A novel mass spectrometry-based proteomic and phosphoproteomic strategy for the quantitative analysis of a leukemic cell line infected by an oncolytic virus

by

Matthew Huebsch

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: 978-0-494-81700-1
Our file Notre référence
ISBN: 978-0-494-81700-1

"To succeed as a theorist, you have to be good. To succeed as an experimentalist, you can go through life kicking over a lot of stones, and if you're lucky, you'll find something."

–John B. Fenn

Abstract

Mass spectrometry has become the tool of choice for large proteomic experiments to investigate the molecular dynamics of biological systems. A strategy for the proteomic and phosphoproteomic quantitative analysis of samples in different biological states has been developed. The strategy incorporates tryptic digestion of cellular lysates, on-column stable-isotope dimethyl labeling, and proteomic and phosphoproteomic investigations by SAX/SCX-LC-MS/MS and IMAC-LC-MS/MS, respectively. The strategy was applied to study altered protein expression and signaling pathways induced by infection of a leukemic cell line, K562, with the oncolytic vesicular stomatitis virus for 30 minutes. A total of 53 proteins and 8 phosphorylation sites were found to be upregulated and 11 proteins and 9 phosphorylation sites were found to be downregulated in VSV-infected K562 cells versus controls. Possible roles of differentially regulated proteins and phosphorylation sites in VSV infection are discussed and the performance of the developed quantitative strategy is evaluated.

Acknowledgments

I'd like to thank my supervisor Jeff Smith for his guidance and patience throughout my time as a graduate student, as well as all other members of the group who have provided me with helpful feedback. I'd like to thank my collaborator Harry Atkins for ideas and resources and Girija Waghray of the Atkins laboratory for culturing and infecting the cells used in this work.

I'd also like to extend my gratitude to Michel Dumontier for writing the script used in this work for the analysis of large datasets; without his insight and dedication, this work would not have been possible.

An additional thanks to all professors and fellow students not mentioned herein who have enriched my experience as a graduate student.

I'd also like to thank friends and family for always providing me with support. Thanks to Tim for always being a great brother and to my parents for their unconditional support and understanding. A specific thanks to Chris Bura for not only being a close friend, but for his willingness to proofread, give technical advice, and provide useful scripts. Finally, I'd like to thank Jinal Patel for encouragement and insight; you have been a good colleague and great friend.

Table of Contents

List of Tables

List of Illustrations

List of Appendices

List of Abbreviations

ABC, ammonium bicarbonate

BCA, bicinchoninic acid

CID, collisionally induced dissociation

ETD, electron transfer dissociation

GO, gene ontology

HDGF, hepatoma-derived growth factor

HPLC, high-performance liquid chromatography

IMAC, immobilized metal ion affinity chromatography

LC, liquid chromatography

m/z, mass to charge ratio

MOI, multiplicity of infection

MS, mass spectrometry

MS/MS, tandem mass spectrometry

RPLC, reversed-phase liquid chromatography

SAX, strong anion exchange

SCX, strong cation exchange

SILAC, stable isotope labeling by amino acids in cell culture

VSV, vesicular stomatitis virus

XIC, extracted ion chromatogram

CHAPTER 1

Introduction

The late 20th century is considered to be the genomic era, highlighted by the complete sequencing of the human genome in 2001 by the Human Genome Project (HGP) at a cost of $3 billion.[1] With this era came a change in attitudes towards how biological systems are understood: no longer are they investigated piece by piece, but rather in large-scale experiments in an attempt to understand how different parts work together as a whole. Microarrays that can profile gene expression have been used to analyze entire organisms, including yeast[2] and human.[3]

The central dogma of molecular biology, illustrated in Figure 1.1, dictates that proteins are the final products of genes. Therefore, when gene expression is measured, it is assumed that one is actually measuring protein expression. The success of identifying genes involved in disease states is a testament to the utility of these studies, however the assumption that gene expression and protein expression are directly correlated is not always correct. For example, proteins can be modified by post-translational modifications or degraded rapidly in response to stimuli. Gene expression studies are also not able to probe how proteins interact with each other; its utility in studying the dynamics of complex organisms is therefore limited.

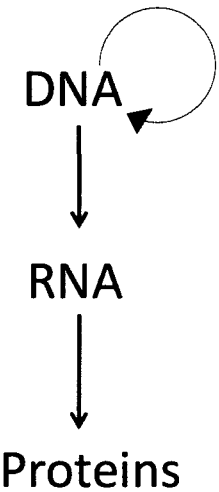


Figure 1.1. Central dogma of molecular biology with arrows indicating flow of information.

For these reasons, the large-scale study of proteins – termed *proteomics* after the word genomics – was a necessary advancement in studying biological systems. This goal was realized with the 2002 Nobel Prize-winning work[4] of John Fenn[5] and Koichi Tanaka[6] for their development of ionization techniques suitable for the analysis of biomolecules, thus allowing proteomic samples to be analyzed by mass spectrometry. Since proteomics and genomics have allowed large-scale, multi-faceted investigations of complex biological systems, this area of study is now referred to as *systems biology*.

The central aspects of mass spectrometry-based proteomics are presented in detail in this chapter, beginning with the fundamentals of protein biochemistry, followed by instrumentation, data gathering and analysis, and separation methods, and finally an introduction to the biochemistry of the oncolytic vesicular stomatitis virus.

1.1 Protein biochemistry

The central dogma of molecular biology, as has been stated, governs the flow of genetic data into functional biomolecules. Briefly, DNA is transcribed into mRNA, which is translated into proteins by ribosomes using triplet codons. Proteins are polymer chains of monomer subunits, termed amino acids, joined together by peptidic amide bonds. There are 20 common proteinogenic amino acids, shown and classified in Figure 1.2. All amino acids contain an amine group, a carboxylic group, and a functional group (typically denoted by –R) bonded to an alpha carbon (C_α). As C_α is bonded to 4 distinct groups, it is therefore isomeric; all amino acids (with the exception of glycine) have two enantiomeric forms, D and L. However, biological systems synthesize and utilize the L-amino acids exclusively. They are often classified by the hydrophobicity of their functional group: there are 8 hydrophobic amino acids and 9 hydrophilic amino acids (3 basic, 2 acidic, and 4 uncharged species). The remaining 3 amino acids are termed special as they have unique properties. Proline is the only amino acid that contains a ring structure around C_α and therefore can form very rigid structures in proteins. Glycine has 2 hydrogen atoms bound to C_α and is therefore non-chiral (i.e. has no D and L isomers). Cysteine is listed as a special amino acid since it contains a primary thiol group that can form disulfide bridges with other cysteine residues in close proximity.

Condensation reactions of the amine and carboxylic acid groups on the individual amino acids form long polypeptide chains, termed the backbone of the protein, and the functional groups dictate the properties of the protein. A

polypeptide will therefore have an end containing a free amine group (termed the N-terminus) and an opposite end containing a free carboxylic acid group (C-terminus). Convention states that polypeptide sequences are listed from the N-terminus to the C-terminus. These termini are basic and acidic, respectively, and are therefore charged under physiological conditions.

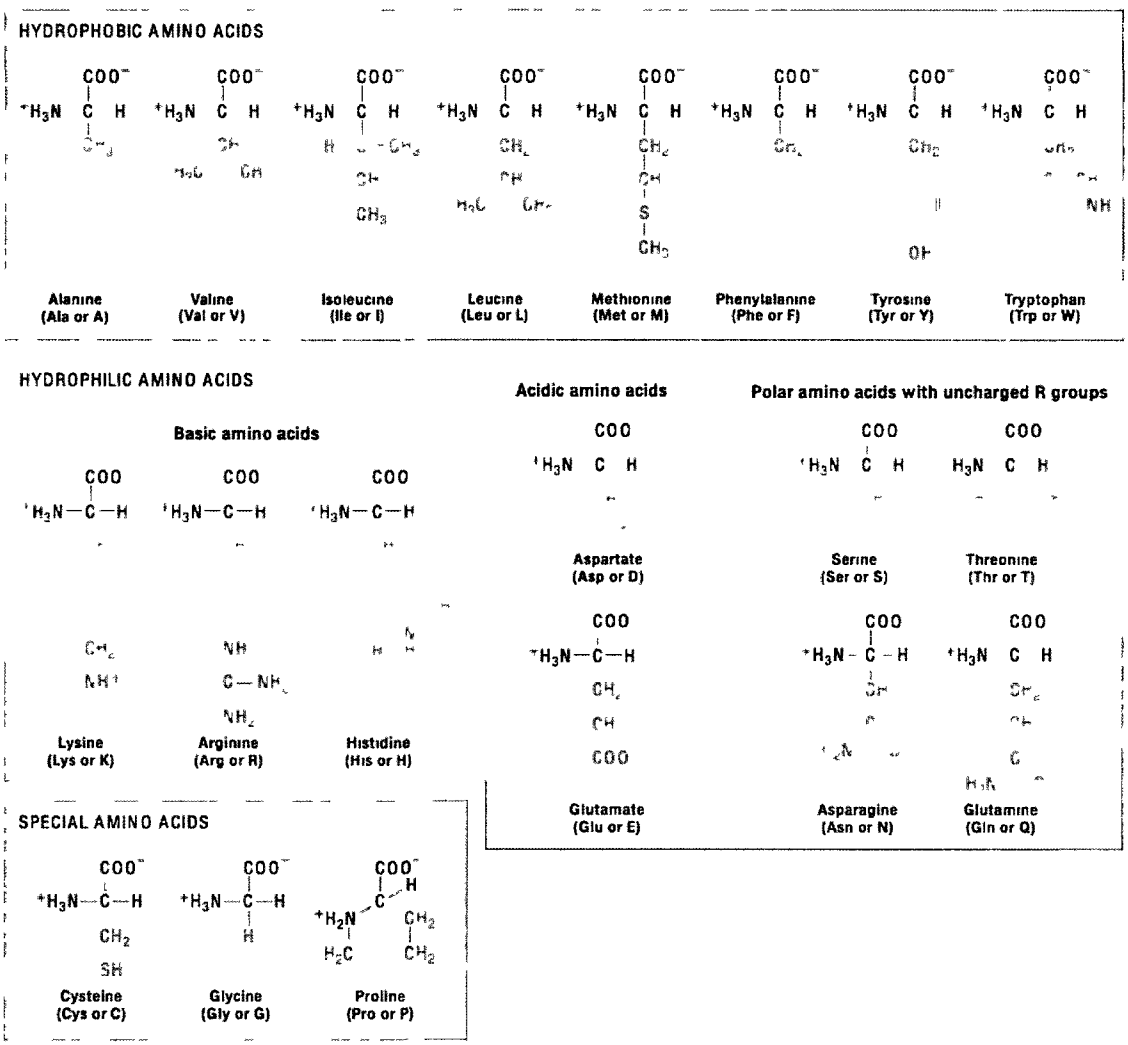


Figure 1.2. The common 20 amino acids found in proteins, classified into hydrophobic amino acids, hydrophilic amino acids, and amino acids with special properties. Charged structures at physiological pH are shown. Reproduced from Lodish *et al.* [7]

The hydrophilic amino acids form hydrophilic, polar regions on a protein: they are typically found on the outside of a protein to maintain solubility in their environment and can undergo non-covalent (e.g. hydrogen bonding) with water-soluble substrates. Conversely, hydrophobic amino acids form hydrophobic regions that are often on the inside of a protein to form the insoluble core. The properties of these amino acids and their locations are responsible for the structure of proteins; these structures are categorized as primary, secondary, tertiary, and quaternary.

These categories of structures are illustrated in Figure 1.3. The primary structure is defined as the linear arrangement of amino acids in a peptide (< ~30 amino acids) or polypeptide (< ~2000 amino acids), termed the sequence. The secondary structure is the local spatial arrangement in a protein: either α-helices, β-sheets, or turns, illustrated in Figure 1.3. An α-helix is a helical structure held together by hydrogen bonds between the amide and carboxylate groups on the peptide backbone and contain ~3.6 residues per turn. As the functional groups of the peptides are outward facing, α-helices may be hydrophobic or hydrophilic structures. A β-sheet is a planar, pleated structure consisting of linear strands of 5-8 amino acid residues. Like α-helices, they are also formed by hydrogen-bonding of the peptide backbone and can be hydrophobic or hydrophilic. Turns are secondary structures that are short (3-4 residues) and whose purpose is to change the direction of the polypeptide backbone. They commonly consist of glycine or proline residues, as these amino acids are able to form rigid, U-shaped turns. There are also weakly defined elements of secondary structure such as loops and random coils.

Tertiary structures define the total, 3-dimensional conformation of a protein. It contains all the secondary structures, along with all other interactions (hydrophobic, hydrogen bonding, ionic interactions, disulfide bridges, etc.), which give the protein its overall shape. While the overall folding of a protein is considered its tertiary structure, it may have a quaternary structure if it is multimeric. For example, hemoglobin is a tetramer[8] with 2 alpha and 2 beta units containing their own tertiary structures; together they co-ordinate with heme-containing iron and form the quaternary structure of hemoglobin.

Protein structure defines their function and thus folding can be considered the most important property of a protein; failure to correctly fold can result in fatal diseases in humans by the accumulation and aggregation of misfolded proteins.[9] In addition to the factors listed above, protein structure is dependent on pH, temperature, and hydration state. The physical state of biological systems thus has to be tightly regulated with a buffered pH and set temperature (termed physiological conditions). Conformational changes can be induced by changes in pH or binding of a substrate onto an active site of a protein. A protein exposed to heat or a strong denaturing agent such as urea will *denature* proteins; this process may be irreversible but has been shown to often be reversible with the protein returning to its Gibbs free energy minimum with a return to physiological conditions. [10]

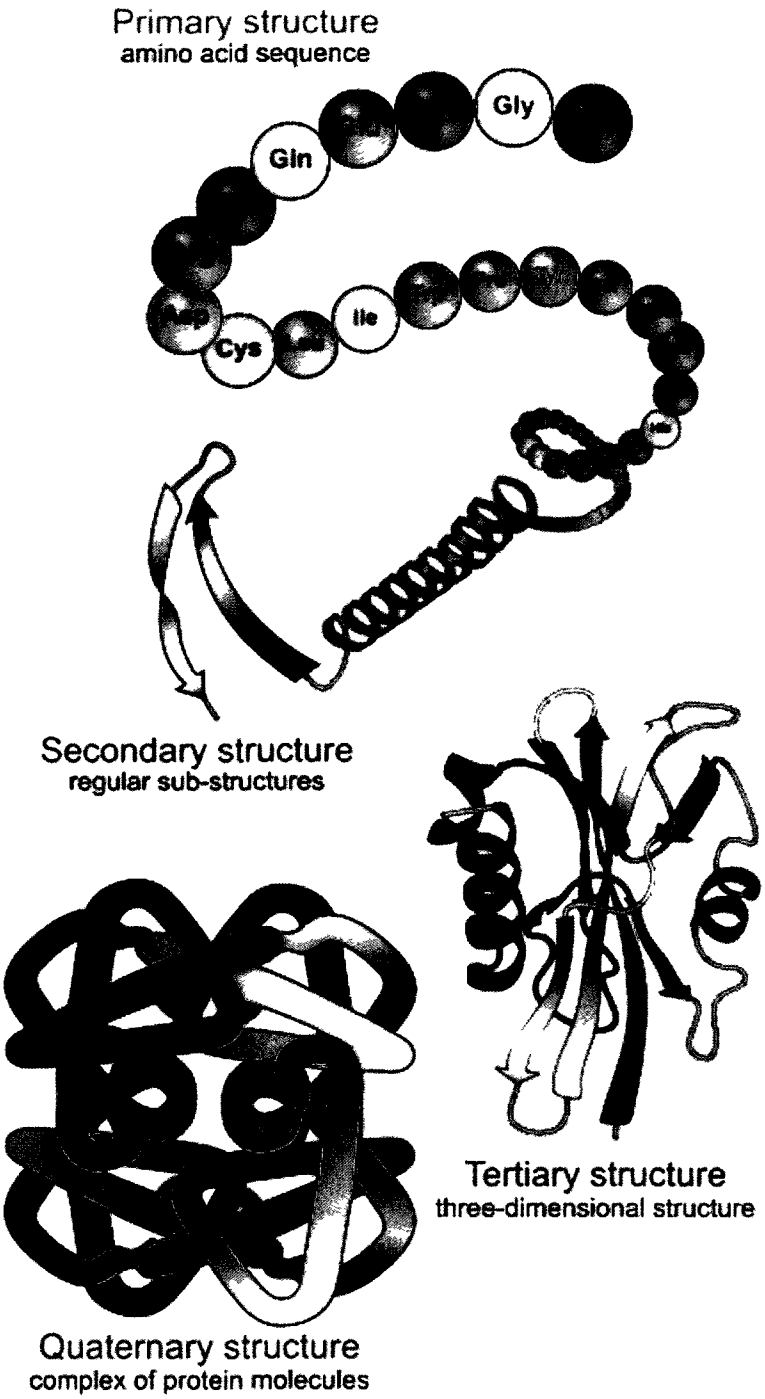


Figure 1.3. The four structural levels of proteins. Examples shown are of a random primary sequence, common secondary structures (alpha helix and beta sheet), a representative tertiary structure from PI3 showing multiple secondary structure elements, and the quaternary structure of hemoglobin. Reproduced from Wikimedia Commons.[11]

Proteins have motif-containing regions called domains, which serve both functional and structural uses and are the fundamental units of nearly all large proteins as they dictate the tertiary structure.[12] Functional domains have a specific purpose; relevant examples include the kinase catalytic domain, the death effector domain, and the catalytic domain of the serine protease trypsin. The kinase catalytic domain is responsible for the post-translational modification of proteins by addition of a phosphate group, discussed in detail in the next section (*vide infra*). The death effector domain regulates many signaling pathways, however it is primarily responsible for activating caspases to induce programmed cell death (termed *apoptosis*).[13] The motifs in these domains often have consensus sequences and motif analysis can reveal information with regards to which enzyme or family of enzymes recognizes that motif. This is particularly useful in phosphoproteomics as the diversity of kinases is very small relative to the number of known phosphorylation sites in humans (*vide infra*). Therefore motif analysis can classify phosphorylated residues into groups that are likely recognized by the same kinase or kinase family.[14]

A particularly important class of proteins in the field proteomics is that of the proteases, which are enzymatic proteins that can hydrolyze the covalent peptide bond. Proteases can break these bonds either specifically or non-specifically and are therefore grouped into specific and non-specific proteases. Trypsin is a specific protease as it cleaves in a defined matter: C-terminal to arginine and lysine residues unless they are followed immediately by a proline residue. Conversely, non-specific proteases hydrolyze amide bonds at non-specific locations in a typically poorly

defined manner. Proteases with high specificity are important for the formation of mass spectrometry amenable peptides in bottom-up proteomics (*vide infra*).

1.1.1 Post-translational modifications

The human genome contains between 20,000-25,000 genes,[15] far fewer than originally expected and approximately the same number as *C. elegans*. The higher complexity in *H. sapiens* is explained by two factors: one, the high use of alternative splicing, in which the same genes can be spliced differently at the RNA level, and two, the larger number of post-translational modifications (PTMs) on proteins. There is a large diversity among PTMs: they can be reversible or permanent, and can be as simple as an oxidation of a residue (e.g. oxidation of sulfur on methionine) to the covalent addition of large molecules onto residues (e.g. SUMOylation, ubiquitination). There are currently over 400 PTMs listed in the UniProt list of post-translational modifications.[16] Common modifications and their functions are listed in Table 1.1; of these, phosphorylation will be discussed in more detail.

Table 1.1. Common protein post-translational modifications. Adapted from Mann *et al.*[17]

PTM	Δm (Da)	Function and notes
Phosphorylation (pS, pT, pY)	+80	Reversible. Used in signaling pathways, activation/inactivation of enzymes, structural changes.
Acetylation	+42	Regulation of gene expression by histone modification, N-terminal modification.
Methylation	+14	Regulation of gene expression.
Glycosylation (N- or O-linked)	+203, >800	Reversible. Structural role, signaling, regulatory functions.
Ubiquitination	>1,000	Signal tag for proteasomal degradation.
Disulfide bridge	−2	Protein structure.
Sulfation (sY)	+80	Regulation of protein-protein and receptor-ligand interactions

Phosphorylation is a ubiquitous, reversible modification in which a phosphate group is added to the hydroxyl group of a serine, threonine, or tyrosine residue. The phosphates are added by an ATP-driven mechanism by a family of enzymes called kinases, and removed by phosphatases, as shown in Figure 1.4. Given that there are 518 known kinases, 65 known phosphatases,[18] and an estimated 100,000 phosphorylation sites in the human proteome,[19] many phosphorylation substrates share homology. It has also been estimated that 30% of proteins are phosphorylated at any given time and 50% of proteins are phosphorylated during their lifetime[19] in a ratio of approximately 1800:200:1 for phosphoserine:phosphothreonine: phosphotyrosine.[20] Phosphorylation sites modify the structure and function of proteins, thus regulating signaling pathways, cell division, and metabolic maintenance.[21] Aberrant phosphorylation has been shown to be involved in disease states[22] and has led to the development of protein kinases

as therapeutic targets.[23] The development, approval, and success of the tyrosine kinase inhibitor drug imatinib to treat chronic myelogenous leukemia clearly demonstrates how protein phosphorylation can affect human health. The identification of phosphorylation sites by mass spectrometry is later discussed in detail (*vide infra*).

