undifferentiated PC-12Adh sample.

Protein Name	Gene Symbol	Normalized Ratio
Acyl-CoA-binding protein	LOC100365425;Dbi	0.116366
BEIGE	Lyst	0.129296
Isoform 1 of Glutamate receptor-interacting protein 1	Grip1	0.116366
LAG1 homolog, ceramide synthase 5-like	Cers6	0.172394
Aire autoimmune regulator	Aire	0.133605
Transmembrane protein 164	Tmem164	0.172394
Anaphase promoting complex subunit 2	Anapc2	0.185324
Interferon kappa	Ifnk	0.120676
Transforming, acidic coiled-coil containing protein 2	Tacc2	0.150845
Tubulin, delta 1-like	LOC100363487	0.120676

Table 4-2: Down-regulated proteins in undifferentiated PC-12Adh sample

Protein Name	Gene Symbol	Normalized Ratio
SET and MYND domain containing 1	Smyd1	25.26003
Sialic acid binding Ig-like lectin 10	Siglec10	13.05023
Glucocorticoid induced transcript 1	Glcci1	12.63217
Membrane primary amine oxidase	Aoc3	9.908345
ATP-binding cassette, sub-family A (ABC1), member 13	Abca13	8.951558
Family with sequence similarity 71, member E2	Fam71e2	5.383003

Appendix B

Table 4.3 and 4.5 present a list of differentially up- and down- regulated proteins observed within differentiated PC-12Adh neuronal cells upon treatment of hypoxia.

Table 4-3: Up-regulated proteins in the differentiated PC-12Adh sample.

Protein Name	Gene Symbol	Normalized Ratio
Nuclear fragile X mental retardation-interacting protein 1	Nufip1	0.131837
7SK snRNA methylphosphate capping enzyme	Mepce	0.125559
Calicin	Ccin	0.144393
Cd300 molecule-like family member G isoform 1	Cd300lg	0.182061
coiled-coil domain containing 150-like	Ccdc150	0.175783
Connective tissue growth factor	Ctgf	0.175783
CUB and Sushi multiple domains 3-like isoform 3	LOC314942	0.182061
CWF19-like 2, cell cycle control	Cwf19l2	0.200895
DIS3 mitotic control homolog (S. cerevisiae)-like 2 isoform 2	Dis3l2	0.056502
GTP-binding protein 1	Gtpbp1	0.169505
Heat shock protein beta-3	Hspb3	0.119281
Heme oxygenase 1	Hmox1	0.182061
Intermediate conductance calcium-activated potassium channel protein 4	Kcnn4	0.156949
Isoform 1 of Nicastrin	Ncstn	0.188339
Isoform Alpha-7X1B of Integrin alpha-7	Itga7	0.169505
Kinesin-like protein KIF15	Kif15	0.056502
Liver carboxylesterase 4	Ces1f	0.200895
Mitogen-activated protein kinase 12	Mapk12	0.131837
Myosin light chain kinase 2, skeletal/cardiac muscle	Mylk2	0.175783
non-SMC condensin II complex, subunit D3	Ncapd3	0.125559
Nuclear receptor coactivator 2	Ncoa2	0.075336
Oligosaccharyltransferase complex subunit OSTC	Ostc	0.169505
PDZ domain containing 7 (predicted)-like	LOC681677	0.200895
peptidase inhibitor 16	Pi16	0.043946
peptidyl-prolyl cis-trans isomerase FKBP3	Fkbp3	0.150671
PHD finger protein 16	RGD1563945	0.081614
phosphatidylinositol glycan, class Z isoform 1	Pigzl1	0.156949
PIN2/TERF1-interacting telomerase inhibitor 1	Pinx1	0.050224
Polypeptide N-acetylgalactosaminyltransferase 10	Galnt10	0.125559
Protein turtle homolog A	Igsf9	0.144393

Rab effector MyRIP	Myrip	0.119281
retinoblastoma-associated protein 140-like isoform 1	LOC680155	0.03139
RPTOR independent companion of MTOR, complex 2	Rictor	0.043946
seizure threshold 2	Szt2	0.131837
Solute carrier family 23 member 2	Slc23a2	0.03139
Sphingosine-1-phosphate lyase 1	Sgpl1	0.144393
Transmembrane and ubiquitin-like domain-containing protein 2	Tmub2	0.018834
E3 ubiquitin-protein ligase MIB2	LOC474147	0.018834
benzodiazapine receptor associated protein 1	Bzrap1	0.119281
Uncharacterized protein	=-	0.06278
lipoyltransferase 1	Lipt1	0.194617
spermine oxidase	Smox	0.119281
G protein-coupled receptor 162	Gpr162	0.138115
gamma-glutamyl cyclotransferase	Ggct;Ggctl1	0.06278
intraflagellar transport protein 74 homolog	Ift74	0.069058
matrilin-3	Matn3	0.163227
ankyrin repeat and SAM domain-containing protein 1A	Anks1a	0.069058
holocytochrome-c synthase	Hccs	0.138115
endoplasmic reticulum protein 44	Erp44	0.125559
Uncharacterized protein	Vom2r33	0.194617
ATP-dependent RNA helicase	Ddx18	0.081614
uncharacterized protein LOC31412	RGD1304624	0.119281
Uncharacterized protein	Myh14	0.207173
serine/arginine repetitive matrix	Srrm2	0.150671
uncharacterized protein LOC302773	RGD1563104	0.200895
molybdenum cofactor biosynthesis protein 1	Mocs1	0.169505
protein EFR3 homolog A	Efr3a	0.06278
neutral amino acid transporter A	Slc1a4	0.043946
rCG53773-like	LOC100363783	0.056502
phosphatidylinositol-4-phosphate 3-kinase C2 domain-containing subunit gamma	Pik3c2g	0.169505
Uncharacterized protein	=-	0.056502
histone-lysine N-methyltransferase SETD2	Setd2	0.188339
Uncharacterized protein	=-	0.075336
ELAV-like protein 1	Elavl1	0.188339
probable global transcription activator SNF2L1	Smarca1	0.03139
Uncharacterized protein	=-	0.07115
Uncharacterized protein	LOC499643	0.138115
Uncharacterized protein (Fragment)	=-	0.172644

Uncharacterized protein (Fragment)	=-	0.150671

Table 4-4: Down-regulated proteins in differentiated PC-12Adh sample.

Protein Name	Gene Symbol	Normalized Ratio
Uncharacterized protein	Prc1	9.912915
Peroxisomal membrane protein 11A	Pex11a	4.714755
RT1 class I, A3	RT1-A3	9.190948
Very long-chain acyl-CoA synthetase	Slc27a2	12.58733
Isoform 1 of Ameloblastin	Ambn	5.04121
Histone H3.3	H3f3b; LOC100361558; LOC100365096	5.700397
solute carrier family 5 (inositol transporters), member 3	Slc5a3	7.45195
Uncharacterized protein	Eaf1	27.71724
Uncharacterized protein	Urb2	5.650173
Cardiac troponin C	Tnnc1	13.60436
Uncharacterized protein	Filip1l	5.141657
Uncharacterized protein	Zfp710	6.478865
Uncharacterized protein	RGD1561916	5.989183
Collagen alpha-1(III) chain	Col3a1	9.435789
MAM domain-containing glycosylphosphatidylinositol anchor protein 1	Mdga1	29.27417
Protein unc-45 homolog A	Unc45a	8.669877
Uncharacterized protein	Uvrag	6.315638
Uncharacterized protein	Unc45b	8.41248
Glyceraldehyde-3-phosphate dehydrogenase	=-	33.58086
Origin recognition complex subunit 4	Orc4	10.71022
Dermcidin	=-	11.71469
hypothetical protein	LOC100364723	6.315638
Uncharacterized protein (Fragment)	=-	11.9972
Uncharacterized protein	=-	15.1613
hypothetical protein	Fam193a	6.930879
UPF0723 protein C11orf83 homolog	LOC690344;LOC686324	13.87431
A disintegrin and metalloproteinase with thrombospondin motifs 14	Adamts14	4.865427

Uncharacterized protein (Fragment)	Unc5d	6.585591
nascent polypeptide-associated complex subunit alpha isoform 2	Naca	8.334005
Uncharacterized protein	Chd4	5.461834

Appendix C

Table 4.5 and 4.6 list the differentially up- and down-regulated proteins observed within SH-SY5Y neuronal cells upon treatment of hypoxia.