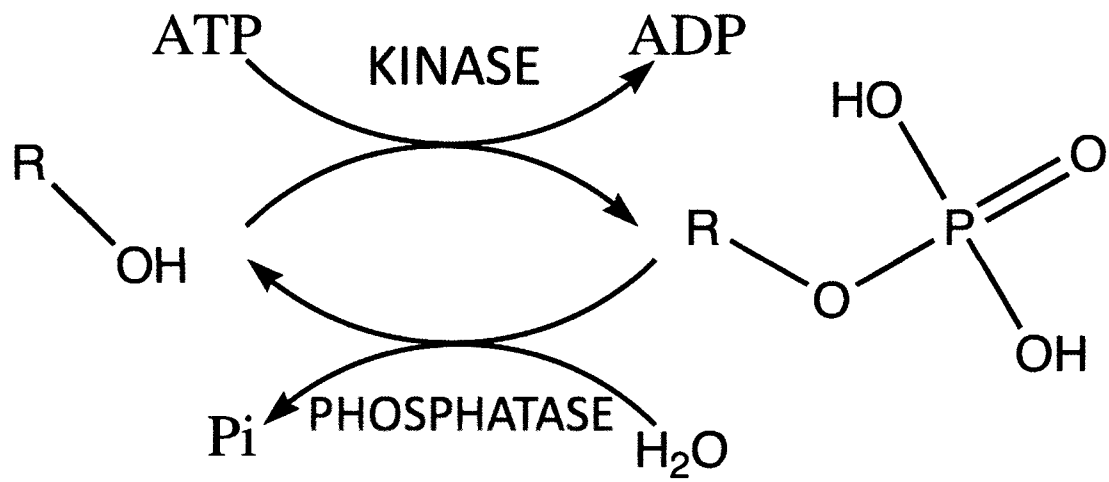


Figure 1.4. Addition and removal of phosphate groups by kinases and phosphatases, respectively. R designates a serine, threonine, or tyrosine residue.

1.2 Mass Spectrometry

All mass spectrometers are comprised of three components: an ionization source, mass analyzer(s), and a detector. These components will be discussed in detail in the following sections, but a brief description of the components is helpful.

First, all mass spectrometers operate by analyzing ions in the gas-phase; therefore, an ionization source is necessary. Many ionization techniques have been developed for all classes of molecules, such as electron impact (EI) for organic molecules, atmospheric-pressure chemical ionization (APCI) for medium-polarity molecules, and matrix-assisted laser desorption/ionization (MALDI) and electrospray ionization (ESI) for biomolecules. There is diversity among the types of ions these techniques form; their only common property is that they all ionize analytes for gas-phase analysis.

Mass analyzers are necessary to determine the mass-to-charge ratio (m/z) of the ions of interest. They can separate ions in space (ion trap), in time (time-of-flight) or as a mass filter (quadrupole). They are generally classified into their ability to resolve ions, defined by

$$R_{FWHM} = \frac{m}{\Delta m} \tag{1.1}$$

where R is the resolution, m is the mass of the ion and Δm is the peak width at half maximum (FWHM). A higher resolution therefore allows peaks closely spaced to be resolved and gives narrower peaks.

A detector is needed to convert ions into an electrical signal that can be used to make a mass spectrum. A detector should be sensitive, able to resolve to at least the same degree as the mass analyzer, have a very rapid response time, be stable over time, have a large dynamic range, and have low noise.[24] Detectors are classified

into focal-point and focal-plane detectors; an example of the latter, the microchannel plate detector, will be discussed.

By scanning a range of masses and gathering their intensities, a *mass spectrum* can be created, shown in Figure 1.5. This is the basis for mass spectrometry and will be used extensively in this work. Note that Da or amu are often listed as the x-axis units in mass spectra; this is not technically correct as these are units of mass and not m/z. The Thomson (Th) is a unit referring to m/z; however, it has not been universally adopted.

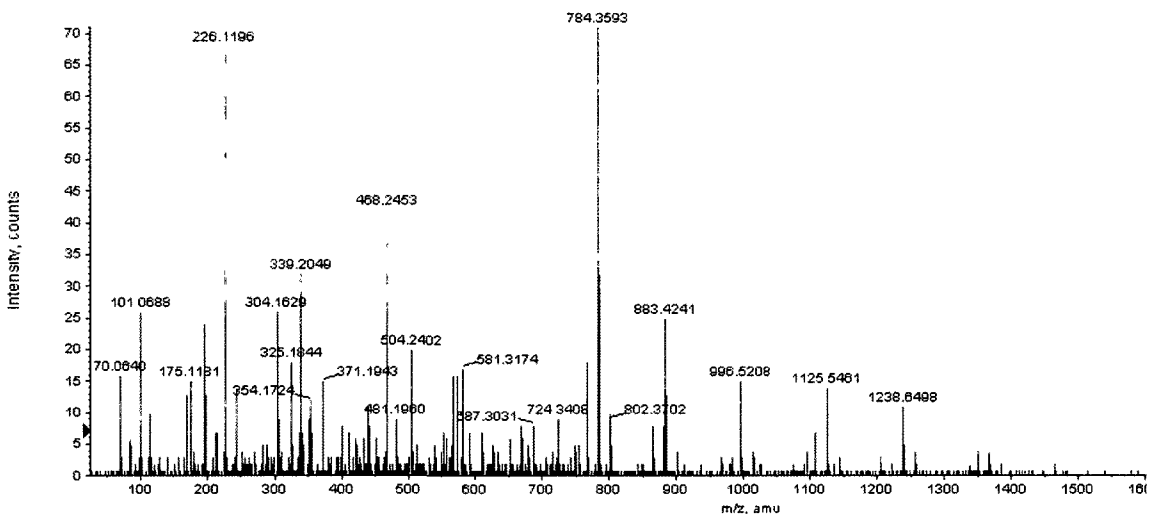


Figure 1.5. A mass spectrum showing intensity of ions (count) versus mass to charge ratio (Th).

1.2.1 Electrospray ionization

Electrospray ionization (ESI) is an ionization method invented by Malcolm Dole[25] and adapted for biomolecules by John Fenn.[5] The principle is illustrated in Figure 1.6. A capillary has a large positive voltage of +2-5kV applied (note: all voltages will be stated as they apply to positive ion mode) and, since the capillary is very thin, the capillary will experience a very large electric field as explained by the equation

$$E_c = \frac{2V_c}{r_c \ln\left(\frac{4d}{r_c}\right)} \qquad (1.2)$$

where V_c is the voltage applied to the capillary, r_c the radius of the capillary, d the distance between the capillary and counterelectrode, and E_c the resultant electric field experienced by the capillary. As a polar liquid containing the analyte is flowed through the capillary, positive ions will flow towards the highest field at the capillary tip and negative ions away from the tip. The positive ions are ejected from the tip due to the downfield force and form a Taylor cone.[26] The cone is composed of a jet followed by a plume formed due to the electrostatic repulsions of the positive ions. As the droplets travel towards the counterelectrode, solvent evaporation occurs until they become bare analyte ions. There are competing mechanisms by which this is proposed to happen: the charged residue model (CRM) and the ion evaporation model (IEM). The CRM explains the desolvation as the evaporation of solvent from the droplets until the Rayleigh limit is reached, when they then undergo Coulombic fission into smaller droplets.[25] Evaporation and fission repeats

until the bare ion is reached – that is, until the droplets contain on average one analyte molecule. There is evidence that this is the primary mechanism in which large molecules like proteins become gas-phase ions in electrospray ionization.[27] The competing IEM model states that ions are ejected directly from small droplets of less than 10 nm and experimental evidence has been shown to support this theory.[28] This model, however, only works for small molecules (e.g. salts) and is not believed to be a mechanism associated with the ionization of biomolecules.[29]

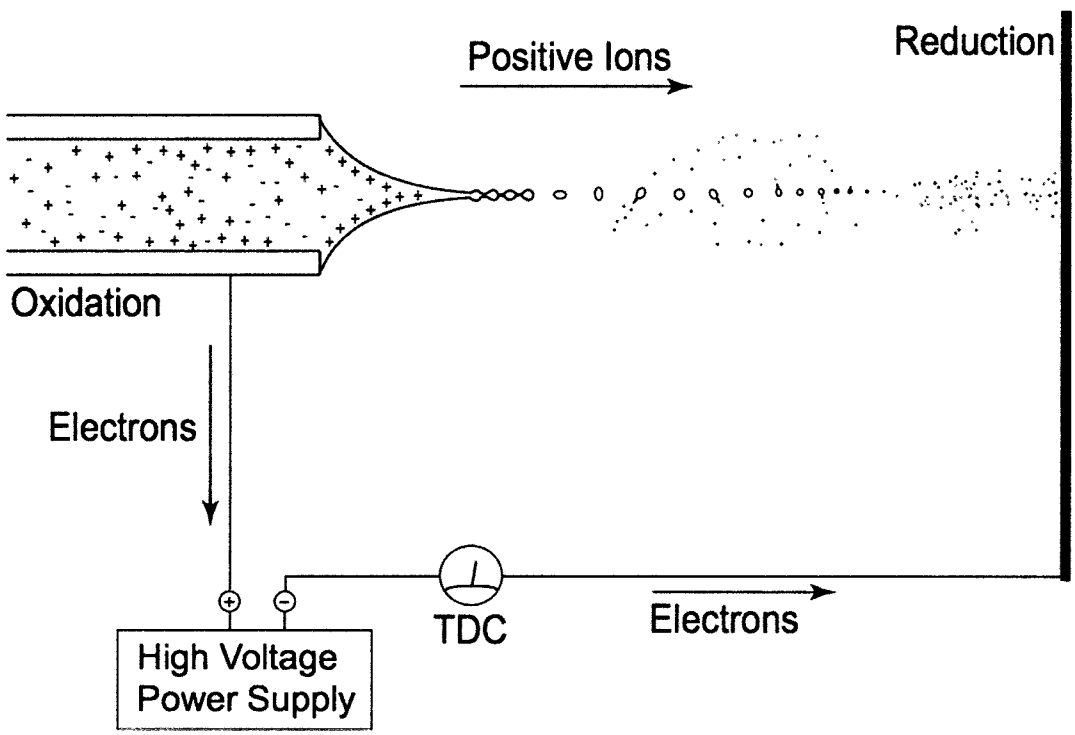


Figure 1.6. Electrospray ionization. Reproduced from Kebarle *et al.*[29] Note the reduction electrode would have an orifice when inline with a mass spectrometer.

A variation of electrospray is nanospray: electrospray with a flow rate on the order of nanolitres per minute. Concurrently developed by Smith[30] and Mann,[31] the use of nanospray has become standard for biological mass spectrometry. Nanospray tips are typically pulled fused silica with a capillary outer diameter of 5-8μm and an orifice size of ~1μm. Coupled with a nanoHPLC, these tips deliver electrospray rates of 20-40 nL/min, giving droplet sizes two to three orders of magnitude smaller than conventional electrospray.[32] The tips are available either Au/Pd coated for direct infusion analysis at ionspray voltages of 0.7-1 kV (Figure 1.7a, left) or uncoated fused silica for operation online with nanoHPLC at electrospray voltages of 4-5 kV (Figure 1.7a, right).

While these systems are more expensive, electrospray ionization mass spectrometers function as a concentration-sensitive detector and, since only a small number of droplets are sampled at high flow rates, sensitivity is often poor. Consequently, compared to traditional electrospray, nanospray has increased sensitivity,[33] as well as increased ionization efficiency,[34] increased transmission efficiency,[31] and decreased ionization suppression. It has been estimated that the increase in ion transmission rates (up to a maximum of ~10^8 charges/s) may exceed the capacity of ion trap instruments by >100 fold; however, this is not a problem for tandem quadrupole-time-of-flight instruments.[35] Band broadening can also be substantially reduced in nanospray as the tips can be purchased pre-fritted and packed in-house with chromatographic resin, essentially integrating the analytical column with the nanospray emitter, thus eliminating dead volume.

The setup shown in Figure 1.7b is used for mass calibration by direct infusion. Several microliters of a calibrant such as CsI (concentration ~20 μM) is put into the coated nanospray capillary. Pressure is generally added to ensure liquid is present at the tip of the capillary. These tips may come without an orifice; in this case, the tip is gently bent against the interface plate so liquid may pass. A voltage is applied to give a spray on the order of nL/minute and the corresponding mass spectrum is used to calibrate the instrument.

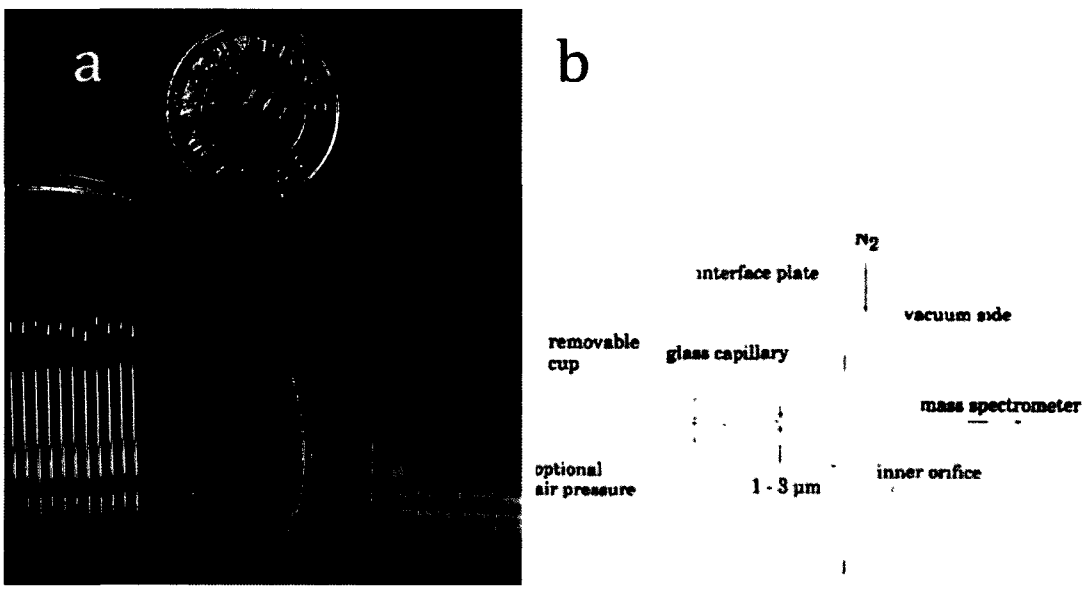


Figure 1.7. (a) Commercially available nanospray tips for direct sample analysis (left) and nanoHPLC systems (right). (b) Nanospray ion source setup for direct infusion.[31]

1.2.2 Quadrupole mass analyzers

A quadrupole is a mass filter that theoretically consists of four parallel, electrically-connected hyperbolic rods, although due to engineering costs the rods

are usually circular. As shown in Figure 1.8, opposite rods are given a superimposed direct current (DC) potential (defined as U) and radio-frequency (RF) potential (defined as $V \cos \omega t$). The other pair of rods is given the opposite potential (-U) and a RF field that is 180° out of phase ($-V \cos \omega t$). The effect of applying these voltages is an oscillating field given by

$$\Phi_{(x,y,t)} = (U - V \cos \omega t)\left(\frac{x^2 - y^2}{r_0^2}\right) \tag{1.3}$$

where r_0 is the inscribed radius inside the quadrupoles, x and y are the distances from the center of the field in the *x*- and *y*-planes, respectively. Ions travel through the quadrupole in the z-direction and, depending on their m/z, either have a stable or unstable trajectory depending on the operating parameters.

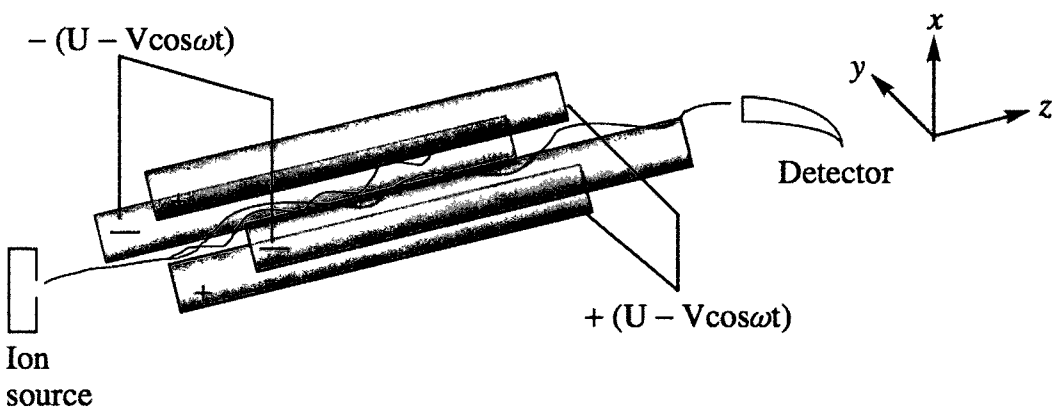


Figure 1.8. Applied voltages to quadrupole rods. Reproduced from Dass *et al.*[24]

The quadrupole acts as a narrow bandpass mass filter rather than a traditional mass analyzer as the ions are not separated in time (e.g. time of flight MS) or in space (e.g. magnetic sector MS). The positive rods (i.e. the *xz* plane) act as a high-pass filter: positive ions will converge towards the central *z*-axis during the positive half-cycle of the RF field, but lighter ions will be rapidly accelerated and exit the field entirely during the negative half-cycle. Conversely, the negative rods (*yz* plane) act as the low-pass filter: ions will diverge during the negative half-cycle, but only light ions will again converge during the positive half-cycle due to the high momentum of the heavier ions. The total effect of this filtering is that of a narrow bandpass filter, illustrated in Figure 1.9.

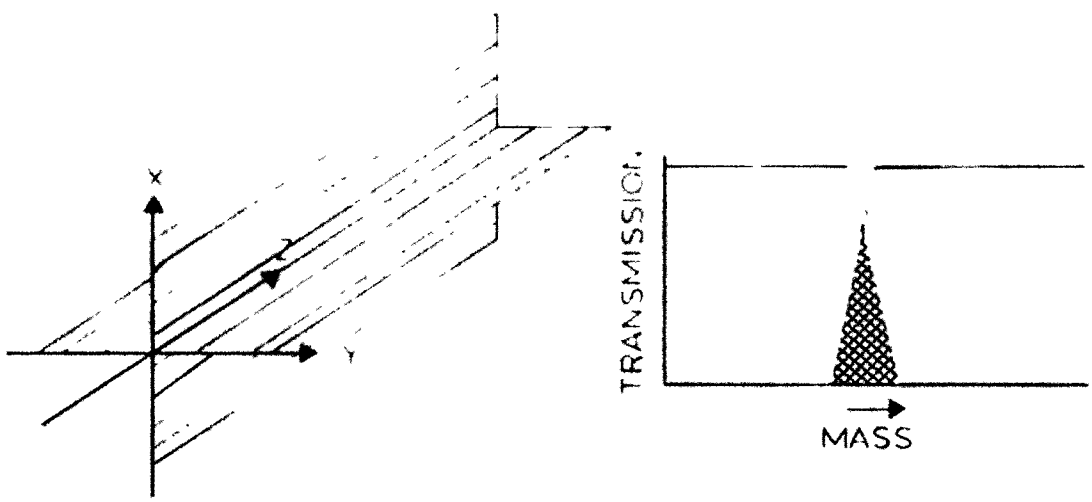


Figure 1.9. Low-bandpass filtering in the yz plane and high-bandpass filtering in the xz plane. Reproduced from Miller *et al.*[36]

The ion motion in the *x*- and *y*-directions is described by the canonical form of the 2nd-order differential Mathieu equation, first described by Émile Mathieu to model the movement of vibrating elliptical membranes[37]:

$$\frac{d^2u}{d\xi^2} + (a - 2q\cos 2\xi)u = 0 \qquad (1.4)$$

where u is the transverse displacement from the center of the field in the *x*- and *y*-directions, and ξ is the time of the applied field ($\omega t/2$) in radians. The solutions to this equation are given by

$$a = a_x = -a_y = \frac{8zeU}{m\omega^2 r_0^2} \qquad (1.5)$$

and

$$q = q_x = -q_y = \frac{4zeV}{m\omega^2 r_0^2} \qquad (1.6)$$

where a and q are dimensionless quantities. The solutions to this equation can either be bounded or unbounded; a bounded solution is stable in that the *x* and *y* values are finite and thus physically the ion will remain in the quadrupole along the length of the quadrupoles. As the values of a and q have reduced six dimensions (U, V, m, e ,ω, r_0) into 2 dimensions, a stability diagram can be easily plotted, as shown in Figure 1.10. There are 4 possible stability regions; only the first is shown.