Table 4-5: Up-regulated proteins in the SH-SY5Y sample

Protein Name	Gene Symbol	Normalized Ratio
N(G),N(G)-dimethylarginine dimethylaminohydrolase 2	DDAH2	0.298038
14 kDa protein	RANBP1	0.289272
40S ribosomal protein S20	SNORD54;RPS20	0.27174
Aldehyde dehydrogenase family 1 member A3	ALDH1A3	0.201614
Aldose reductase	AKR1B1	0.298038
Aryl hydrocarbon receptor	AHR	0.219146
ATP-binding cassette sub-family B member 10, mitochondrial	ABCB10	0.184082
ATP-dependent Clp protease ATP-binding subunit clpX-like, mitochondrial	CLPX	0.306804
Cardiomyopathy-associated protein 5	CMYA5	0.184082
cDNA FLJ57349	PTCD3	0.043829
cDNA FLJ61112, highly similar to BTB/POZ domain-containing protein KCTD15	KCTD15	0.166551
Cofilin-2	CFL2	0.306804
DnaJ homolog subfamily B member 11	DNAJB11	0.298038
DnaJ homolog subfamily C member 14	DNAJC14	0.175316
Doublesex- and mab-3-related transcription factor A1	DMRTA1	0.306804
E3 ubiquitin-protein ligase MIB2 isoform 3	MIB2	0.157785
Eukaryotic translation initiation factor 2 subunit 3	EIF2S3	0.289272
FERM and PDZ domain-containing protein 1	FRMPD1	0.219146
Histone H1.3	HIST1H1D	0.306804
Histone H1.4	HIST1H1E	0.192848
Histone H2B type 1-A	HIST1H2BA	0.184082
Histone H2B type 1-K	HIST1H2BK	0.149019
Isoform 1 of 60S ribosomal protein L12	RPL12	0.306804
Isoform 1 of Calcium/calmodulin-dependent 3',5'-cyclic nucleotide phosphodiesterase 1A	PDE1A	0.122721

Isoform 1 of Cdc42 effector protein 1	CDC42EP1	0.280506
Isoform 1 of Coiled-coil domain-containing protein 116	CCDC116	0.100807
Isoform 1 of Eukaryotic translation initiation factor 2-alpha kinase 4	EIF2AK4	0.078892
Isoform 1 of G protein-regulated inducer of neurite outgrowth 1	GPRIN1	0.254209
Isoform 1 of Histone-lysine N-methyltransferase MLL	MLL	0.175316
Isoform 1 of Outer dense fiber protein 2-like	ODF2L	0.192848
Isoform 1 of Rho guanine nucleotide exchange factor 10	ARHGEF10	0.026297
Isoform 1 of SH3 domain-binding protein 4	SH3BP4	0.298038
Isoform 1 of Sialic acid-binding Ig-like lectin 10	SIGLEC10	0.262975
Isoform 1 of Transcription factor E2F7	E2F7	0.21038
Isoform 1 of WD repeat-containing protein 52	WDR52	0.306804
Isoform 2 of Formin-binding protein 1	FNBP1	0.10519
Isoform 2 of Golgin subfamily A member 2-like protein 2	AGSK1;LOC727849	0.27174
Isoform 2 of Rho GTPase-activating protein 5	ARHGAP5	0.245443
Isoform 2 of Rho guanine nucleotide exchange factor 1	ARHGEF1	0.236677
Isoform 5 of WD repeat- and FYVE domain-containing protein 4	WDFY4	0.289272
Isoform Long of Uncharacterized protein C21orf2	C21orf2	0.192848
jerky protein homolog isoform a	JRK	0.298038
Macrophage migration inhibitory factor	MIF	0.201614
MARCKS-related protein	MARCKSL1	0.306804
NEDD4-like E3 ubiquitin-protein ligase WWP2	WWP2	0.306804
Non-histone chromosomal protein HMG-17	HMGN2	0.236677
Osteopetrosis-associated transmembrane protein 1	OSTM1	0.236677
Parathymosin	PTMS	0.298038
Peroxiredoxin-1	PRDX1	0.227911
Pre-mRNA-splicing factor SLU7	SLU7	0.254209
Protein AMN1 homolog	AMN1	0.166551
Protein cTAGE-2	CTAGE1	0.298038
Putative uncharacterized protein DKFZp761P221 (Fragment)	ARHGEF10	0.245443
SLIT-ROBO Rho GTPase-activating protein 2	SRGAP2	0.078892
Src substrate cortactin	CTTN	0.070127