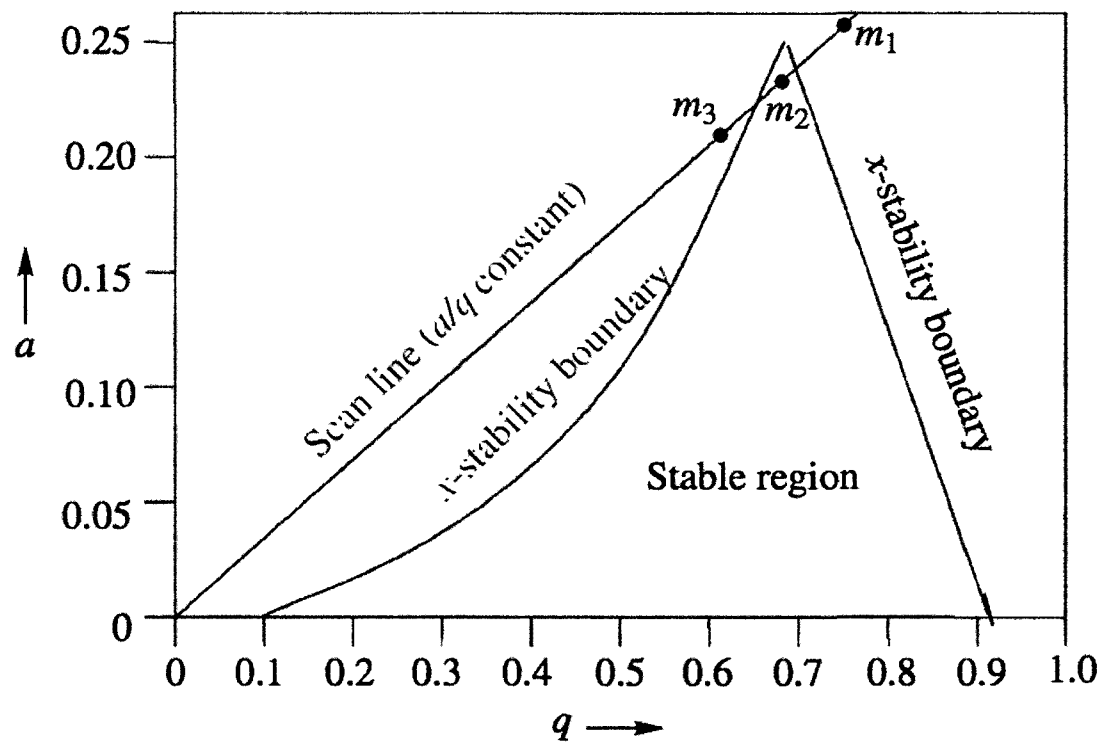


Figure 1.10. Quadrupole first stability diagram. Reproduced from Dass *et al.*[24]

In practice, the values of a and q are chosen such that the stability region only intersects the tip of the scan line. The a/q ratio is then held constant and U and V are simultaneously increased so as to change the mass that the quadrupole allows to be transmitted. By sweeping the voltages, a quadrupole can sequentially collect a series of spectra of increasing m/z values to create a mass spectrum, up to a practical limit of approximately 4,000 Th. For modern quadrupole instruments, the values for a and q are typically 0.237 and 0.706, respectively.[38]

Quadrupoles can also operate in RF-only mode, where no DC potential is applied. According to equation 1.5, $a = 0$ because $U = 0$. Thus, a large portion of the

mass scan line will fall within the stable region; this is physically interpreted as allowing ions of all m/z through the quadrupole. Generally, a cutoff value of q is used, rejecting all masses below this.[38] This type of quadrupole is of importance for hybrid quadrupole time-of-flights instruments as the q0 and q2 quadrupoles operate in this mode, as can the Q1 quadrupole (*vide infra*).

The quadrupole mass analyzer is cheap, mechanically simple, and has high scan speeds and transmission efficiencies; however, it suffers from poor mass resolution (~4,000).[24]

1.2.3 Time-of-flight

Time-of-flight (TOF) mass analyzers are based on a simple concept that heavy ions will travel slower than lighter ions when accelerated by an electric field and allowed to travel down a field-free drift tube, as shown in Figure 1.11. The kinetic energy given to an accelerated ion is defined by

$$KE = qV = \frac{mv^2}{2} \tag{1.7}$$

and the velocities of the ions are then given by

$$v = \sqrt{\frac{2qV}{m}} \tag{1.8}$$

Since the ions have an initial velocity prior to entering the drift region, their acceleration after will be given by

$$KE = qV + \frac{1}{2}mv_0^2 \qquad (1.9)$$

The second term in equation 1.9 limits the resolution of the instrument as it widens the time in which ions of the same m/z reach the detector. Spatial distribution also decreases resolution since ions closer to the flight tube will experience the acceleration field for a shorter period of time and thus will have a slower velocity than those that remain in the field for a longer period of time. To reduce the effects of these phenomena, reflectron and/or orthogonal acceleration can be used, or the drift tube length can be increased (*vide infra*).

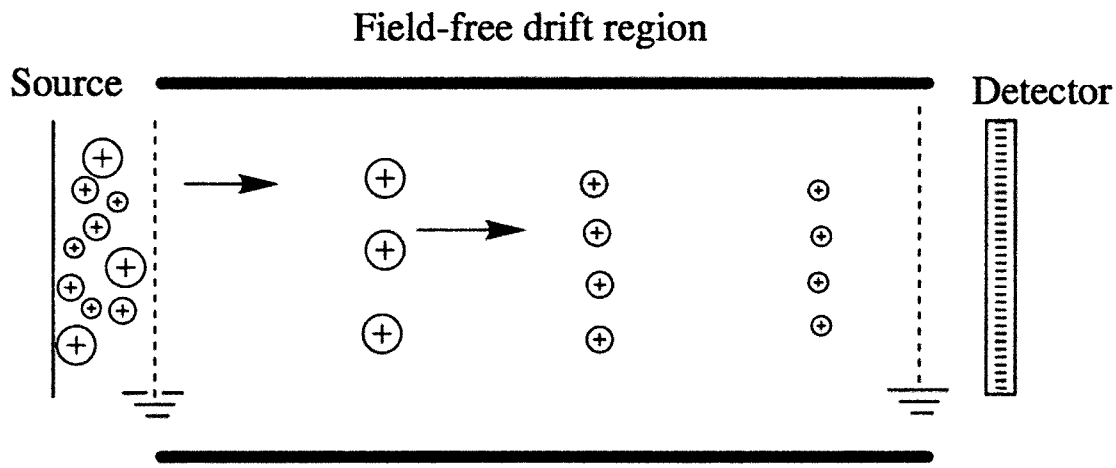


Figure 1.11. General schematic of time-of-flight mass spectrometry. Reproduced from from Dass *et al.*[24]

Any collisions between analyte ions and atmospheric molecules will transfer energy away from the analyte ion and affect the time it spends in the flight-tube. For this reason, a TOF has to operate under vacuum generated by a turbomolecular pump, typically on the order of 10^{-7} torr. At this pressure, the mean free path is very long, given by

$$\lambda = \frac{1}{N\sigma} \tag{1.10}$$

where λ is the mean free path, N is the gas number density, and σ is the collisional cross-section between an ion and an atmospheric molecule. For a pressure of 5×10^{-7} torr and assuming a collisional cross-section of 50 Å^2, equation 1.10 gives a mean free path of ~100 m. That is, an ion at this pressure would be able to travel an average of 100 m before striking an atmospheric gas molecule. Since the length of a TOF is 1-2m in length, it is therefore unlikely that ions will experience collisions with any atmospheric molecules.

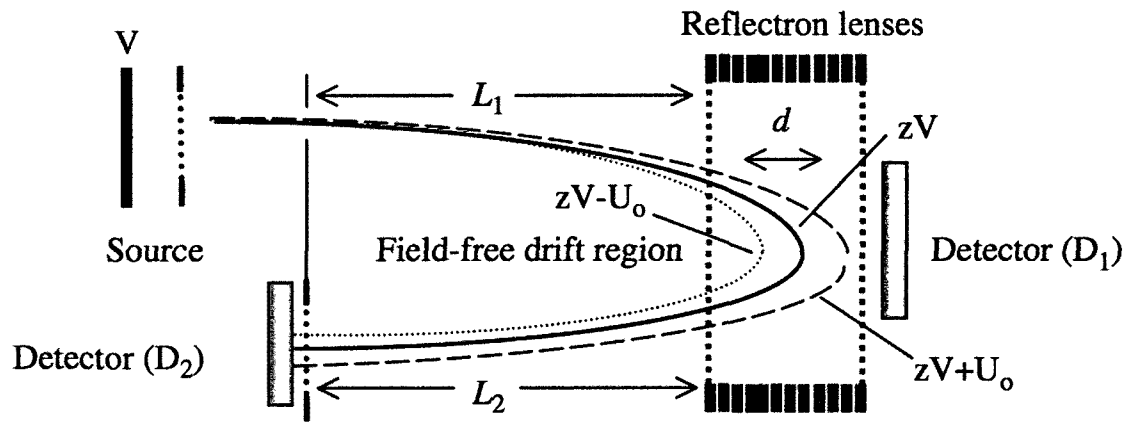


Figure 1.12. Reflectron time-of-flight mass spectrometer. Reproduced from Dass *et al.*[39]

A reflectron TOF improves resolution over a traditional TOF through multiple mechanisms. It effectively doubles the drift tube length, which increases resolution by increasing the time difference that similar ions would reach the detector. The reflectron also corrects small differences in velocity of ions of equal m/z. For instance, shown in Figure 1.12, ions with an excess of energy $+U_0$ will penetrate a further distance, *d*, into the reflectron's field; conversely, ions lacking energy will penetrate a smaller distance. This will correct the ions' variation in velocity and they should travel in unison through L_2 and arrive simultaneously at the detector.

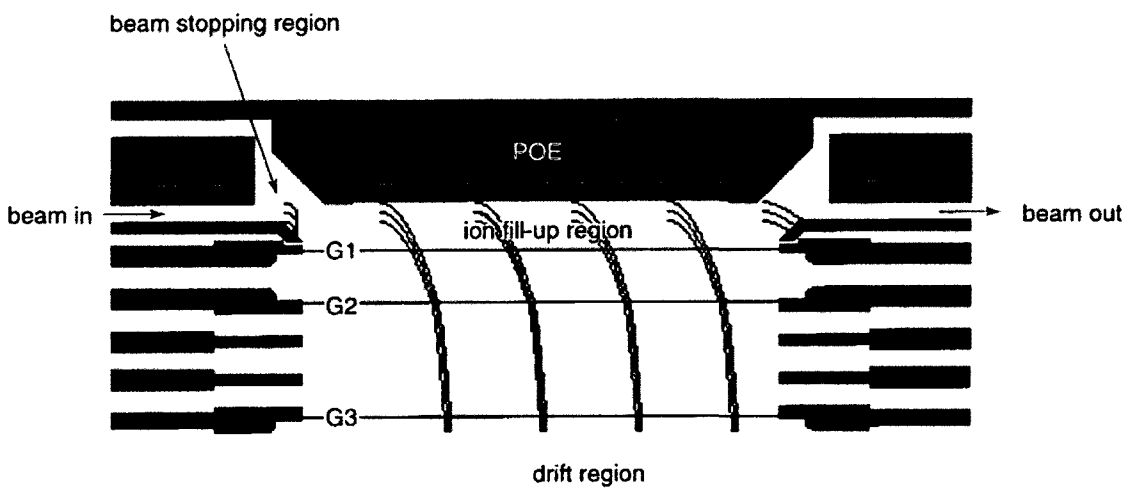


Figure 1.13. Acceleration region of an orthogonal acceleration time-of-flight mass spectrometer. Reproduced from Dawson *et al.*[40]

Orthogonal acceleration, first proposed in 1989[40] and implemented in 1994 for biomolecules,[41] is a technique for introducing continuous ionization sources like

ESI into a TOF-MS. After passing through the final quadrupole (q2), the ion beam is directed into a fill-up region where the ions are very briefly stored. An electric potential is then applied very rapidly (charge up time of order ~100 ns, pulse time ~10 μs) by a simultaneous positive pulse on a push-out electrode (POE) behind the fill-up region and negative pulse on the first grid (G1) in Figure 1.13 (~+600 V and ~−600 V, respectively).[42] These potentials in the acceleration region give the ions a velocity orthogonal to their initial path from q2. The final grid in the acceleration region (G3) is set as the voltage of the TOF; after this grid, the field is constant and thus the ions feel no potential and travel at a constant velocity. After the ions leave the acceleration region, the POE and G1 are set back to a zero potential and the next set of ions are stored into the fill-up region. The acceleration region typically operates at a frequency of 1-2 kHz, however this varies according to the mass range being analyzed. When a higher mass range is being scanned, the frequency must be reduced so as light ions do not overlap with the arrival times of heavier ions from a previous pulse. The advantage to the orthogonal acceleration configuration is that, since the acceleration is orthogonal to the ions' original motion, the spatial and energy variations of the ions will be reduced, resulting in higher resolution.[43]

TOF instruments have a high sensitivity (femtomolar, *vide infra*), high resolution (15,000), and high mass accuracy of <50 ppm with external calibration and <5 ppm with internal calibration (*vide infra*).[24] Their pulsed nature also makes them relatively fast for most applications. Furthermore, they have no theoretical upper mass limit, making them suitable for detecting both peptides and entire proteins.

1.2.4 Ion detectors

There are many detectors that can be used in mass spectrometry, however, the class of detectors known as focal plane detectors are more desirable than their focal point detector counterparts due to sensitivity. The most ubiquitous of the focal plane detectors is the microchannel plate (MCP) detector. First developed reliably in 1962, these detectors have very high sensitivity towards positively-charged ions, low dark current and noise, and ultra-high time resolution.[44] A MCP detector is an array of electron multiplier channels arranged in parallel with a diameter of $6 - 10$ μm and spaced ~15 μm apart. A metallic, conductive coating is added to the front and back of the MCP, which serve as electrodes. When an ion strikes the front plate, it creates a cascade of electrons down one of the channels, which eventually is detected as an ion current. The gain is related to the applied voltage on a single MCP by the equation

$$G = \left(\frac{KV_0^2}{4V\alpha^2}\right)^{4V\alpha^2 V_0^{-1}} \tag{1.11}$$

where G is the gain, V_0 is the applied voltage, V is the initial energy of the secondary electron, α is the ratio of length to diameter of the microchannels, and K is from the relationship $\delta = KV_c$, where δ is the secondary emission coefficient and V_c is the collision energy. It follows from this relationship that as the applied voltage increases, as does the gain, until saturation is reached and the gain levels off in response to higher applied voltages.[45]

Typically in modern mass spectrometers there are 2 MCPs with a gap separating them: this is known as a chevron configuration, which can increase the gain of the system by >2 orders of magnitude.[45] As electrons exiting a microchannel in the first MCP cross the gap, they spread radially and cause electron cascades in multiple microchannels in the second MCP, substantially increasing the gain. Although this results in a loss of spatial definition in the *xy* direction, such resolution is generally not important in TOF instruments and the increase in sensitivity is substantial.

1.2.5 Hybrid quadrupole-time-of-flight mass spectrometer

The schematic of the hybrid QqTOF mass spectrometer with an electrospray source is shown in Figure 1.14. After being ionized by the ESI source, ions pass through the front end of the mass spectrometer into sequentially lower regions of pressure: through the orifice, followed by the skimmer, and into q0, a sequence in which the pressure drops from atmospheric pressure (~760 torr) to 10 mTorr. q0, the first quadrupole, is an RF-only quadrupole used to collisionally cool and focus the ions: a phenomenon used to increase ion transmission and remove axial kinetic energy from the ions.[46] The re-accelerated ions, with near identical axial kinetic energies, then pass through Q1 operating in either r.f.-only mode for survey MS scans or as a mass filter for product ion scans. Ions are then accelerated towards q2, which is filled with an inert gas such as argon, at either low energies (~10 eV) for collisional cooling or high energies (20-200 eV) for collisionally-induced

dissociation (CID). The fragments and/or the parent ions are again collisionally cooled prior to entry into the pulsing region of the TOF to ensure a uniform energy distribution and focused ion beam.

The ions are separated in the TOF mass analyzer as detailed previously, with a few clarifications. Quadrupoles are generally operated near ground, requiring that the TOF not operate from a high positive voltage to ground, but instead operates at ground to a high negative voltage at the detector. This configuration requires a "liner" shown in Figure 1.14 to shield the electric field inside the TOF from any external fields.[38]

The ions are detected by a MCP with a 4-anode detector rather than the traditional single-anode detector described previously. The 4-anode detector has been found to increase detection efficiency by ~2.5 times.[47] The ion current from the MCP detector over time must be converted to a mass spectrum; in the TOF instruments this can be done via a time-to-digital converter (TDC) or transient recorder (TR). Transient recorders convert ion current directly to a digital signal; however, this conversion method leads to high noise levels.[38] TDCs have become the standard for QqTOF instruments due to their low noise but are limited by their dynamic range. Time-to-digital converters record the time at which a packet of ions is accelerated down the flight-tube and records the times at which ions arrive at the anode. These arrival times are on the order of nanoseconds and are placed into discrete time bins. As the pulses occur at a frequency of several kHz, the arrival

times can be summed and converted into a mass spectrum of ion counts versus mass-to-charge ratio.

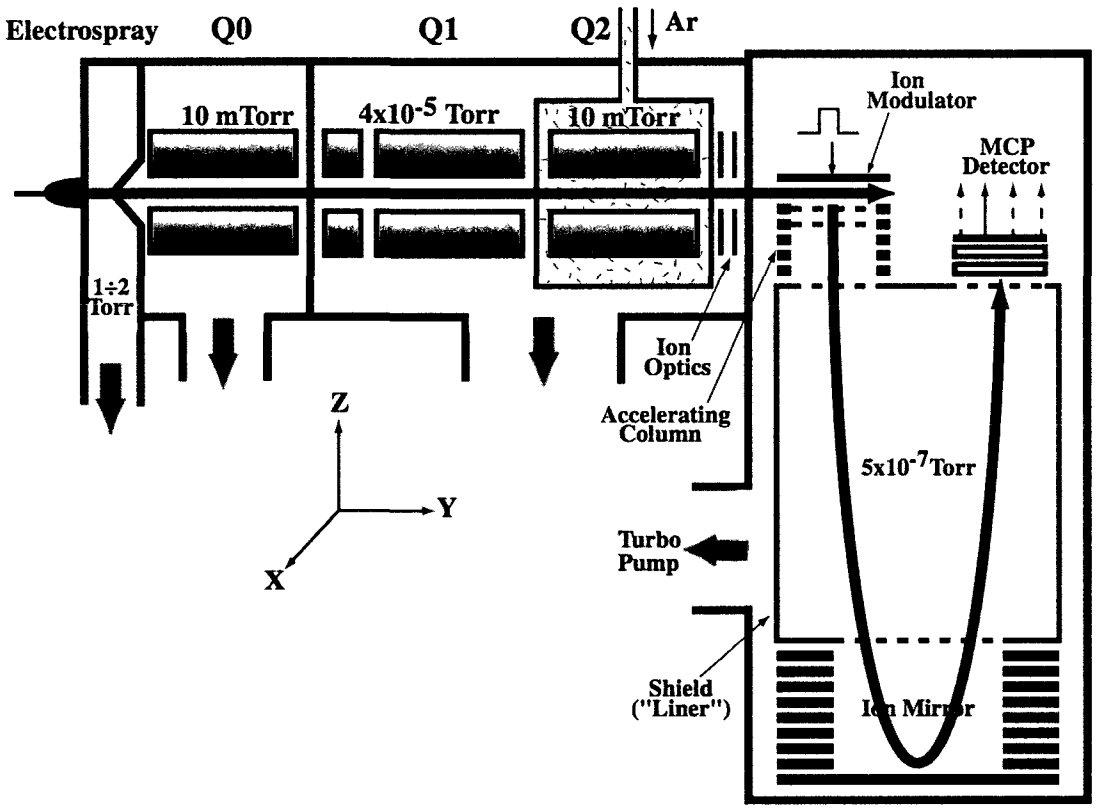


Figure 1.14. General schematic of QqTOF mass spectrometer. Reproduced from Chernushevich *et al.*[38]

1.3 MS-based proteomics

Mass spectrometry-based proteomics falls into two distinct methods: bottom-up and top-down proteomics. Top-down proteomics refers to the analysis of proteomes containing intact proteins, and directly performing MS/MS on entire proteins. While this gives the molecular weight of the protein and a theoretical 100% sequence coverage, it is not the preferred analysis method due to several

factors. First, it is nearly impossible to fragment an entire protein using CID, as the energy is not easily transferred entirely along the protein backbone since (a) a small amount of energy is transferred to large gas-phase protein ions hitting small argon atoms presents a center-of-mass problem and (b) the large number of degrees of freedom in the protein results in poor energy transfer. These phenomena lead to incomplete fragmentation and the large complexity of the product ion spectrum results in difficult peak assignment. ESI of large proteins results in high charge states that are difficult to assign with all but the highest resolution mass spectrometers. Post-translational modifications also present a difficulty for top-down proteomics as a protein can have many PTMs, creating a large number of identical proteins which differ only in their PTM state, and thus their mass.

The bottom-up proteomic approach involves converting proteins into smaller peptides by digestion with a proteolytic enzyme. These digestions can be carried out by a variety of enzymes; among these, trypsin is by far the most common. Trypsin has the advantage of creating peptides of ideal length for MS analysis: the peptides are comprised of approximately 6-20 amino acids, averaging ~8.4 for yeast.[48] This enzyme cleaves amide bonds on the C-terminal side of all lysine and arginine residues, unless they are followed immediately by a proline residue. Trypsin also shows poor activity around acidic regions, such as those containing phosphorylation sites (*vide infra*). The peptides that trypsin creates are very well suited to positive ion-CID-MS/MS as the C-terminal basic residue ensures protonation and allows at least one additional mobile proton along the peptide backbone, leading to complete fragmentation.[49] The lack of internal basic residues

both improves fragmentation and guarantees no higher charge state (>4+) peptides. Lower charge state peptides (2+, 3+) have been found to be more suitable for CID[50] rather than more basic, higher charged peptides, which are more suitably fragmented using electron-transfer dissociation (ETD).[51]

The fragmentation chemistry of peptides is fundamental to bottom-up proteomics and the common fragments in CID are shown in Figure 1.15. When a peptide of sufficient energy strikes an inert gas molecule in the collision cell of an instrument, kinetic energy is transferred along the peptide backbone into internal vibrational energy.[52] When sufficient internal energy is transferred to the peptide, a covalent amide bond may break and yield b- and y-type ions. Protonation weakens the amide bonds and, since protons can move along the peptide backbone according to the mobile proton theory,[49] a collection of fragments corresponding to a series of b- and y-ions will be formed. As shown in Figure 1.15, b-ions are numbered sequentially from the N-terminus to the C-terminus; conversely, the y-ions are numbered sequentially from the C-terminus to the N-terminus.

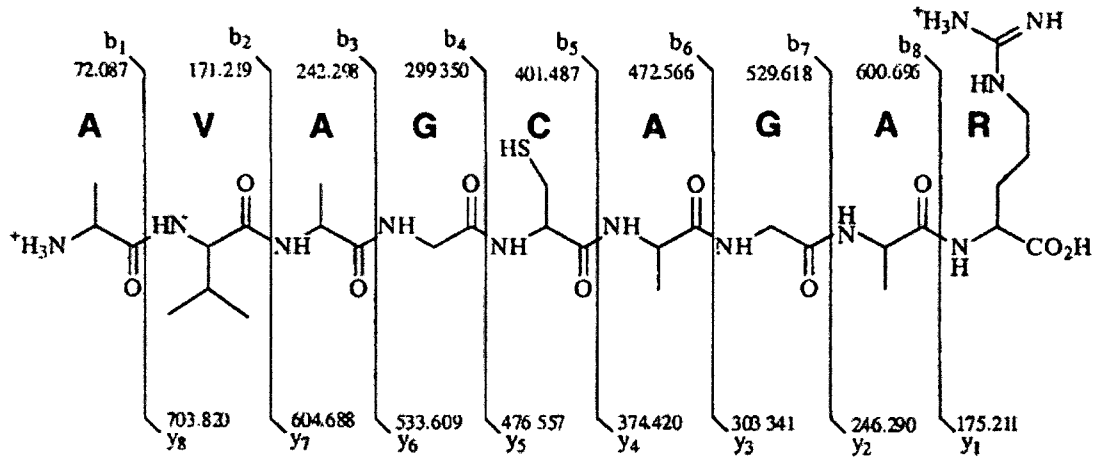


Figure 1.15. Formation of b- and y-ions from fragmentation of peptide AVAGCAGAR. Reproduced from Liebler *et al.*[53]

To collect both MS and MS/MS spectra for bottom-up proteomics experiments, an information-dependent acquisition method is typically used. A mass spectrum is taken, typically in the range of 400 to 2000 Th for peptides (*vide infra*). The MS software then will analyze charge states of all ions in the spectrum by looking at the isotopic envelope of the ion. Since carbon has a naturally occurring isotope (^{13}C) of ~1% abundance and that is 1 Da greater than ^{12}C, it is probable that a substantial portion of the peptides present will contain 1 or more ^{13}C isotopes. A heavier isotopic peak will then occur according to the equation

$$\frac{M + n\mathrm{H}^+}{n\mathrm{H}^+} \tag{1.12}$$

where n is the charge on the peptide and M is the mass of the peptide; for example, a peptide of charge 2+ will have isotopic spacing of 0.5 Th. Once the software has determined the charge states of the ions in the MS, it will select a pre-determined number of ions of charge 2+ or greater for MS/MS. It will then repeat this series of

MS and MS/MS events, typically storing the masses already collected on an exclusion list for a defined period of time or until the experiment is over. The spectra are then compiled for peptide identification and data analysis (*vide infra*). This process is often referred to as *shotgun proteomics*.

1.3.1 Identification of peptides from spectra

Peptides can be identified by two strategies, either *de novo* sequencing or computational database searching. *De novo* sequencing, pioneered by Donald Hunt and colleagues,[52] is the traditional method as it is relatively simple and can be done manually with no computational aspect. A MS/MS spectrum can be sequenced by subtracting masses of the ions in the spectrum; these differences in mass sequentially reveal the amino acid sequence. For example, for the peptide AVAGCAGAR, peaks at 600.7 and 529.6 corresponding to the b_7 and b_6 ions, respectively, would be subtracted to obtain a mass of 71.1, revealing alanine as the 8^{th} residue. This process is repeated for the entire spectrum, both b- and y-ions, to obtain the sequence. *De novo* sequencing, however, suffers from several drawbacks. First, it is rare for a sequence to contain all fragment ions (especially when internal basic residues are present), making it difficult to get an entire sequence. Second, leucine and isoleucine are isobaric and impossible to differentiate based on mass alone. Third, this process is very time consuming to be done by hand, but developments in technology have allowed the process to be done computationally. Finally, this method becomes difficult when spectra are complicated by PTMs or

multiple species are present as neutral loss or contaminant peaks can be present. For these reasons, *de novo* sequencing is not done unless the genome of the species that a proteomic sample belongs to has *not* been sequenced.

When the genome of a species has been sequenced, all proteins (both known and theoretical) can be compiled and used for database searching, as originally shown by Yates and colleagues.[54] Peptides and their corresponding proteins are identified by comparing experimental MS/MS spectra to an *in-silico* digested database, which can be automated through tools like Mascot, X!Tandem, or SEQUEST. Mascot scores matches through a variation of the Mowse scoring algorithm, first formulated in 1993 by Pappin and colleagues.[55] Mass ranges in the database are divided into protein sequence intervals of 10 kDa, *i*, and peptide intervals of 100 Da, *j*, to give a frequency matrix **F**. Theoretical peptides are put into (*i,j*) classes depending on their peptide and parent protein masses for a total number of peptides, $n_{i,j}$. The frequency matrix is then normalized to give the Mowse factor matrix **M** by

$$m_{i,j} = \frac{f_{i,j}}{|f_{i,j}|_{\text{max column j}}} \tag{1.13}$$

The Mowse score for a peptide is then

$$S = \frac{50{,}000}{M_{protein} \times \prod_n m_{i,j}} \tag{1.14}$$

where $M_{protein}$ is the molecular weight of the protein. The product term calculates the Mowse factor elements for every match between experimental and theoretical data. Mascot utilizes a probability-based model of the Mowse algorithm, so all matches have a corresponding probability of whether the match is significant or not, that is, the probability that the match is due to a random event. These probabilities are converted to a Mascot score using

$$S_{Mascot} = -10\log(P) \tag{1.15}$$

where S_{Mascot} is the Mascot score and P is the absolute probability that the match is a random event. There are 2 distinct scores within Mascot: the ions score and the protein scores. The ions score is based on the probability that an experimental MS/MS spectrum matches a theoretical spectrum; manual inspection of the hit will show the matched b- and y-ions and a plot of their corresponding errors. The protein score is, for small data sets, a simple sum of the ions scores for each unique sequence of the protein. For large datasets, however, this method would create high protein scores if multiple low-scoring ions were present; therefore, when datasets exceed the size of the databases being searched, Mascot does not let ions scores below the significance threshold(s) contribute to the proteins scores.

1.3.2 Quantitation

The quantitation of proteins by mass spectrometry can be categorized into 3 general methods: MS-based, MS/MS-based, and label-free, described graphically in

Figure 1.16. Label-free methods are the simplest, provide the best sequence coverage, and span the largest dynamic range of all the methods, but are considered unreliable as ionization conditions may change run-to-run.[56] MS/MS-based methods have become very popular with the commercialization of iTRAQ (isobaric tag for relative and absolute quantitation). This technique works by the simple addition of an iTRAQ reagent to each sample, combining them, and running them simultaneously. The MS/MS reporter ions give the relative ratios between the peptides. Advantages of this technique include the ability to compare up to 8 samples and the increased sensitivity for a species due to the isobaric nature of the reagents. The reagents are very expensive, however, and there are concerns that some of the reporter ions may overlap with peptide fragment ions, particularly when 8 samples are being compared on low-resolution instruments.[57]

MS-based quantitation methods are further classified into chemical labeling, enzymatic labeling, and metabolic labeling. Metabolic labeling involves growing cells in media containing light and heavy (^{15}N or ^{13}C) arginine and/or lysine amino acids. After several growth cycles, the light or heavy amino acids are fully incorporated into all proteins of the cells. Termed stable istope labeling by amino acids in cell culture, (SILAC),[58] it is considered one of the most accurate techniques as the labeling is done biologically and protein samples can be mixed immediately after lysis to reduce variance. The labeled amino acids are cost-prohibitive, however, and can only be used on cell cultures. Enzymatic labeling is typically done by incorporation of ^{18}O (2 molecules, 4 Da) into the C-terminal end of a peptide during proteolytic digestion, yet it can be extended to the use of any proteolytic enzyme.

Enzymatic labeling does not introduce chemical artifacts, but it has been shown that complicated algorithms must be used to correct for labeling efficiency during quantitation.[59] Chemical labeling is the most common as it is arguably the most simple, and for this reason several reagents have been commercialized (e.g. TMT, ICAT). This technique derivatizes a functional group on a common amino acid using isotopically light and heavy reagents for multiple samples. Labeling schemes have been developed using cost-effective and widely available chemicals, such as dimethyl labeling by sodium cyanoborohydride/formaldehyde and methyl esterification by thionyl chloride/methanol. Dimethyl labeling will be discussed in detail in this work (*vide infra*). While not as accurate as SILAC or allowing as much proteome coverage or linear dynamic range as label-free methods, chemical peptide labeling is an adequate "middle ground" quantitative method for shotgun proteomics.[56]

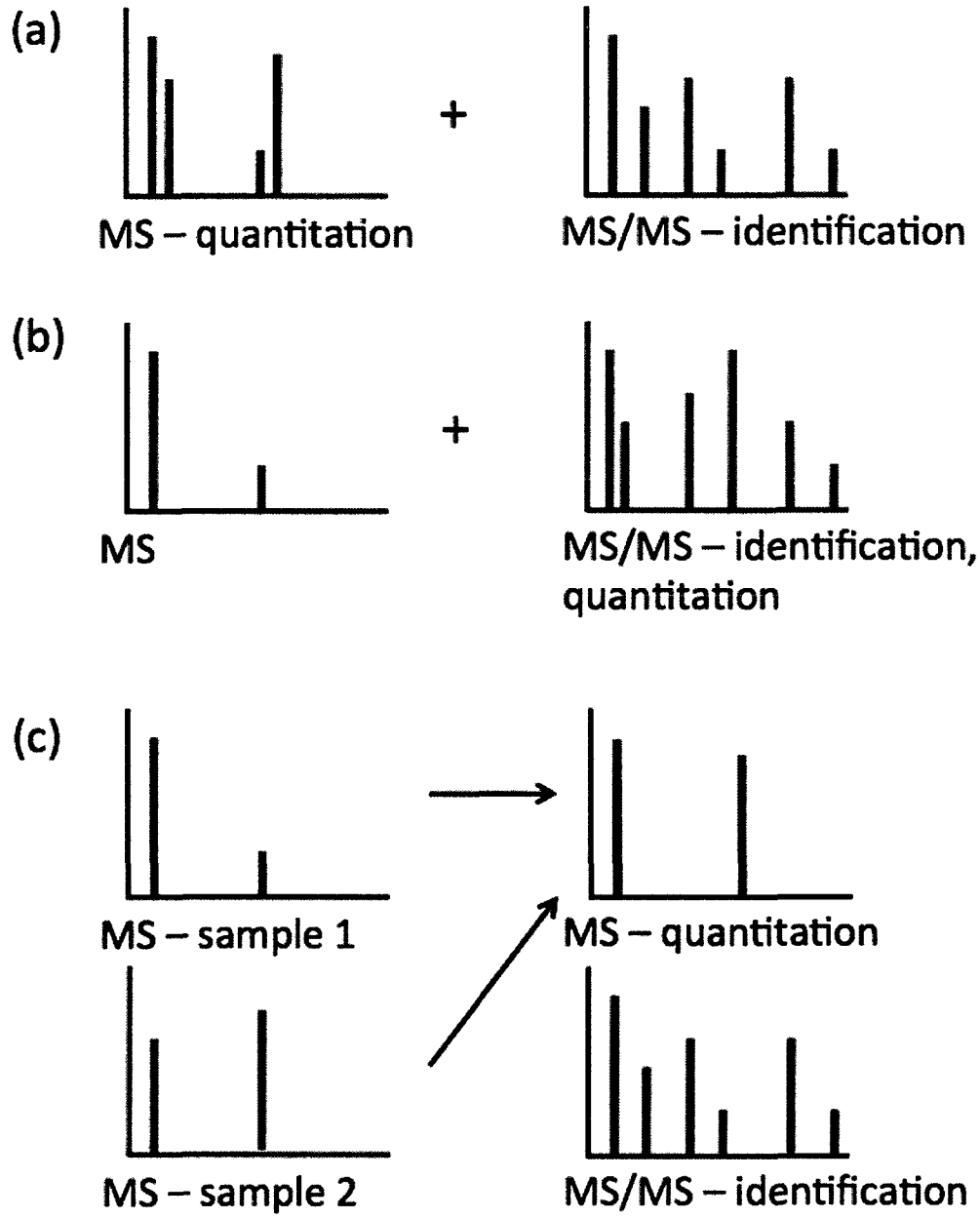


Figure 1.16. A comparison of quantitation methods in mass spectrometry-based proteomics. (a) MS quantitation. Light and heavy peptides are combined and run concurrently. The MS is used to find L/H ratios and MS/MS used for identification. (b) MS/MS quantitation. Samples are labeled such that they are isobaric, but MS/MS of the ions give reporter ions that are used to find relative ratios. (c) Label-free quantitation. Samples are run separately and ion count from the MS is used to find peptide ratios.

1.3.3 Phosphoproteomics

The identification of phosphorylation sites is possible using mass spectrometry as phosphopeptides follow the same structural fragmentation, but with a few difficulties that make phosphoproteomics a challenge. Phosphoserine and phosphothreonine residues can undergo a neutral loss of phosphoric acid (H_3PO_4) corresponding to 98 Da. This neutral loss results in incomplete fragmentation of the peptide and confounds identification and phosphorylation site localization.

Phosphorylated proteins may also have very low stoichiometry compared to their non-phosphorylated counterparts; some residues may be quantitatively phosphorylated, but some may be very low (1-2%), especially for those regulating signaling pathways.[21] The phosphoproteome may contain anywhere between 100,000 distinct phosphorylation sites,[19] with upper-limit estimates reaching 1,000,000 phosphorylation sites.[60] Phosphoproteomics therefore deals with samples with a high degree of complexity and a low abundance of the species of interest.

There is currently some debate concerning the ionization efficiencies and elution times of phosphopeptides in LC-ESI-MS/MS. Steen and colleagues showed that phosphorylated/non-phosphorylated pairs showed similar ionization efficiencies[18] while contrary evidence demonstrated phosphopeptides having only 50% of the ionization efficiency compared to their non-phosphorylated counterpart.[61] There is also research showing that phosphorylated/non-

phosphorylated pairs elute at approximately the same times,[18] while others have shown that phosphopeptides elute very early on and a decrease in loading acetonitrile concentration (2% versus 5%) in fact increases phosphopeptide identification by up to 26%.[62]

Enrichment is needed due to low stoichiometry and the dynamic range limitations of QqTOF mass spectrometers. Enrichment can be accomplished by immobilized metal affinity chromatography (IMAC) using various metals such as Fe(III),[63] Ga(III),[64] or Zr(IV); metal oxide affinity chromatography (MOAC) using titanium dioxide (TiO_2)[65] or zirconium dioxide (ZrO_2)[66]; or chemical methods such as ß-elimination[67] or phosphoridate chemistry (PAC).[68] Development of enrichment techniques remains an active area of research; recent developments have been thoroughly discussed in a recent review by Dunn and colleagues.[69] Many of these techniques are said to be "complementary" towards each other in that they enrich for distinct sections of the phosphoproteome. A study comparing PAC, IMAC, and MOAC showed that the overlap of phosphopeptide identification between these techniques was only in the range of 33-35%.[70]

IMAC has become the enrichment method of choice among these due to its specificity, price, and commercial availability. Recently, Gygi and colleagues have identified in excess of 5,500 and 13,000 phosphorylation sites in mouse liver[14] and Drosophila embryos,[71] respectively, using a SCX/IMAC protocol: a testament to the power of IMAC coupled to mass spectrometry. IMAC works on the principle that an immobilized metal – Fe(III) being the most common – can co-ordinate with a

phosphate group, as shown in Figure 1.17. The Fe(III) ions are held in place by a chelating agent such as iminodiacetic acid (IDA) that is covalently bound to solid support beads, such as silica. The deprotonated phosphate groups from phosphopeptides co-ordinate with the Fe(III) and the non-phosphopeptides washed from the resin. The phosphopeptides can the be specifically eluted from the resin by a phosphate buffer that provides competitive binding with the phosphopeptides; as the phosphates from the buffer are in excess, the phosphopeptides are eluted, giving a phosphoproteomic sample. During this process, many factors come into play, such as buffer composition, concentration, and pH. The factors influencing specific and non-specific binding and pH control have been investigated systematically by Tsai and colleagues.[72]

As discussed in section 1.3, CID is outperformed by ETD when the charge state of the peptides is greater than 2+; this remains true for phosphopeptides. ETD also has the advantage over CID in that it does not lose the labile phosphate group as a neutral loss and therefore ETD may aid in phosphorylation site localization. There is also evidence that phosphate groups may undergo gas-phase migration when undergoing CID in ion traps, confounding phosphosite localization.[73] This effect has been shown, however, to contribute negligibly to mass spectra[51] and is not believed to occur in QqTOF instruments where the activation time in the collision cell is much shorter. In large-scale phosphoproteomics, CID has still been found to perform far better than ETD due to the lack of internal basic residues in tryptic peptides. Studies in yeast using the SCX-IMAC approach have shown that CID can identify ~10,000 unique phosphopeptides versus ~2,200 found by ETD.

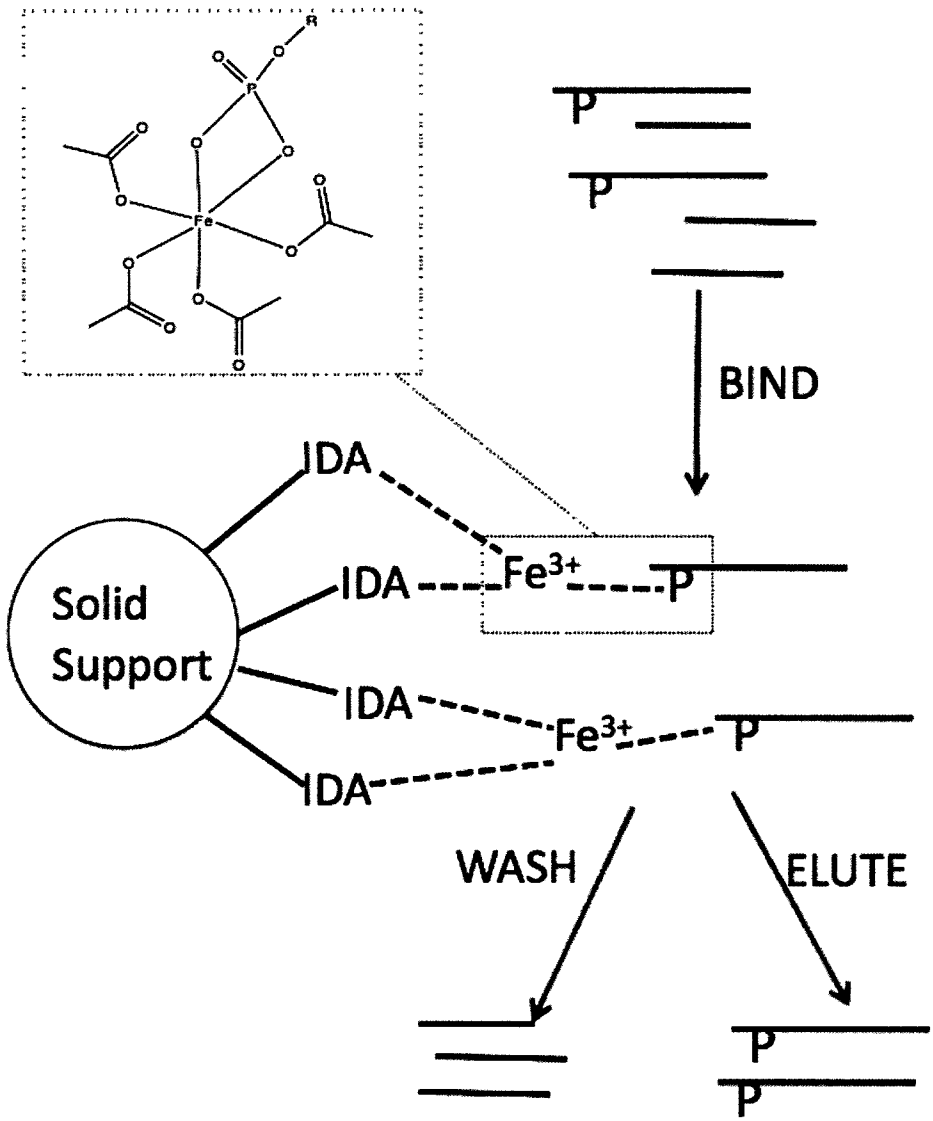


Figure 1.17. General schematic for enrichment of phosphopeptides from peptide mixture using Fe(III) IMAC.

1.4 Separation methods

Peptides can be separated by exploiting the diversity of their physicochemical properties. Strong-cation exchange (SCX) and strong-anion exchange (SAX) chromatography can separate peptides based on their isoelectric

point (pI), reverse-phase chromatography can separate based on hydrophobicity, and immobilized metal affinity chromatography (previously discussed in section 1.3.3) can separate based on the presence of phosphate groups.

As previously discussed in section 1.3, analyses by shotgun proteomics are conducted on very complex samples and thus require a high degree of separation. As electrospray (or nanospray) is an online, liquid-based technique, samples must be introduced by a solvent pumping system called a high-performance liquid chromatograph (HPLC). As the HPLC is used solely to deliver a gradient to the analytical column, the details of this instrument will not be further discussed; it is sufficient to say that two solvents are pumped through the system at a defined flow rate and change in solvent composition over time. Since electrospray is generally intolerant towards salts due to clustering, reverse-phase chromatography has become the standard for hyphenated instrumental setups with mass spectrometry (RPLC-MS). An aqueous sample containing peptides can be loaded onto a reverse-phase chromatographic column (via pressure vessel, *vide infra*). Reverse-phase resins are generally C_{18} covalently bound to silica beads and thus only molecules with hydrophobicity have affinity towards these resins. Salts have no hydrophobicity and do not bind or are quickly removed with aqueous flow.