Stathmin (Fragment)	STMN1	0.280506
Threonyl-tRNA synthetase, cytoplasmic	TARS	0.201614
TPR and ankyrin repeat-containing protein 1	TRANK1	0.149019
Translational activator GCN1	GCN1L1	0.131487
tRNA (cytosine(34)-C(5))-methyltransferase	NSUN2	0.306804
centriolin	CNTRL	0.122721
tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein	YWHAB	0.219146
N-lysine methyltransferase SETD6	SETD6	0.262975
1-phosphatidylinositol-3-phosphate 5-kinase	PIKFYVE	0.245443
TATA element modulatory factor	TMF1	0.262975
X-ray repair cross-complementing protein 5	XRCC5	0.245443
Zinc finger protein 831	ZNF831	0.21038
Oxidation resistance protein 1	OXR1	0.262975
hypothetical protein LOC100288412	LOC100288412	0.078892
DNA polymerase subunit gamma-2	POLG2	0.298038
heterogeneous nuclear ribonucleoprotein A1 pseudogene	LOC120364	0.061361
#N/A	LOC729501	0.166551

Table 4-6: Down-regulated proteins in the SH-SY5Y

Protein Name	Gene Symbol	Normalized Ratio
Leucine-rich repeat transmembrane protein FLRT2	FLRT2	3.874492
Isoform Long of Antigen KI-67	MKI67	4.27772
Isoform 1 of Myb-binding protein 1A	MYBBP1A	6.495473
U4/U6 small nuclear ribonucleoprotein Prp3	PRPF3	3.743005
Probable ATP-dependent RNA helicase DDX20	DDX20	3.260885
Similar to Integrin alpha-4 subunit	ITGA4	3.532625
Isoform 2 of Zinc finger protein 43	ZNF43	3.444967
Brain-specific angiogenesis inhibitor 1	BAI1	3.444967
Isoform 1 of Kinesin-like protein KIF15	KIF15	7.38082

Zinc finger MYND domain-containing protein 15	ZMYND15	9.08139
Isoform 1 of Interleukin-1 receptor accessory protein	IL1RAP	3.532625
Isoform 1 of Oligoribonuclease, mitochondrial (Fragment)	REXO2	9.668699
TMEM181 protein (Fragment)	TMEM181	3.997214
Isoform 1 of Genetic suppressor element 1	KIAA0182	6.679555
Isoform 1 of Armadillo repeat-containing protein 3	ARMC3	3.576455
Isoform 1 of Centromere-associated protein E	CENPE	5.355916
Isoform 1 of Protein HEG homolog 1	HEG1	3.357309
Isoform 1 of Scavenger receptor class A member 3	SCARA3	3.444967
Taste receptor type 2 member 42	TAS2R42	4.69848
Isoform 1 of RanBP2-like and GRIP domain-containing protein 4	RGPD4	12.97341
Isoform 2 of LSM domain-containing protein 1	LSMD1	3.532625
Isoform 2 of Uncharacterized protein C3orf67	C3orf67	6.35522
Isoform 1 of KN motif and ankyrin repeat domain-containing protein 4	KANK4	10.16835
Isoform 2 of GRIP1-associated protein 1	GRIPAP1	3.646581
ras-specific guanine nucleotide-releasing factor 1	RASGRF1	5.285789
Isoform 1 of Synaptonemal complex protein 2-like	SYCP2L	6.574365
411 kDa protein	SMG1	5.899397
Methyltransferase-like protein 14	METTL14	3.243353
Centromere protein F	CENPF	4.312783
cDNA FLJ50334	CCDC152	3.453733
SECISBP2 protein (Fragment)	=-	12.83316
Isoform 3 of Putative helicase MOV-10	MOV10	8.459016
Homeobox protein Hox-D9	HOXD9	5.329619
POU domain, class 6, transcription factor 2	POU6F2	3.269651

nebulette	NEBL	4.286486
#N/A	NSFP1	4.207594

Appendix D

Bar graph showing the functional annotations, biological processes and molecular function, enriched within the differentially regulated SH-SY5Y dataset. The figure was generated by GeneCodis2.