The separation of the analytes is accomplished by gradient chromatography, in which the least hydrophobic peptides elute first at high aqueous concentrations and hydrophobic peptides elute later at higher organic (e.g. methanol, acetonitrile) concentrations. Therefore larger peptides abundant in hydrophobic amino acids (A,

I, L, M, F, W, Y, V, see Figure 1.2) will have more affinity, and thus elute later, than smaller peptides or those abundant in hydrophilic peptides (R, H, K, D, E).

The chromatographic separation can be described in terms of number of theoretical plates, N, and plate height, H, a theory first described by Martin and Synge.[74] These parameters are related via column length, L, by the equation

$$N = \frac{L}{H} \tag{1.16}$$

Assuming Gaussian peak shape, the number of theoretical plates can be calculated by

$$N = 16\left(\frac{t_R}{W}\right)^2 \tag{1.17}$$

where t_R is the retention time and W is the width of the chromatographic peak. At every plate, there is an equilibration between the analytes in the mobile and stationary phases and thus a higher number of theoretical plates yields a greater chromatographic separation between analytes. The plate theory does not, however, describe band broadening adequately; a rate theory is therefore required. The most simple formulation that has been shown to model column efficiency quite successfully[75] was introduced by Van Deemter and named after him to become the Van Deemter equation.[76] It describes the plate height in the terms

$$H = A + \frac{B}{u} + Cu \tag{1.18}$$

where u is the linear velocity of the mobile phase, A is due to eddy diffusion, B is the longitudinal diffusion coefficient, and C is the mass transfer coefficients (the sum of the mass transfer coefficients from mass transfer to and from the stationary phase, C_s, and in the stationary phase, C_m). A detailed discussion of these terms is beyond the scope of this work, however the immediate effects of these terms are clear. The eddy diffusion term is significant at high flow rates as variation in molecule paths cannot be averaged out by diffusion. The use of nanoflow and of small particle size chromatographic resin, therefore, makes this term have a negligible effect on band broadening. The second term also has a negligible effect on band broadening, as longitudinal diffusion is more likely to occur in gases (i.e. GC) than liquids (HPLC). However, at nanoflow rates, this term may become more substantial. The third, mass-transfer term generally determines the band broadening effect, especially if the flow rate is above the optimal linear velocity determined by a Van Deemter plot of plate height against linear velocity.[77]

It is important to note that in most analytical applications, these chromatographic terms are very important since chromatographic reproducibility is necessary for quantitative applications. In quantitative shotgun proteomics, however, the analytes being quantified are isotopically paired in the same runs and therefore run-to-run reproducibility is of little importance (*vide infra*).

For a second dimension of separation, SCX fractionation can be used. SCX resins contain strongly acidic groups like sulfonic acid bound to a solid support; these groups can bind positively charged ions according to the equilibrium

$$x\mathrm{RSO_3^-H^+}_{(s)} + P^{x+}{}_{(aq)} \rightleftharpoons (RSO_3^-)_x P^{x+}{}_{(s)} + x\mathrm{H^+}_{(aq)}$$

where P is the peptide, x is the charge state of the peptide, and R is the spacer between the solid support and the acidic group. Most modern resins use polysulfoethyl aspartamide, as shown in Figure 1.18. When the pH of a sample is lowered to ~2, all aspartic acid (pKa 3.65) and glutamic acid (pKa 4.25) residues will be neutral and the N-termini of the peptides will be positively charged, resulting in all peptides being overall positively charged. Peptides will then bind to the resin and can be eluted sequentially either by buffers of increasing salt concentration or of increasing pH. An increase in salt concentration will result in a competition for the binding sites of the SCX resin and elute in the order of binding strength. An alternative is pH elutions, in which an increasing pH will elute the peptides in increasing isoelectric points; the peptides will elute from the SCX after being neutralized.

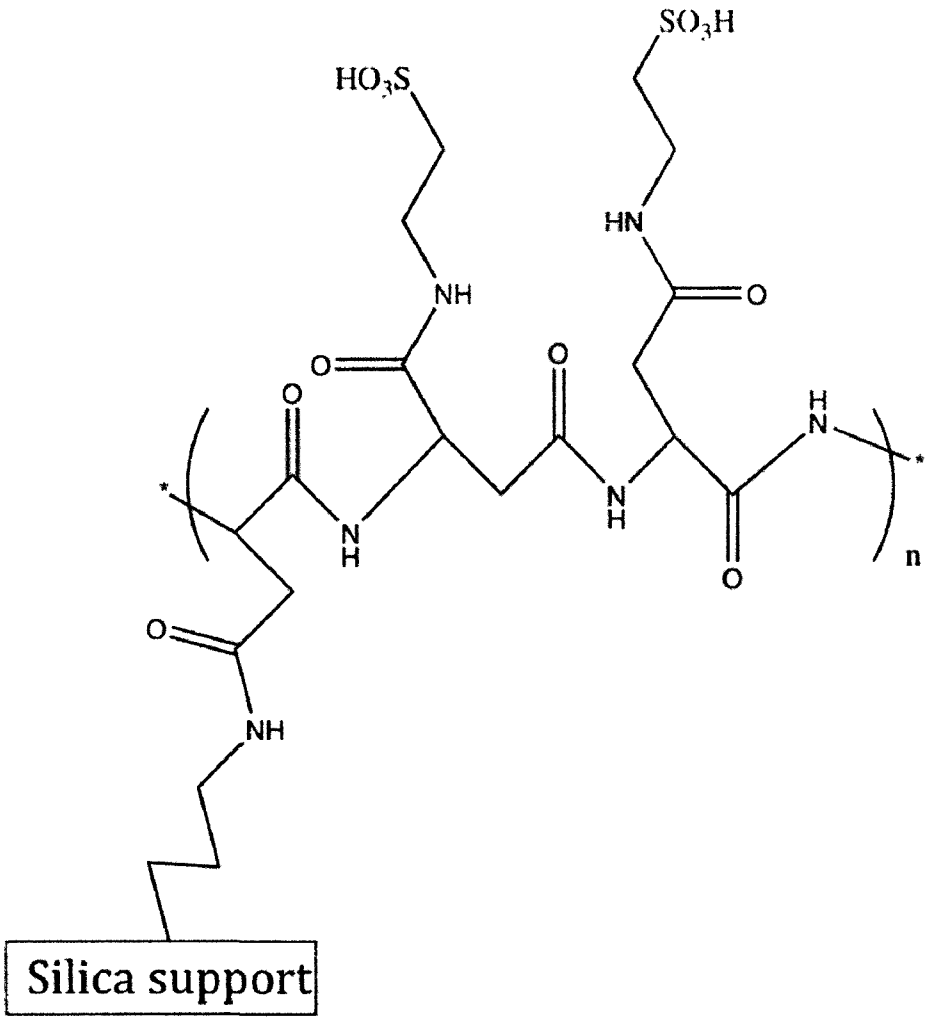


Figure 1.18. Typical SCX resin consisting of solid support and polysulfoethyl aspartamide. The sulfate groups are negative after conditioning and provide the ionic attraction for the positively-charged peptides.

SAX fractionation follows the same principles except the charges are reversed, that is, functional groups on the resin are positively charged, samples are made alkaline, and the peptides are eluted by step-wise decreasing pH.

1.5 Vesicular Stomatitis Virus

Vesicular stomatitis virus (VSV) is a bullet-shaped (180 nm long, 75 mm wide), negative-strand RNA virus of the Rhabdoviridae family, illustrated in Figure 1.19. It has an 11,161 base-pair genome and an RNA polymerase that converts this genomic to subgenomic mRNA. This genome encodes for 5 proteins: nucleocapsid proteins (N) that encapsulate the genome, two polymerase subunits (L, P), an internal matrix protein (M), and a transmembrane glycoprotein (G) that is responsible for cell binding and membrane fusion; these proteins are encoded in the order N-P-M-G-L from 3' to 5'. VSV also encodes two non-structural, basic proteins that have been found in the cytoplasm of infected cells[78] but are not required for viral replication.[79]

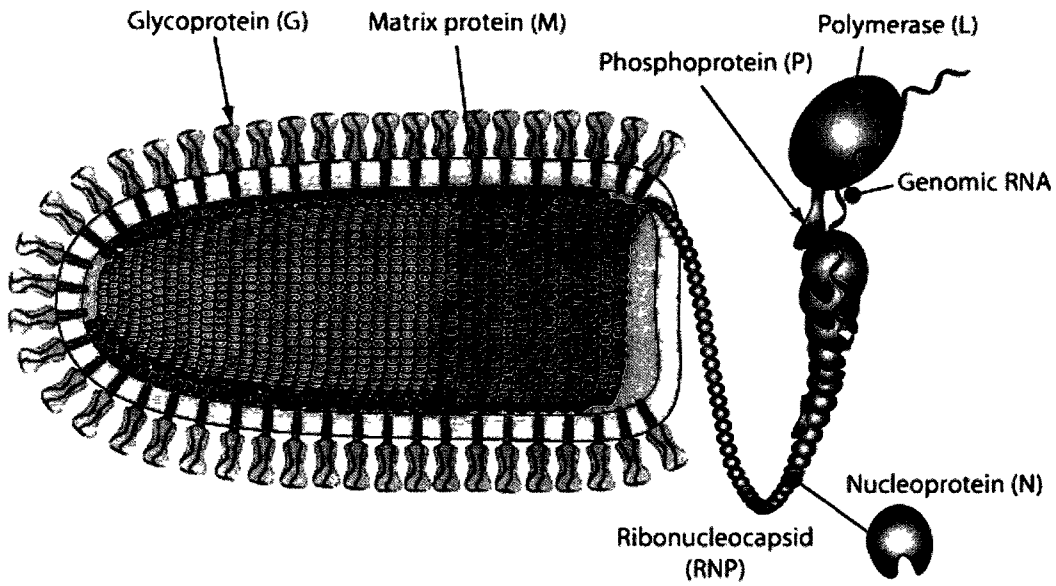


Figure 1.19. Structure of VSV and localization of its 5 proteins. Reproduced from ViralZone, "Vesicular stomatitis virus".[80]

The N protein is the nucleocapsid protein that is responsible for encapsulating the RNA genome into a tight, helical, RNase-resistant core. Relatively little is known about the domains of this protein; however, it has been shown that the C-terminus is required for RNA binding[81] and the N-terminal domain is required for N-P interactions.[82]

The P protein is an acidic phosphoprotein that, along with L, forms part of the viral transcriptase-replicase complex. It is a 265 amino acid protein consisting of 3 domains and a variable region known as a hinge.[83] Phosphorylation on serine and threonine residues in the N-terminal acidic domain is required for polymerase activity and occurs via the action of casein kinase II.[84] Upon phosphorylation, P can form trimers and bind to L and N-RNA to form the active form of the transcriptase.[85] There is evidence that differential phosphorylation of P may switch the polymerase complex between transcription and replication roles.[86]

The L protein is a large protein of approximately 2,100 amino acids in length and is also a part of the RNA polymerase complex. It is a multifunctional protein responsible for transcription and replication, mRNA capping,[87] methylation,[88] and polyadenylation,[89] although the domains of this protein have not been completely mapped.

The G protein is a membrane glycoprotein of approximately 500 amino acids that forms trimeric spikes protruding from the membrane, with a 20 amino acid transmembrane domain and a 29 amino acid cytoplasmic domain.[90] At low pH (< 6.1), the trimers are stabilized and undergo conformational changes that expose a

hydrophobic domain. This hydrophobic domain targets host membranes and mediates membrane fusion.[91] This process is believed to also occur at low temperatures so, as a practical matter, most infections are done at low pH and/or 4°C rather than 37°C to increase adsorption.[92, 93] Note that after binding, the virus is incorporated into the cell by typical endocytosis through a clathrin-coated vesicle mechanism.[93] The RNP complex is then released into the host-cell followed by subsequent transcription of the genome.

The matrix protein is the smallest and most abundant protein in VSV, however its functions are critical and diverse. While its exact location in the viral structure has not been confirmed,[94] it does play roles in virus assembly, virus budding, and nearly all effects on host function described below. The M protein is a phosphoprotein, however the function(s) of the phosphorylation sites are not currently known.[95]

Recently it has been shown that VSV induces cytolysis of malignant cell lines but is unable to do so in primary, non-malignant cell lines.[96] Several properties of VSV made it a good candidate as a potential lytic agent for cancer cells and therefore has been an active area of research with hopes the virus can be adapted for cancer treatment in humans. The virus is non-pathogenic in humans and exposure is not common.[94]

The helical structure of the nucleocapsid and modular, start-stop nature of the genome (with conserved start and end sequences) allow the insertion of foreign genes into VSV. It has been shown that insertion of genes smaller than 2,000

nucleotides does not affect the growth or replication efficiency of VSV.[97, 98] In fact, even the insertion of two HIV genes totaling 4,200 nucleotides (~40% increase in genome size) only resulted in a three-fold decrease in titer.[99]

Animal models show that exposure to VSV results in strong immune responses, both innate and adaptive, that are required for elimination of the virus from the body. VSV triggers a cytotoxic T-lymphocyte immune response against the G and N proteins of VSV; without this innate response, the virus will replicate to lethal levels *in vivo* prior to the adaptive immune response being able to respond.[100] Conversely, mice deficient in B-cells (i.e. without an adaptive immune response) will also experience neuropathy and death after approximately 9 days, even with an intact innate immune response.[100]

The immune response to VSV is believed to be an *interferon-mediated* response. Interferons (IFNs) are a class of antiviral cytokines discovered in the late 1950s[101] that activate over 30 cellular genes and inhibit the replication of viruses. The interferons have been classified into two types: type I (α/β) and type II (γ). The viral RNA activates the signaling cascades, via NF-κB and IFN regulatory factor 3 (IRF-3) pathways, which leads to the transcriptional activation and secretion of IFN-β.[102, 103] The secreted species then bind to cell-surface receptors, which in turn activate a family of kinases known as the Janus protein kinases and the STAT1/2 pathways (signal transducer and activator of transcription).[104] STAT1 and STAT2 heterodimers bind IFN-stimulated response elements and activate a number of genes encoding for proteins that either cause apoptosis or that directly prevent

virus replication, including heat shock proteins, death ligand TRAIL, dsRNA-dependent protein kinase (PKR), more IRF species, and major histocompatibility (MHC) antigens, as well as many others.[105] The IFNs also link innate and adaptive immune responses through the activation of natural killer and T-lymphocyte cells and triggering immunity by differentiation into dendritic cells (DCs). These IFNs have been shown to be critical for immune response to viral infection: studies have shown that both type I and type II IFNs are essential for survival in mice infected with viruses and with otherwise normal immune systems.[106] Without a functioning IFN system, VSV is able to target and block nuclear export of host mRNA, preventing antiviral immune responses and hijacking the translational machinery of the cell.[107, 108] Translational machinery is required for a virus to replicate in the cell; it is not surprising that defects have been found in transformed cell lines. The oncogenic elongation factor, eEF1A, has been found to be mutated in some ovarian tumors,[109] the mRNA cap-binding factor eIF1A has been found to be upregulated in some transformed cell lines,[110] and defective signaling of PKR and eIF2α have been found to facilitate oncolysis by VSV.[111] PKR is an IFN-inducible threonine serine kinase – as such, upon interaction with a dsRNA species, it is activated by autophosphorylation and in turn catalyzes the phosphorylation of eIF2α on serine 51. This process sequesters eIF2B and causes a dramatic inhibition in translation.[112] It has further been found that non-immortalized cells lacking STAT1 and PKR were infected by VSV after 30 hours; however, VSV infects immortalized cells lacking STAT1 or PKR after only 3 hours, indicating that other mechanisms are involved in enhancing VSV replication.[111] The importance of the kinase PKR and the phosphorylation site on

eIF2α indicates that other phosphorylation processes may be involved during the infection and replication of VSV; as such, a phosphoproteomics approach could reveal other potential co-operative pathways.

While this section has so far dealt with the factors affecting replication of the virus, it is important to consider how oncolysis occurs. It has been shown that oncolysis occurs via apoptosis by activation of caspase-3 and caspase-9,[113] although there is evidence for roles of both the intrinsic and extrinsic apoptotic pathways. A summary of apoptotic pathways shown to be involved in VSV-induced oncolysis are shown in Figure 1.20. The aforementioned discussed blocking of nuclear mRNA transport leads to the depolarization of the mitochondrial membrane, subsequent mitochondrial membrane permeabilization and release of cytochrome c and apopotosis-inducing factor (AIF).[113] Cytochrome c activates the apoptosome complex as well as downstream effector caspases,[114] leading to cell death. Antiapoptotic factor BCL-2 regulates the mitochondrial-mediated apoptotic pathway by interfering with the ability of proapoptotic Bax/Bak to oligomerize and release cytochrome c.[115] Apoptosis can also be induced by the extrinsic pathway; it has been shown that cross-talk between caspase-8 and caspase-9 allows the activation of caspase-9 from the extrinsic pathway.[116] Note that there are other apoptotic pathways linked to VSV infection as this is an active area of research: relevant pathways will be further discussed with regards to the work done in this thesis (*vide infra*).

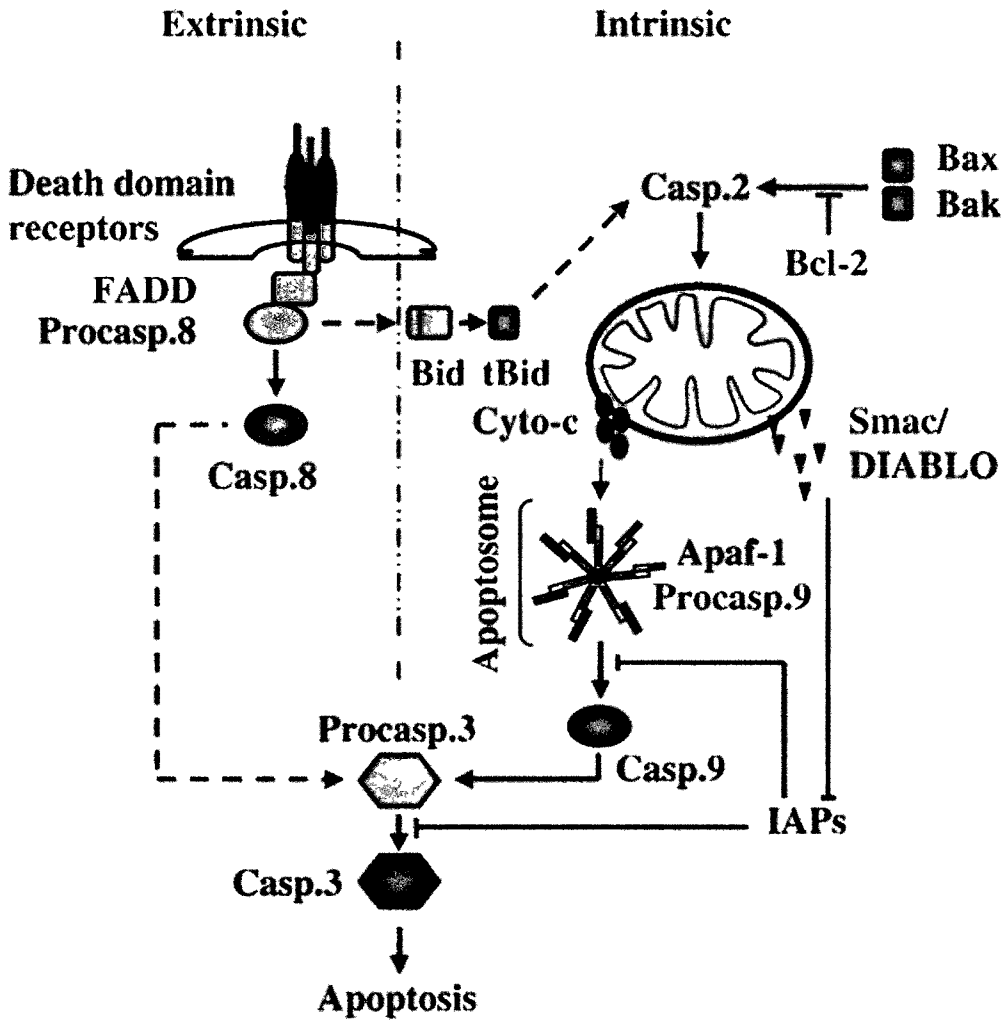


Figure 1.20. Apoptotic pathways believed to be involved in apoptosis by VSV. Pathways believed to not be involved are shown as dashed lines. Reproduced from Sharif-Askari *et al.*[117]

Information on the mechanisms of VSV infection could lead to improved methods to selectively kill cancer cells and simultaneously attenuate VSV towards normal cells. As mentioned, VSV is a good candidate for foreign gene insertion and groups have successfully inserted genes such as the herpes virus thymidine kinase suicide cassette (TK), interleukin-4 (IL-4), or IFNβ gene with success.[118, 119] VSV

oncolysis has also been shown to be more effective when combined with an inhibitor of the antiapoptotic protein BCL-2.[120, 121] The proteins expressed by the virus (i.e. G, M) can be mutated to increase attenuation towards healthy cells and be hypervirulent in cancer cells, as demonstrated on the Maraba-strain Rhabdovirus.[122] Natural strains of VSV (AV1 and AV2) have been found to be attenuated towards normal cells compared to wild-type VSV due to mutations in the G protein[123]; these mutations lead to a strain failing to block the transport of host mRNA from the nucleus to the cytoplasm, and therefore normal cells have a stronger antiviral response. In fact, while primary leukemic cell lines have been shown to undergo time-dependent oncolysis, primary chronic lymphocytic leukemia (CLL) lymphocytes *ex vivo* do not undergo oncolysis, surprisingly.[124]

The mechanisms of VSV infection in malignant cells and the pathways leading to apoptosis are thus of great interest. Interestingly, while it has been observed there is a great diversity in the molecular mechanism of different types of cancer, VSV has been shown to effectively infect a remarkably large number of cancer cell types. VSV has shown success in glioma, breast cancer (MCF-7, TS/A), prostate cancer (PC3), B-lymphoma (Karpas-422, A20), and melanoma (B16-F10, BHK-21), among many others.[118, 120, 125] Therefore, the choice in cell line was not deemed of high importance as these malignancies must have common mechanisms with regards to VSV infection and induced apoptosis. Regardless, a relatively rarely used cell line with regards to VSV research was chosen, the human CML (chronic myelogenous leukemia) K562 cell line, derived in the early 1970s from a 53-year old

CML patient.[126] It is easily grown as a suspension culture (non-adherent) and was chosen to be used as a model malignant cell line.

The importance of phosphorylation in signaling pathways and the induction of antiviral or apoptotic genes by VSV suggest a joint proteomic/phosphoproteomic investigation into VSV infection could lead to novel information regarding its molecular dynamics. To date, no such study has been performed using a shotgun proteomics/mass spectrometry-based approach. Initial work involved discovering and optimizing a quantitative strategy for analysis of complex proteomes and appropriate lysis, labeling, separation, enrichment, and MS analysis steps were researched. A background proteome and phosphoproteome was obtained from K562 cells using SCX and IMAC approaches, respectively. Finally, the mechanisms of VSV infection and apoptosis were investigated using the developed quantitative proteomic/phosphoproteomic strategy as well as a rigorous bioinformatics analysis.

References

1. Lander, E. S. et al. Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921 (2001).

2. Lashkari, D. A. et al. Yeast microarrays for genome wide parallel genetic and gene expression analysis. *Proc Natl Acad Sci U S A* **94**, 13057-13062 (1997).

3. Ishkanian, A. S. et al. A tiling resolution DNA microarray with complete coverage of the human genome. *Nat Genet* **36**, 299-303 (2004).

4. Nobel Prize 2002. *http://nobelprize.org/nobel_prizes/chemistry/laureates/2002/*

5. 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 (1989).

6. 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).

7. Lodish, H. et al. *Molecular Cell Biology* (W. H. Freeman, 2003).

8. Fermi, G., Perutz, M. F., Shaanan, B. & Fourme, R. The crystal structure of human deoxyhaemoglobin at 1.74 A resolution. *J Mol Biol* **175**, 159-174 (1984).

9. Selkoe, D. J. Folding proteins in fatal ways. *Nature* **426**, 900-904 (2003).

10. Anfinsen, C. B. Principles that govern the folding of protein chains. *Science* **181**, 223-230 (1973).

11. Wikimedia Commons. *http://en.wikipedia.org/wiki/File:Main_protein_structure_levels_en.svg*

12. Wodak, S. J. & Janin, J. Location of structural domains in proteins. *Biochemistry* **20**, 6544-6552 (1981).

13. Valmiki, M. G. & Ramos, J. W. Death effector domain-containing proteins. *Cell Mol Life Sci* **66**, 814-830 (2009).

14. Villen, J., Beausoleil, S. A., Gerber, S. A. & Gygi, S. P. Large-scale phosphorylation analysis of mouse liver. *Proc Natl Acad Sci U S A* **104**, 1488-1493 (2007).

15. Finishing the euchromatic sequence of the human genome. *Nature* **431**, 931-945 (2004).

16. UniProt. UniProt Knowledgebase: Controlled vocabulary of posttranslational modifications. (2011).

17. Mann, M. & Jensen, O. N. Proteomic analysis of post-translational modifications. *Nat Biotechnol* **21**, 255-261 (2003).

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

19. Kalume, D. Tackling the phosphoproteome: tools and strategies. *Curr Opin Chem Biol* **7**, 64-69 (2003).

20. Hunter, T. & Sefton, B. M. Transforming gene product of Rous sarcoma virus phosphorylates tyrosine. *Proc Natl Acad Sci U S A* **77**, 1311-1315 (1980).

21. Reinders, J. & Sickmann, A. State-of-the-art in phosphoproteomics. *Proteomics* **5**, 4052-4061 (2005).

22. Lim, Y. P. Mining the tumor phosphoproteome for cancer markers. *Clin Cancer Res* **11**, 3163-3169 (2005).

23. Cohen, P. Protein kinases--the major drug targets of the twenty-first century? *Nat Rev Drug Discov* **1**, 309-315 (2002).

24. Dass, C. *Fundamentals of Contemporary Mass Spectrometry (Wiley - Interscience Series on Mass Spectrometry)* (Wiley-Interscience, 2007).

25. Dole, M. et al. Molecular beams of macroions. *J Chem Phys* **49**, 2240 (1968).

26. Taylor, G. I. The stability of horizontal fluid interface in a vertical electric field. *J Fluid Mech* **2**, 1-15 (1965).

27. 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).

28. Iribarne, J. V. & Thomson, B. A. On the evaporation of small ions from charged droplets. *J Chem Phys* **64**, 2287 (1976).

29. 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).

30. Gale, D. C. & Smith, R. D. Small volume and low flow-rate electrospray lonization mass spectrometry of aqueous samples. *Rapid Commun Mass Spectrom* **7**, 1017-1021 (1993).

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

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

33. Schmidt, A., Karas, M. & Dulcks, T. Effect of different solution flow rates on analyte ion signals in nano-ESI MS, or: when does ESI turn into nano-ESI? *J Am Soc Mass Spectrom* **14**, 492-500 (2003).

34. Smith, R. D., Shen, Y. & Tang, K. Ultrasensitive and quantitative analyses from combined separations-mass spectrometry for the characterization of proteomes. *Acc Chem Res* **37**, 269-278 (2004).

35. Smith, R. D. Future directions for electrospray ionization for biological analysis using mass spectrometry. *Biotechniques* **41**, 147-148 (2006).

36. Miller, P. E. & Denton, M. B. The quadruple mass filter: Basic operating concepts. *J. Chem. Educ.* **63**, 617 (1986).

37. Mathieu, E. Le Mouvement Vibratoire d'une Membrane de Forme Élliptique. *Journal des Mathématiques Pures et Appliquées* 137-203 (1868).

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

39. Dass, C. *Principles and Practice of Biological Mass Spectrometry* (Wiley-Interscience, 2000).

40. Dawson, J. H. J. & Guilhaus, M. Orthogonal-acceleration time-of-flight mass spectrometer. *Rapid Commun Mass Spectrom* **3**, 155-159 (1989).

41. Verentchikov, A. N., Ens, W. & Standing, K. G. Reflecting time-of-flight mass spectrometer with an electrospray ion source and orthogonal extraction. *Anal Chem* **66**, 126-133 (1994).

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

43. Charles, L. Influence of internal standard charge state on the accuracy of mass measurements in orthogonal acceleration time-of-flight mass spectrometers. *Rapid Commun Mass Spectrom* **22**, 151-155 (2008).

44. Adams, J. & Manley, B. W. Microchannel plate detectors. *Nuclear Instruments and Methods* **162**, 587-601 (1979).

45. Adams, J. & Manley, B. W. The mechanism of channel electron multiplication. *Nuclear Science, IEEE Transactions on* **13**, 88-99 (1966).

46. Douglas, D. J. & French, J. B. Collisional focusing effects in radio frequency quadrupoles. *J Am Soc Mass Spectrom* **3**, 398-408 (1992).

47. Barbacci, D. C. et al. Multi-anode detection in electrospray ionization time-of-flight mass spectrometry. *J Am Soc Mass Spectrom* **9**, 1328-1333 (1998).

48. 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).

49. Dongre, A. R., Jones, J. L., Somogyi, A. & Wysocki, V. H. Influence of peptide composition, gas-phase basicity, and chemical modification on fragmentation efficiency: Evidence for the mobile proton model. *J Am Soc Mass Spectrom* **118**, 8365-8374 (1996).

50. Swaney, D. L., McAlister, G. C. & Coon, J. J. Decision tree-driven tandem mass spectrometry for shotgun proteomics. *Nat Methods* **5**, 959-964 (2008).

51. Aguiar, M., Haas, W., Beausoleil, S. A., Rush, J. & Gygi, S. P. Gas-phase rearrangements do not affect site localization reliability in phosphoproteomics data sets. *J Proteome Res* **9**, 3103-3107 (2010).

52. Hunt, D. F., Yates, J. R. r., Shabanowitz, J., Winston, S. & Hauer, C. R. Protein sequencing by tandem mass spectrometry. *Proc Natl Acad Sci U S A* **83**, 6233-6237 (1986).

53. Liebler, D. C. *Introduction to Proteomics: Tools for the New Biology* (Humana Press, 2001).

54. Eng, J. K., McCormack, A. L. & Yates III, J. R. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectrom* **5**, 976-989 (1994).

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

56. 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).

57. Ow, S. Y. et al. iTRAQ underestimation in simple and complex mixtures: the good, the bad and the ugly; *J Proteome Res* **8**, 5347-5355 (2009).

58. Ong, S.-E. et al. Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics. *Mol Cell Proteomics* **1**, 376-386 (2002).

59. Ramos-Fernandez, A., Lopez-Ferrer, D. & Vazquez, J. Improved method for differential expression proteomics using trypsin-catalyzed 18O labeling with a correction for labeling efficiency. *Mol Cell Proteomics* **6**, 1274-1286 (2007).

60. Boersema, P. J. et al. In-depth qualitative and quantitative profiling of tyrosine phosphorylation using a combination of phosphopeptide immunoaffinity purification and stable isotope dimethyl labeling. *Mol Cell Proteomics* **9**, 84-99 (2010).

61. Gropengiesser, J., Varadarajan, B. T., Stephanowitz, H. & Krause, E. The relative influence of phosphorylation and methylation on responsiveness of peptides to MALDI and ESI mass spectrometry. *J Mass Spectrom* **44**, 821-831 (2009).

62. Peterson, A. et al. Analysis of RP-HPLC loading conditions for maximizing peptide identifications in shotgun proteomics. *J Proteome Res* **8**, 4161-4168 (2009).

63. Andersson, L. & Porath, J. Isolation of phosphoproteins by immobilized metal (Fe3+) affinity chromatography. *Anal Biochem* **154**, 250-254 (1986).

64. Posewitz, M. C. & Tempst, P. Immobilized gallium(III) affinity chromatography of phosphopeptides. *Anal Chem* **71**, 2883-2892 (1999).

65. Pinkse, M. W. H., Uitto, P. M., Hilhorst, M. J., Ooms, B. & Heck, A. J. R. Selective Isolation at the Femtomole Level of Phosphopeptides from Proteolytic Digests Using 2D-NanoLC-ESI-MS/MS and Titanium Oxide Precolumns. *Anal. Chem.* **76**, 3935-3943 (2004).

66. Kweon, H. K. & Hakansson, K. Selective zirconium dioxide-based enrichment of phosphorylated peptides for mass spectrometric analysis. *Anal Chem* **78**, 1743-1749 (2006).

67. Oda, Y., Nagasu, T. & Chait, B. T. Enrichment analysis of phosphorylated proteins as a tool for probing the phosphoproteome. *Nat Biotechnol* **19**, 379-382 (2001).

68. Zhou, H., Watts, J. D. & Aebersold, R. A systematic approach to the analysis of protein phosphorylation. *Nat Biotechnol* **19**, 375-378 (2001).

69. Dunn, J. D., Reid, G. E. & Bruening, M. L. Techniques for phosphopeptide enrichment prior to analysis by mass spectrometry. *Mass Spectrom Rev* **29**, 29-54 (2010).

70. Bodenmiller, B., Mueller, L. N., Mueller, M., Domon, B. & Aebersold, R. Reproducible isolation of distinct, overlapping segments of the phosphoproteome. *Nat Methods* **4**, 231-237 (2007).

71. Zhai, B., Villen, J., Beausoleil, S. A., Mintseris, J. & Gygi, S. P. Phosphoproteome analysis of Drosophila melanogaster embryos. *J Proteome Res* **7**, 1675-1682 (2008).

72. Tsai, C. F. et al. Immobilized metal affinity chromatography revisited: pH/acid control toward high selectivity in phosphoproteomics. *J Proteome Res* **7**, 4058-4069 (2008).

73. Palumbo, A. M. & Reid, G. E. Evaluation of gas-phase rearrangement and competing fragmentation reactions on protein phosphorylation site assignment using collision induced dissociation-MS/MS and MS3. *Anal Chem* **80**, 9735-9747 (2008).

74. Martin, A. J. & Synge, R. L. A new form of chromatogram employing two liquid phases: A theory of chromatography. 2. Application to the micro-determination of the higher monoamino-acids in proteins. *Biochem J* **35**, 1358-1368 (1941).

75. Katz, E., Ogan, K. L. & Scott, R. P. W. Peak dispersion and mobile phase velocity in liquid chromatography: the pertinent relationship for porous silica. *J Chrom A* **270**, 51-75 (1983).

76. Van Deemter, J. J., Zuiderweg, F. J. & Klinkenberg, A. Longitudinal diffusion and resistance to mass transfer as causes of nonideality in chromatography. *Chem Eng Sci* **5**, 271-289 (1956).

77. Skoog, D. A., Holler, F. J. & Crouch, S. R. *Principles of Instrumental Analysis* (Brooks Cole, 2006).

78. Spiropoulou, C. F. & Nichol, S. T. A small highly basic protein is encoded in overlapping frame within the P gene of vesicular stomatitis virus. *J Virol* **67**, 3103-3110 (1993).

79. Kretzschmar, E., Peluso, R., Schnell, M. J., Whitt, M. A. & Rose, J. K. Normal replication of vesicular stomatitis virus without C proteins. *Virology* **216**, 309-316 (1996).

80. Expasy ViralZone Database. Vesiculovirus Entry. *http://www.expasy.org/viralzone/all_by_species/21.html*

81. Das, T., Chakrabarti, B. K., Chattopadhyay, D. & Banerjee, A. K. Carboxy-terminal five amino acids of the nucleocapsid protein of vesicular stomatitis virus are required for encapsidation and replication of genome RNA. *Virology* **259**, 219-227 (1999).

82. Takacs, A. M., Das, T. & Banerjee, A. K. Mapping of interacting domains between the nucleocapsid protein and the phosphoprotein of vesicular stomatitis virus by using a two-hybrid system. *Proc Natl Acad Sci U S A* **90**, 10375-10379 (1993).

83. Bilsel, P. A., Rowe, J. E., Fitch, W. M. & Nichol, S. T. Phosphoprotein and nucleocapsid protein evolution of vesicular stomatitis virus New Jersey. *J Virol* **64**, 2498-2504 (1990).

84. Barik, S. & Banerjee, A. K. Cloning and expression of the vesicular stomatitis virus phosphoprotein gene in Escherichia coli: analysis of phosphorylation status versus transcriptional activity. *J Virol* **65**, 1719-1726 (1991).

85. Gao, Y., Greenfield, N. J., Cleverley, D. Z. & Lenard, J. The transcriptional form of the phosphoprotein of vesicular stomatitis virus is a trimer: structure and stability. *Biochemistry* **35**, 14569-14573 (1996).

86. Pattnaik, A. K. et al. Phosphorylation within the amino-terminal acidic domain I of the phosphoprotein of vesicular stomatitis virus is required for transcription but not for replication. *J Virol* **71**, 8167-8175 (1997).

87. Abraham, G., Rhodes, D. P. & Banerjee, A. K. The 5' terminal structure of the methylated mRNA synthesized in vitro by vesicular stomatitis virus. *Cell* **5**, 51-58 (1975).

88. Hercyk, N., Horikami, S. M. & Moyer, S. A. The vesicular stomatitis virus L protein possesses the mRNA methyltransferase activities. *Virology* **163**, 222-225 (1988).

89. Schubert, M., Keene, J. D., Herman, R. C. & Lazzarini, R. A. Site on the vesicular stomatitis virus genome specifying polyadenylation and the end of the L gene mRNA. *J Virol* **34**, 550-559 (1980).

90. Doms, R. W., Keller, D. S., Helenius, A. & Balch, W. E. Role for adenosine triphosphate in regulating the assembly and transport of vesicular stomatitis virus G protein trimers. *J Cell Biol* **105**, 1957-1969 (1987).

91. Florkiewicz, R. & Rose, J. A cell line expressing vesicular stomatitis virus glycoprotein fuses at low pH. *Science* **225**, 721-723 (1984).

92. Fredericksen, B. L. & Whitt, M. A. Attenuation of recombinant vesicular stomatitis viruses encoding mutant glycoproteins demonstrate a critical role for maintaining a high pH threshold for membrane fusion in viral fitness. *Virology* **240**, 349-358 (1998).

93. Matlin, K. S., Reggio, H., Helenius, A. & Simons, K. Pathway of vesicular stomatitis virus entry leading to infection. *J Mol Biol* **156**, 609-631 (1982).

94. Field, B. N. et al. *Fields Virology (2-Volume Set)* (Lippincott Williams & Wilkins, 1996).

95. Kaptur, P. E., McKenzie, M. O., Wertz, G. W. & Lyles, D. S. Assembly functions of vesicular stomatitis virus matrix protein are not disrupted by mutations at major sites of phosphorylation. *Virology* **206**, 894-903 (1995).

96. Balachandran, S. & Barber, G. N. Vesicular stomatitis virus (VSV) therapy of tumors. *IUBMB Life* **50**, 135-138 (2000).

97. Roberts, A., Buonocore, L., Price, R., Forman, J. & Rose, J. K. Attenuated vesicular stomatitis viruses as vaccine vectors. *J Virol* **73**, 3723-3732 (1999).

98. Schnell, M. J., Buonocore, L., Whitt, M. A. & Rose, J. K. The minimal conserved transcription stop-start signal promotes stable expression of a foreign gene in vesicular stomatitis virus. *J Virol* **70**, 2318-2323 (1996).

99. Haglund, K., Forman, J., Krausslich, H. G. & Rose, J. K. Expression of human immunodeficiency virus type 1 Gag protein precursor and envelope proteins from a vesicular stomatitis virus recombinant: high-level production of virus-like particles containing HIV envelope. *Virology* **268**, 112-121 (2000).

100. Thomsen, A. Cooperation of B cells and T cells is required for survival of mice infected with vesicular stomatitis virus. *Int Immunol* **9**, 1757-1766 (1997).

101. Isaacs, A. & Lidnenmann, J. Virus interference. I. The interferon. *Proc R Soc Lond B Biol Sci* **147**, 258-267 (1957).

102. Stark, G. R., Kerr, I. M., Williams, B. R., Silverman, R. H. & Schreiber, R. D. How cells respond to interferons. *Annu Rev Biochem* **67**, 227-264 (1998).

103. Haque, S. J. & Williams, B. R. Signal transduction in the interferon system. *Semin Oncol* **25**, 14-22 (1998).

104. Darnell, J. E. J. Studies of IFN-induced transcriptional activation uncover the Jak-Stat pathway. *J Interferon Cytokine Res* **18**, 549-554 (1998).

105. Barber, G. N. Host defense, viruses and apoptosis. *Cell Death Differ* **8**, 113-126 (2001).

106. Muller, U. et al. Functional role of type I and type II interferons in antiviral defense. *Science* **264**, 1918-1921 (1994).

107. Enninga, J., Levy, D. E., Blobel, G. & Fontoura, B. M. Role of nucleoporin induction in releasing an mRNA nuclear export block. *Science* **295**, 1523-1525 (2002).

108. Desforges, M. Different host-cell shutoff strategies related to the matrix protein lead to persistence of vesicular stomatitis virus mutants on fibroblast cells. *Virus Research* **76**, 87-102 (2001).

109. Lamberti, A. et al. The translation elongation factor 1A in tumorigenesis, signal transduction and apoptosis: review article. *Amino Acids* **26**, 443-448 (2004).

110. Holland, E. C., Sonenberg, N., Pandolfi, P. P. & Thomas, G. Signaling control of mRNA translation in cancer pathogenesis. *Oncogene* **23**, 3138-3144 (2004).

111. Balachandran, S. & Barber, G. N. Defective translational control facilitates vesicular stomatitis virus oncolysis. *Cancer Cell* **5**, 51-65 (2004).

112. Browne, G. J. & Proud, C. G. Regulation of peptide-chain elongation in mammalian cells. *Eur J Biochem* **269**, 5360-5368 (2002).

113. Gadaleta, P., Perfetti, X., Mersich, S. & Coulombie, F. Early activation of the mitochondrial apoptotic pathway in Vesicular Stomatitis virus-infected cells. *Virus Res* **109**, 65-69 (2005).

114. Li, P. et al. Cytochrome c and dATP-Dependent Formation of Apaf-1/Caspase-9 Complex Initiates an Apoptotic Protease Cascade. *Cell* **91**, 479-489 (1997).

115. Scorrano, L. Mechanisms of cytochrome c release by proapoptotic BCL-2 family members. *Biochem Biophys Res Comm* **304**, 437-444 (2003).

116. Gaddy, D. F. & Lyles, D. S. Vesicular stomatitis viruses expressing wild-type or mutant M proteins activate apoptosis through distinct pathways. *J Virol* **79**, 4170-4179 (2005).

117. Sharif-Askari, E. et al. Bax-dependent mitochondrial membrane permeabilization enhances IRF3-mediated innate immune response during VSV infection. *Virology* **365**, 20-33 (2007).

118. Fernandez, M., Porosnicu, M., Markovic, D. & Barber, G. N. Genetically engineered vesicular stomatitis virus in gene therapy: application for treatment of malignant disease. *J Virol* **76**, 895-904 (2002).

119. Obuchi, M., Fernandez, M. & Barber, G. N. Development of Recombinant Vesicular Stomatitis Viruses That Exploit Defects in Host Defense To Augment Specific Oncolytic Activity. *J Virol* **77**, 8843-8856 (2003).