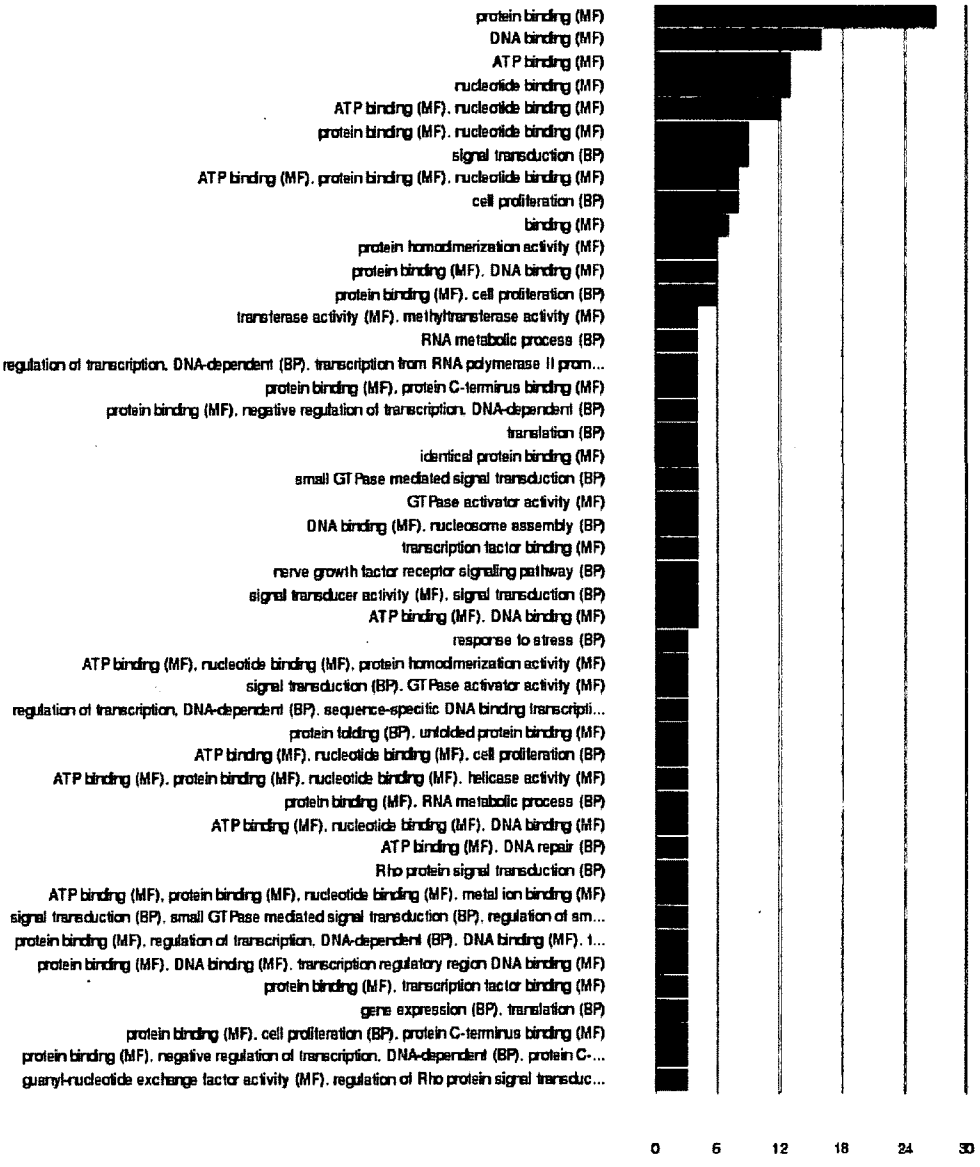


Figure 4-1: Number of genes within the differentially regulated SH-SY5Y datasheet associated with biological proccess (BP) and moleculare function (MF) annotations. Figure generated by GeneCodis2.