120. Samuel, S. et al. VSV oncolysis in combination with the BCL-2 inhibitor obatoclax overcomes apoptosis resistance in chronic lymphocytic leukemia. *Mol Ther* **18**, 2094-2103 (2010).

121. Tumilasci, V. F. et al. Targeting the apoptotic pathway with BCL-2 inhibitors sensitizes primary chronic lymphocytic leukemia cells to vesicular stomatitis virus-induced oncolysis. *J Virol* **82**, 8487-8499 (2008).

122. Brun, J. et al. Identification of genetically modified Maraba virus as an oncolytic rhabdovirus. *Mol Ther* **18**, 1440-1449 (2010).

123. Rots, M. G., Curiel, D. T., Gerritsen, W. R. & Haisma, H. J. Targeted cancer gene therapy: the flexibility of adenoviral gene therapy vectors. *J Control Release* **87**, 159-165 (2003).

124. Cesaire, R. et al. Oncolytic activity of vesicular stomatitis virus in primary adult T-cell leukemia. *Oncogene* **25**, 349-358 (2006).

125. Leveille, S., Samuel, S., Goulet, M. L. & Hiscott, J. Enhancing VSV oncolytic activity with an improved cytosine deaminase suicide gene strategy. *Cancer Gene Ther* (2011).

126. Lozzio, C. B. & Lozzio, B. B. Human chronic myelogenous leukemia cell-line with positive Philadelphia chromosome. *Blood* **45**, 321-334 (1975).

CHAPTER 2

Experimental

Reagents

HPLC-grade solvents (acetonitrile, 0.1% formic acid in water, 0.1% formic acid in acetonitrile) were purchased from Fisher Scientific (Waltham, MA). All other reagents were purchased from Sigma-Aldrich (St. Louis, MO) or Bioshop (Burlington, Ontario) unless otherwise noted.

Cell Culturing

K562 human myeloid leukemia cells were grown in Iscove's Modified Dulbecco's Medium (IMDM) containing 10% fetal bovine serum (FBS) and penicillin-streptomycin (100 IU/mL) and to a density of 1.0×10^6 cells/mL in a total of 200 mL at 37°C. The cells were split into two aliquots and centrifuged at 1000 rpm for 5 minutes and supernatant discarded. Treated cells were resuspended in 2 mL IMDM, infected with vesicular stomatitis virus (VSV) at a multiplicity of infection (MOI) 10 PFU/cell on ice and maintained on ice for 1 hr alongside a control. Cells were diluted to 10 mL with IMDM containing 10% (FBS), incubated at 37°C and harvested after 30 minutes. Samples were centrifuged at 1000 rpm for 5 minutes at 4°C, supernatant discarded and pellet washed with phosphate buffered saline (PBS). Cell pellets were frozen in liquid nitrogen and stored at -80°C.

For the background proteomic analysis, K562 was grown in the same manner and harvested without treatment.

Cell Lysis and Digestion

Cells were resuspended in chilled lysis buffer consisting of 8 M urea, 75 mM NaCl, 1 tablet protease inhibitor cocktail (Complete Mini, Roche, Basel, Switzerland) per 10 mL and 1 tablet phosphatase inhibitor cocktail (PhosSTOP, Roche) per 10 mL. For every 5×10^7 cells, 1 mL of lysis buffer was used. Cells were placed on ice and lysed by sonication for 5×60s at 50% power with 2 min intervals using a Vibra-Cell VCX 130-Watt sonicator (Sonics, Newton, CT). Lysates were centrifuged at 10,000*g* at 4°C for 20 minutes to remove cell debris. Protein concentrations in the lysates were quantified using a BCA assay (Sigma-Aldrich, St. Louis, MO) and a 1:10 dilution of the sample to give concentrations of 5.87 μg/μL (virus-infected) and 6.32 μg/μL (control).

Aliquots containing 1 mg total protein of the virus-infected and control samples were reduced by addition of dithiothreitol (DTT) to a concentration of 5 mM and incubated at 56°C for 25 minutes. Cysteine residues were then alkylated by addition of iodoacetamide to a concentration of 14 mM and incubated at room temperature and in darkness for 30 minutes, followed by addition of 5 mM DTT and incubation for 15 minutes at the same conditions to destroy any remaining iodoacetamide.

Samples were then diluted 1:5 with 25 mM Tris, pH 8.2. Calcium chloride was added to a concentration of 1 mM and sequencing-grade trypsin was added to 1:25 w/w enzyme:substrate and incubated overnight at 37°C.

The background K562 samples were prepared similarly but with several differences. Lysis buffer was removed by buffer exchange into 25 mM Tris, pH 8.2 using Centricon 3000 Da MWCO filters (Millipore, Billerica, MA) and centrifuging at $5\times14{,}000g$ for 20 min. The BCA assay gave a concentration of 7.00 μg/μL and trypsin was added to w/w 1:50 enzyme:substrate.

Dimethyl Labeling

Peptides from 30-min virus-infected cells were passed through Bond Elut 500 mg C_{18} cartridges (Agilent Technologies, Santa Clara, CA) and desalted with HPLC-grade 0.1% formic acid in water. A volume of 600 μL light dimethyl labeling reagent containing 100 mM sodium acetate (pH 5), 60 mM sodium cyanoborohydride, 0.4% formaldehyde was passed through the column to label N-termini and lysine residues, followed by desalting with 0.1% formic acid. Peptides from the control cells were then bound to the cartridge and desalted. They were labeled in the same manner with heavy dimethyl labeling reagent containing 100 mM sodium acetate (pH 5), 60 mM sodium cyanoborohydride, 0.4% formaldehyde-d_2 (Cambridge Isotopes, Andover, MA), followed by desalting with 0.1% formic acid. The resulting labeled and mixed peptides were eluted with 50% acetonitrile/0.1% formic acid and evaporated to dryness.

Immobilized Metal Ion Affinity Chromatography (IMAC)

Phosphopeptides were enriched using Phos-Select Iron Affinity Gel (Sigma-Aldrich) using the manufacturer's instructions with a few modifications. Peptides were resuspended in 500 μL of IMAC binding buffer (40% acetonitrile/60%

water/0.1% formic acid) and fully solubilized in a sonicator bath. IMAC beads (10 µL) were washed three times with IMAC binding buffer and incubated with the sample for 1 hr. The sample was centrifuged (15,000*g*, 2 min) and the supernatant was collected and kept as the non-phosphorylated peptide-containing flow-through (FT). The resin was washed three times with IMAC binding buffer, added to the FT fraction, and frozen at -20°C for future use. Phosphopeptides were eluted from the resin by incubation for 5 min with 40 µL K_2HPO_4/NH_4OH (pH 10), centrifugation, and collection of supernatant. The elution was repeated twice. The combined phosphopeptide elution fractions were then acidified by addition of 10% formic acid (40 µL) prior to LC-MS/MS analysis.

Strong Cation-Exchange/Strong Anion-Exchange Chromatography (SCX/SAX)

The non-phosphorylated FT fraction was evaporated to near-dryness to remove acetonitrile. Phosphoric acid (H_3PO_4) was added to a concentration of 50 mM. Columns were packed with 12 µm silica-based SCX (Canadian Life Sciences, Peterborough, Ontario) or polymer-based SAX (Polymer Laboratories, Varian Inc.) beads with 300 Å pore size to a length of 13 cm in a 700 µm i.d. column with a fritted union (IDEX, Oak Harbor, WA). For SCX, the sample (250 µg peptide aliquot) was passed through the column by a pressure vessel (see Figure 2.1) while keeping the FT fraction, and washed with 20% acetonitrile/8 mM potassium phosphate buffer (KPB, pH 3) to remove any non-peptidic material. The sample was fractionated step-wise by sequential elutions (100 µL) of increasing pH; in order: 3.5, 4.0, 4.5, 5.0, 5.5, 6.0, 7.0, 8.0. Ammonium bicarbonate (800 mM ABC) was used

as the final elution buffer for the background K562 proteome. SAX was performed in a similar fashion, except the sample (250 μg peptides) was made alkaline by addition of ammonium hydroxide (NH_4OH) to 150 mM prior to binding, the wash buffer was 50 mM NH_4OH/20% acetonitrile, and the pH elution buffer order was: 8.0, 7.0, 6.0, 5.5, 5.0, 4.5, 3.5, 3.0, 2.5.

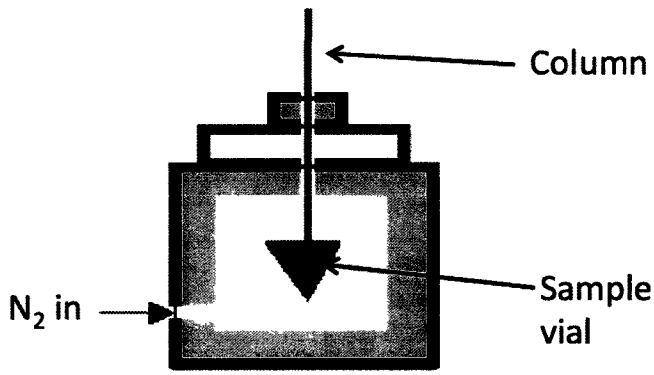


Figure 2.1. Pressure vessel setup used for flow of liquids (samples, reagents, washes) through a column. Typical N_2 pressure range 200-1000 psi depending on column type, length, and inner diameter.

LC-MS/MS Analysis

Either non-phosphorylated samples (1:10 dilution) or phosphorylated (no dilution) fractions were loaded using a pressure vessel onto a 200 μm × 5 cm pre-column fritted in-house with 88 μL Kasil/16 μL formamide and packed with Magic C18AQ (Michrom Bioresources, Auburn, CA). Samples were desalted with HPLC-grade water/0.1% formic acid and placed online with an HP 1090 HPLC (Agilent) and PicoFrit (New Objective, Woburn, MA) 75 μm × 4 cm analytical column packed in-house with Magic C18AQ. Peptides were separated using a gradient of 2% B at 0

min, 5% B at 3 min, 10% B at 8 min, 29% B at 70 min, 80% B at 75 min followed by 10 min of wash at 100% B and 5 min of re-equilibration at 2% B, where A and B are 0.1% formic acid/water and 0.1% formic acid/acetonitrile, respectively.

Peptides were electrosprayed at a voltage of +3,500 V and data were collected on a QStar XL mass spectrometer (AB Sciex, Concord, Ontario) using Analyst QS 2.0 software. Spectra were obtained by 1 survey MS scan of 1 s between m/z 400-1500 Th followed by 4 product ion scans of 3 s of ions of charge >+1. Precursor ions were placed on an exclusion list for 90 s after being selected.

Peptide Identification and Quantitation

Analyst data files (.wiff format) were converted to Mascot Generic Format files (.mgf format) using the Mascot.dll Analyst Plug-in (v1.6b25). Protein identification was accomplished using Mascot 2.3 (Matrix Science, Boston, MA) and searching against the IPI Human protein database. Trypsin was set as the enzyme with 2 allowed missed cleavages. Carbamidomethylation of cysteine residues was set as a fixed modification, and the following variables modifications were selected: oxidation of methionine residues, light and heavy dimethylation on lysine, peptide N-terminus, and protein N-terminus residues, and phosphorylation on serine, threonine, and tyrosine residues. No dimethyl label modifications were selected for the background proteome. Mass error tolerances were set at 100 ppm for peptide mass spectra and 0.2 Da for MS/MS spectra. Only peptides with charge states of 2+, 3+, and 4+ were selected for MS/MS analysis. A Mascot ions score cutoff of 20 was set for proteomic samples or 15 for phosphoproteomic samples. Spectra were

visually inspected for ≥ 3 consecutive b- or y-ions in a series and their mass error tolerances showed an acceptable linear trend; they were rejected if they did not meet these criteria.

Peptides were quantified manually or by a custom script written in PHP. Manual quantitation was accomplished by obtaining the elution time of a light-labeled peptide identified in Mascot and navigating to this time in the total ion chromatogram (TIC) in Analyst. The m/z of the peptide was found in the MS and the first isotopic peak was highlighted. An extracted ion chromatogram (XIC) of the peak from the MS was taken and the peak in the XIC was then highlighted with a width of approximately 1 minute. The maximum height of this peak was noted and the process was repeated for the corresponding heavy-labeled peptide. The light to heavy ratio (L/H) of the respective peak heights was assumed to correspond to the relative concentration of the peptide from the two original biological states.

A computational script was developed to automate this process. An Analyst data file was converted to the open-source mzML format using ProteoWizard. The elution time of a peptide was noted in Mascot as well as its m/z, label (light or heavy), number of modification sites, and charge. The maximum total intensity of the peak was computed within a 0.25 Th interval and 1 min elution interval. The charge, m/z, and number of modification sites were used to compute the expected m/z of its labeled counterpart (light or heavy), and the maximum total intensity of this m/z was recorded as well. Both the light and heavy peaks must have exceeded an ion count threshold of at least 20 to be considered. This process was iterated

until all Mascot identifications were processed; the resulting data was exported into a tab-delimited format. The upregulated and downregulated peptides identified by this script were manually validated.

Bioinformatics analysis

The phosphorylation sites from the background phosphoproteome were assigned into general categories. 13-amino acid peptide sequences centered about the phosphorylated residue were assigned into proline-directed (PD), acidic (A), or basic (B) categories based on the following decision tree[1]: P at +1 (PD); >4 E/D at +1 to +6 (A); R/K at -3 (B); D/E at +1/+2 or +3 (A); >1 R/K at -1 to -6 (B). If a phosphopeptide did not fit into these categories they were assigned to an "others" category. Specific motifs were searched for using the Motif-X algorithm[2] and abundance-based sequence logos were generated by Weblogo software.[3]

Standard gene names of interest were submitted to the GeneCoDis algorithm to identify functional enrichment of GO annotations.[4,5] The search details were varied according to the gene list under investigation, but all searches were performed with the hypergeometric statistical test, FDR p-value correction, and with GO biological processes, molecular functions, and cellular components as the annotations searched.

References

1. Villen, J., Beausoleil, S. A., Gerber, S. A. & Gygi, S. P. Large-scale phosphorylation analysis of mouse liver. *Proc Natl Acad Sci U S A* **104**, 1488-1493 (2007).

2. Schwartz, D. & Gygi, S. P. An iterative statistical approach to the identification of protein phosphorylation motifs from large-scale data sets. *Nat Biotechnol* **23**, 1391-1398 (2005).

3. Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res* **14**, 1188-1190 (2004).

4. Nogales-Cadenas, R. et al. GeneCodis: interpreting gene lists through enrichment analysis and integration of diverse biological information. *Nucleic Acids Res* **37**, W317-22 (2009).

5. 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).

CHAPTER 3

Results and discussion

3.1 Analytical method development

3.1.1 Overview of the workflow

A quantitative workflow for mass spectrometry-based shotgun proteomics was developed and is illustrated schematically in Figure 3.1. A biological sample in two distinct states is chosen – in this case VSV-infected K562 cells and control K562 cells. The cells are lysed and a total protein quantitation assay is done (BCA). An equal amount of proteins from both samples are tryptically digested and the samples are sequentially bound to a reverse-phase resin, desalted, and labeled with light or heavy dimethyl reagents. The samples are combined and IMAC is performed to obtain a phosphorylated and non-phosphorylated fraction. The phosphorylated fraction is directly analyzed by RPLC-MS/MS on a hybrid quadrupole-time-of-flight mass spectrometer. The non-phosphorylated fraction is split, fractionated by SCX and SAX into 6 fractions each, and analyzed in the same way by mass spectrometry. The MS/MS spectra are used to identify the peptides by spectral database searching and the MS are used for quantitation. Of the steps shown in Figure 3.1, the development of appropriate lysis conditions, enzymatic digestion, labeling techniques, IMAC enrichment, SCX/SAX fractionation, and quantitation methods will be discussed. An analysis of the background K562 proteome without any dimethyl labeling or quantitation steps will be used to evaluate the effects of dimethylation on the behavior of peptides in this workflow.

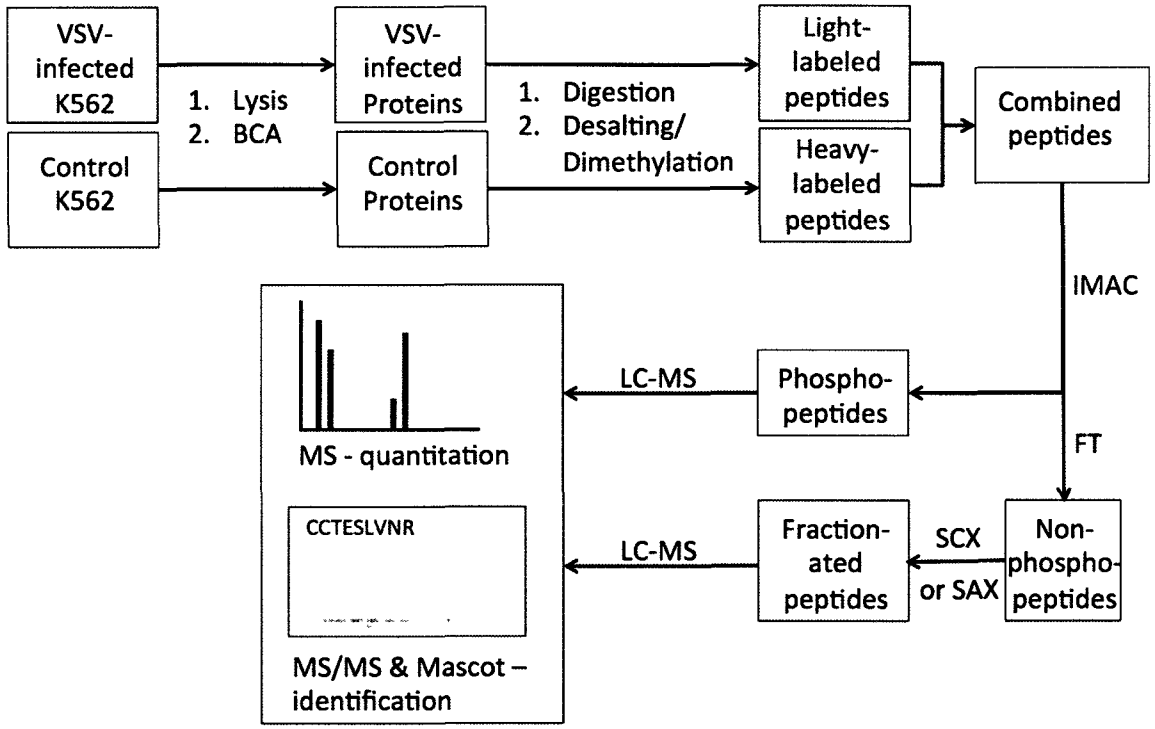


Figure 3.1. Overview of proteomic and phosphoproteomic approach to analyzing VSV-infected K562 cells.

3.1.2 On-column versus in-solution digestion

A protein sample can be tryptically digested by addition of an enzyme to a buffered solution containing the proteins, termed in-solution digestion, or digested after being immobilized onto a chromatographic resin such as SCX, termed on-column digestion. On-column digestion involves acidifying a protein sample to a pH of ~2, trypsinizing at a 1:5 trypsin:substrate ratio and binding to a SCX column. The sample is washed on-column prior to reduction by infusion of DTT, followed by concurrent alkylation and digestion by infusion of iodoacetamide and Tris buffer at pH 8.

It was found that this method had advantages and drawbacks over conventional in-solution tryptic digestion. By adding 20% acetonitrile to the wash buffer, non-proteinogenic material that may cause interference, such as DNA and other nucleic acid species, are removed. This is particularly useful when a phosphoproteomic analysis is being done, as phosphate species will competitively bind along with phosphopeptide species to IMAC resins. On-column digestion also has a speed advantage over in-solution digestion, as full digestion can be completed in 90 minutes rather than the conventional overnight time frame. This is likely due to the smaller volume inside the column and the semi-immobilization of the trypsin. Note that immobilization of trypsin vastly increases reaction kinetics; recent work has increased sample digestion speeds by immobilized trypsin to < 1 minute.[1] Recent evidence suggests peptides can be fractionated by sequential pH elutions directly from the column used for the digest,[2] however it has been shown that this process results in peptides being spread over many pH fractions and is therefore unsuitable for quantitative analyses (data not shown).

On-column digestions do require a large amount of trypsin – 1:5 rather than 1:25 or 1:50 trypsin:substrate for in-solution digestions in this work, or 1:250 reported in other protocols.[3] The analysis of complex samples also necessitates starting with larger quantities of total protein, and therefore an increase in the required amount of chromatographic resin. For 500 μg of starting material, a column size of 13 cm length and 700 μm inner diameter is needed. Due to potential sample contamination and carryover, columns used for digestion are not re-used. Perhaps the biggest drawback for on-column digestion is the associated sample

losses. An experiment was conducted by directly comparing the number of identified peptides by mass spectrometry doing in-solution versus on-column digestion using an equal amount (1 pmol) of protein standards (hemoglobin, myoglobin, albumin, ovalbumin, alpha-casein, beta-casein, concanavalin a). As shown in Figure 3.2, in-solution digestion yielded far more tryptic peptide identifications by Mascot than on-column digestion – nearly 50% more. There could be several possible explanations for this. The use of a digestion buffer at pH 8 means that peptides could be lost while infusing the column with Tris as this pH is well above the pI for many peptides and proteins. Alternatively, peptides and/or proteins could exhibit non-specific binding to the silica support of the SCX resin following digestion, a phenomenon supported both by the literature[4] and by anecdotal evidence.

Due to the non-economical nature of on-column tryptic digestions and associated sample losses, in-solution digestion was found to be a superior method for proteolysis of samples.

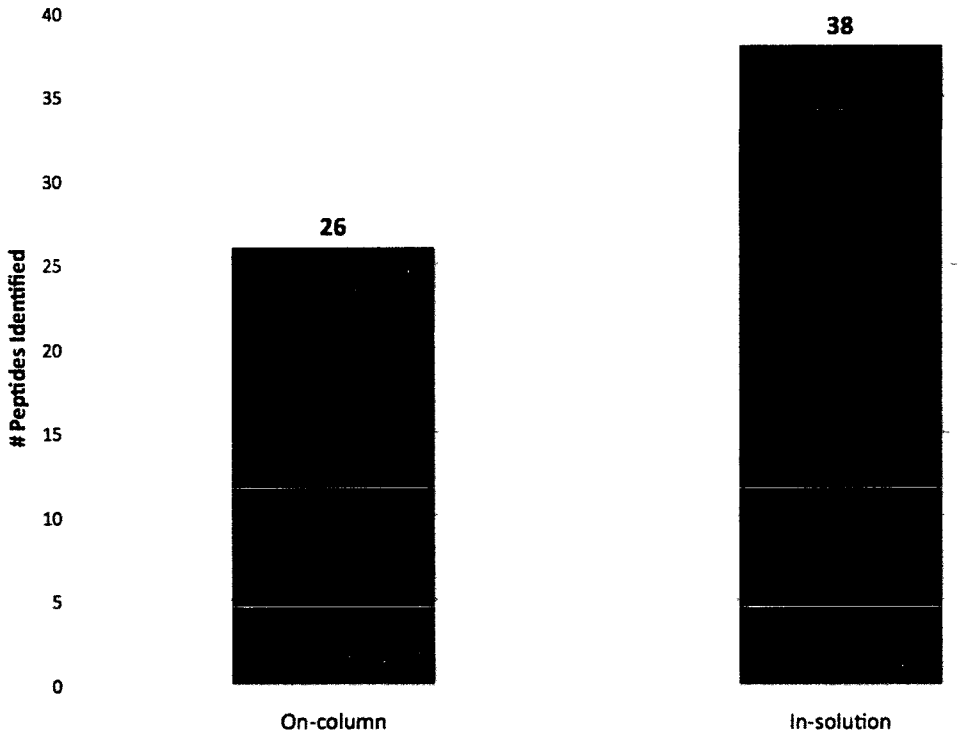


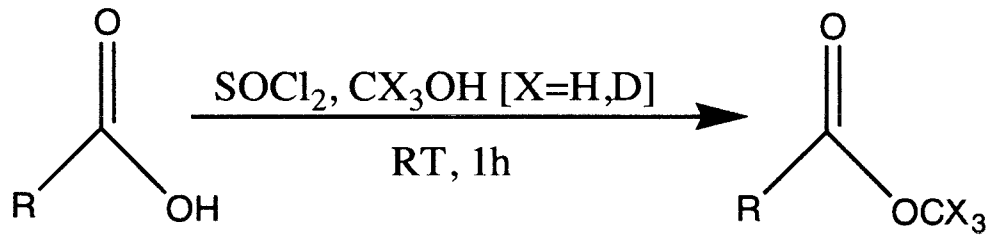
Figure 3.2. Number of identified peptides by LC-MS/MS and Mascot using 1 pmol total protein standards tryptically digested on-column or in-solution.

3.1.3 Labeling strategies for peptide quantitation

A stable-isotope MS-based quantitation strategy was chosen to be developed over MS/MS-based or label-free strategies for the reasons described in 1.3.2. Furthermore, MS-based strategies can be adopted that use inexpensive reagents and that are applicable to any sample.

The first strategy evaluated was methyl esterification by methanolic hydrochloric acid (MeHCl). MeHCl is formed by the addition of thionyl chloride to methanol in a ratio of 1:25. This solution can react with the C-terminal end and aspartic acid and glutamic acid residues of peptides, replacing the hydroxyl group

with a methyl ester, as shown in Scheme 3.1. The use of deuterated methanol will create a mass shift of +3 Da for every labeled residue in a peptide relative to its non-deuterated counterpart – thus, two samples can be differentially labeled, run simultaneously, and quantitated by mass spectrometry.



Scheme 3.1. Methyl esterification of carboxylic acid groups. R designates C-termini, glutamic acid, or aspartic acid residues.

Conventionally, the chemistry is done in solution by the addition of MeHCl to a dried sample and incubation for 1 hour at room temperature, followed by lyophilization to dryness.[5] Methyl esterification also eliminates acidic groups from aspartic acid and glutamic acid residues, making it ideal for phosphoproteomic analyses as it has been observed that the elimination of these acidic groups decreases the binding of non-phosphorylated peptides to IMAC resins.[6, 7]

As a microfluidic methodology was desired, the use of methyl esterification in conjunction with chromatographic resins was evaluated using protein standards. Reverse-phase columns are not compatible, as peptides will be eluted immediately from the column with methanol; therefore on-column labeling was attempted on SCX columns. Tryptic peptides from protein standards were acidified to pH ~2 and

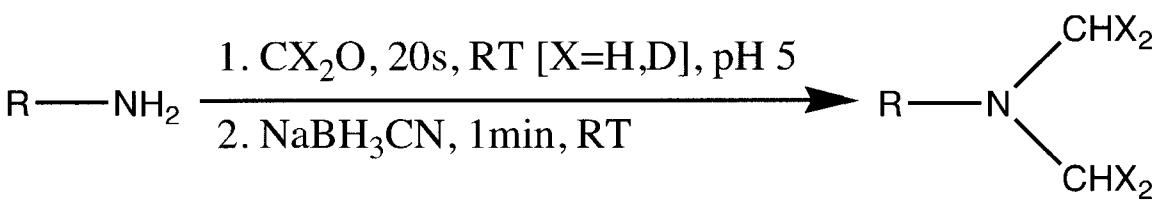
bound to a SCX column (5 cm length, 200 μm inner diameter). MeHCl (100 μL) was then passed through the column, followed by washing (to remove residual MeHCl) and salt elution. No peptides were found in the wash or elution fractions, but 24 peptides were found in the flow-through from the labeling. This indicates that the MeHCl reagent elutes the peptides from the column, making it an unsuitable choice for on-column labeling. This is possibly due to the very strong acid protonating the polysulfoethyl aspartamide functional groups and thus eluting the peptides from the column. Alternating MeHCl infusion/evaporation steps inside the column showed limited success but was considered to not be feasible as a general method.

The methyl esterification method showed problems as an in-vial quantitation as well. There were instances of incomplete esterification, even after a 2-hour reaction time; for instance, the alpha casein peptide shown in Table 3.1 contained both esterified and non-esterified forms. There were also instances of deamidation on the glutamine (Q) and asparagine (N) residues of some peptides (Table 3.1). These incomplete and side-reactions make methyl esterification not ideal as a quantitative labeling method.

Table 3.1. Example of peptides identified that show partial methyl esterification and deamidation on glutamine and asparagine residues.

Protein	Score	Peptide
Hemoglobin	30	K.HLVDEPQNLIK.Q + Methyl (C-term); 2 Methyl (DE); Methyl+Deamidated (Q)
	35	K.HLVDEPQNLIK.Q + Methyl (C-term); 2 Methyl (DE); Methyl+Deamidated (N); Methyl+Deamidated (Q)
Alpha casein	64	R.FFVAPFPEVFGK.E + Methyl (C-term)
	39	R.FFVAPFPEVFGK.E + Methyl (C-term); Methyl (DE)

A second method was developed based on the work of Hsu and colleagues termed dimethyl labeling.[8] The method requires sodium cyanoborohydride, formaldehyde, and an appropriate buffer to maintain pH. The in-solution method described originally is shown in Scheme 3.2. It involves the reductive amination of formaldehyde with an amine from the N-terminus or lysine residue of a peptide.[9] As this reaction can occur twice per amine group, the use of deuterated formaldehyde gives a mass shift of +4 Da compared to its non-deuterated counterpart.



Scheme 3.2. In-solution dimethyl labeling of amine groups. R denotes N-termini of peptide or lysine residues.

The adaptation of this reaction to on-column labeling was promising as the reagents are aqueous and have relatively low salt concentrations, making both reverse-phase and SCX columns possible. The reaction, however, requires a pH between 5 and 6, thereby eluting peptides with a low isoelectric point – tests using peptides from 7 protein standards showed 12 unique, high-scoring peptides in the flow-through fraction. The chemistry was tested using a lower pH but it was found the reaction did not go to completion at pH < 4, particularly on lysine residues. Furthermore, it is more useful to combine this labeling with a peptide desalting step, so C_{18} was the preferred resin on which to optimize on-column labeling.

The on-column labeling was tested on a C_{18} column (5 cm length, 200 μm inner diameter) using tryptic peptides from the same 7 protein standards. As this reaction is typically a one-step reaction, it was hypothesized that reacting with formaldehyde followed by sodium cyanoborohydride was unnecessary – the steps should be able to be performed simultaneously. Adaptation from the in-vial protocol suggested a solution of 30 mM cyanoborohydride/0.2% formaldehyde in 110 μL sodium acetate (100 mM, pH 5). However, peptides did not undergo full dimethylation, nor did an increase in reagent volume affect reaction completion. A doubling of reagent concentration (60 mM sodium cyanoborohydride, 0.4% formaldehyde), however, resulted in full dimethylation on all identified peptides.

The process for dimethyl labeling of two species is then as follows: sample 1 binding, flow of light reagents, desalting, sample 2 binding, flow of heavy reagents, desalting, elution. This process has been successfully adapted for large-scale

proteomic samples: on-column labeling was successfully applied to 2 x 1 mg of cell lysate on a 500 mg Bond Elut C_{18} column (*vide infra*).

3.1.4 Lysis conditions

Some points of discussion should be made regarding the cell lysis steps, as this was a considerable point of difficulty in the developed strategy. Protocols for on-column tryptic digests suggest a detergent lysis is possible, for example with RIPA (radioimmunoprecipitation assay) buffer. This lysis technique is relatively soft and simple, as no mechanical homogenization is performed, and is adequate for K562 cells since they do not have strong cell walls. As the detergents interfere with tryptic activity, they are removed by washing the SCX column with 20% acetonitrile. In-solution digests require an alternate method of removing the detergents from the RIPA buffer, as SCX is not done prior to digestion. Buffer-exchange, molecular-weight cutoff filters (MWCO 3,000 Da) were used prior to digestion to remove all detergents from the lysate. It was found that, even after repeated buffer-exchange (up to 10 repetitions), tryptic digestion failed, evidenced by the formation of a large protein pellet after post-digestion acidification. It is not entirely clear why these buffer-exchange filters did not remove the detergents; it is possible that the detergent polymers used in RIPA exceeded the molecular weight cutoff of 3,000 Da.

An alternative strategy was therefore developed, adapted from a protocol published by Gygi.[3] Briefly, cells were resuspended in a denaturing buffer containing 8 M urea and lysis was performed by the use of a sonicator, which disrupts cell membranes

by ultrasonic pulses. While the sonication is more labor-intensive than detergent lysis, detergents are avoided entirely and the lysis buffer is compatible with tryptic digestion after dilution to a urea concentration of 1.6 M or less.

3.1.5 Peptide fractionation

Unlabeled versus dimethylated peptides

As the dimethylation labeling process replaces all basic amine groups from N-termini and lysine residues, it was predicted that this should change the elution properties of the peptides from SCX resins relative to their non-labeled counterparts. This was investigated by plotting the percentage of peptides identified in each elution fraction against the elution conditions for each fraction, as shown in Figure 3.3. Note the labeled fractions at pH 5.5 and 6 are interpolated for illustrative purposes as these fractions were not analyzed. As expected, the unlabeled peptides show a broad distribution, centered about an approximately neutral pH (6-7). The 800 mM ABC fraction contains the largest percentage of peptides as it was the only buffer used to elute basic peptides; thus it is the sum of all peptides with an isoelectric point greater than 7. The broad distribution and tailings at the extremities indicates good separation of peptides, with an approximate equal number distributed into every fraction, except those that are extremely acidic (and likely basic at the other end) due to the relative lack of peptides with very high or low isoelectric points. The labeled peptides differ in that they show a broad, approximately normal distribution centered about a low pH (5), as shown in grey in

Figure 3.3. This is due to the replacement of the basic groups with neutral ones and thus the reduction of the isoelectric points of all peptides. The effects of this lowering in isoelectric point can be clearly seen: there is a clear shift of the distribution towards a low pH after dimethyl labeling, and there is a nearly complete elimination of any peptides eluting at a basic pH. In this work, since pH elution was done primarily under acidic conditions, the dimethyl labeling in fact improved separation, as the entire normal distribution falls within the pH range of the elution buffers used. Furthermore, very low pH fractions (3.5-4) contain labeled peptides whereas these fractions do not contain any non-dimethylated counterparts (Figure 3.3). Similarly, there are many basic peptides present in the single elution above a neutral pH and thus these peptides are not adequately fractionated.

While sample fractionation benefited from dimethyl labeling in this experiment, the effects of dimethyl labeling are not universally desirable, as many experiments would use a wider pH range for SCX fractionation. The development of polymer-based rather than silica-based beads has allowed SCX resins to resist silica-functional group bond cleavage and they can therefore tolerate higher pHs. Thus, as the dimethyl labeling compresses the isoelectric points of all peptides into a narrower range, the 2nd dimension of peptide separation (SCX or SAX) would yield some fractions that are very complex and some that are nearly devoid of peptides. As a result, dimethyl labeling would result in the SCX/SAX to be reduced into half its original efficiency.

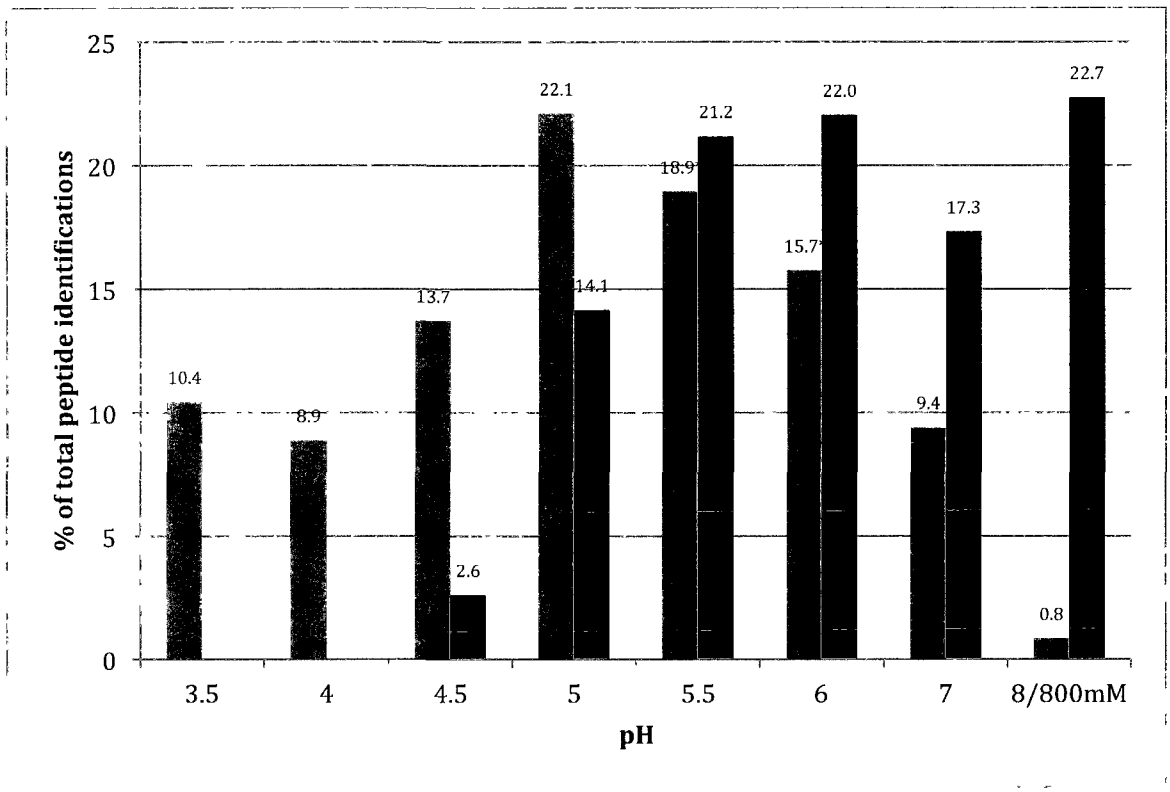


Figure 3.3. Effect of dimethylation on SCX fractionation. Labeled (grey) and unlabeled (black) peptides are shown as a percentage of the total identifications. Numbers marked with an asterisk (*) were interpolated and the final elution for labeled and unlabeled peptides was done with pH 8 buffer and 800 mM ABC, respectively.

The reduced isoelectric point after dimethyl labeling also has a substantial effect on phosphopeptides. Attempts at fractionating enriched phosphopeptide samples by SCX were unsuccessful as none of the elution fractions contained phosphopeptides; all phosphopeptides were contained in the SCX flow-through fraction. Due to the removal of basic groups and the abundance of acidic residues (aspartic, glutamic acid) as well as the very strongly acidic phosphate group, the isoelectric points of the phosphopeptides were below the pH of the acidified phosphoproteomic sample (50 mM H_3PO_4, pH ~2). It is likely that SAX would be better for separation of phosphoproteomic samples based on the low isoelectric

points of the phosphopeptides, however, the lack of diversity would likely make it impossible to fractionate the phosphopeptides nonetheless. Therefore, while useful for some proteomics samples, the developed method is perhaps not adequately robust for generalization to all large-scale proteomic and phosphoproteomic analyses.

SAX versus SCX fractionation

Three total datasets from fractionated samples were obtained and each contained 6 individual fractions. Therefore, the separation efficiency of the different methods can be estimated from the number of fractions that each peptide is identified within. A histogram from all 6 fractions of the unlabeled, SCX-fractionated K562 background proteome and from the labeled, SCX- or SAX-fractionated K562 proteome was constructed. Duplicate peptides were removed and the number occurring in one, two, three, etc. fractions were counted. As the majority of peptides are contained within one or two fractions, a logarithmic plot was made of number of peptides versus number of fractions in which the peptides appear, shown in Figure 3.4. This evidence shows that SCX fractionation of the unlabeled, background proteome showed the smallest amount of overlap between fractions and therefore the highest separation efficiency. This is likely since this group is the only one in which peptides still contain their basic amino groups, thereby imparting a greater diversity in isoelectric points. This is in contrast to the conclusion that separation is improved in the acidic region after peptides are dimethylated; perhaps while there

are more identifications, the actual separation efficiency has decreased due to the decrease in isoelectric point diversity.

Figure 3.4 also shows that, in a direct comparison of SAX to SCX on dimethylated peptides, SAX outperforms SCX in separation efficiency. SAX fractionation resulted in more peptides that occur in only 1 fraction and fewer that occur in 2 fractions when compared with SCX fractionation. This may be due to the more acidic properties of dimethylated peptides and their resulting affinity for SAX (*vide infra*). Interestingly, the improved separation efficiency of SAX over SCX has been previously observed for the separation of phosphopeptides[10]; this data shows that the phenomenon can be extended to other low isoelectric point peptide populations such as dimethylated peptides.

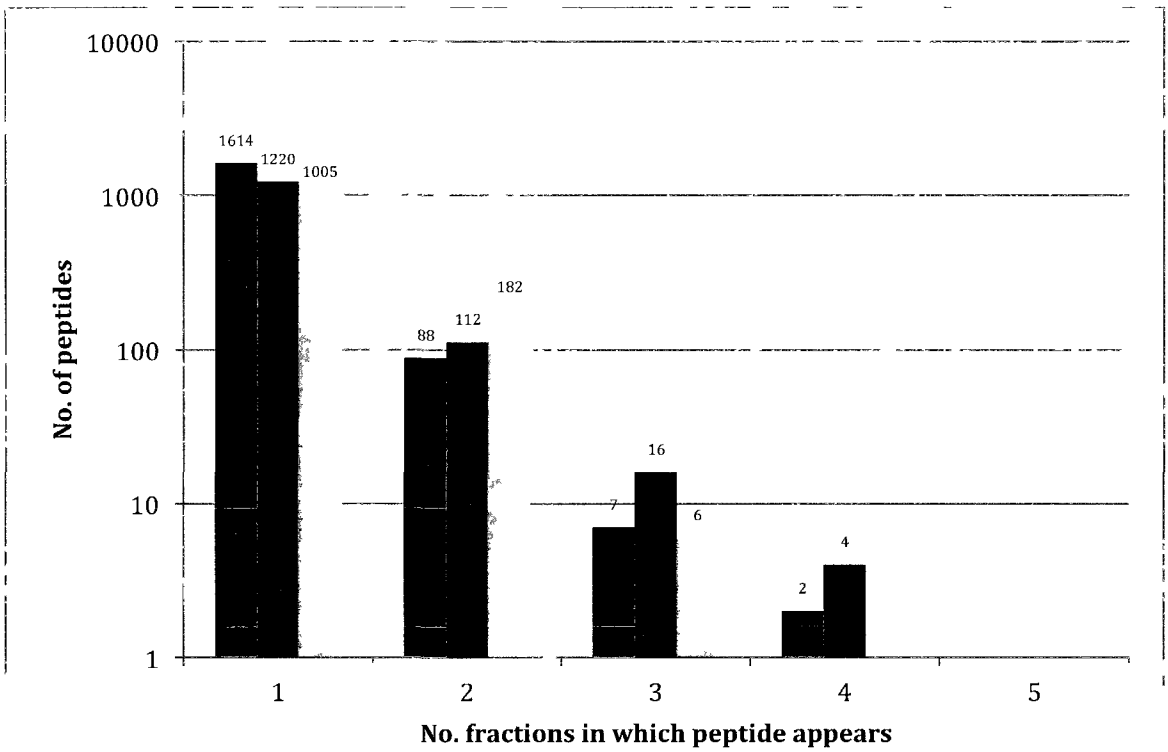


Figure 3.4. Histogram showing resolution of fractionation methods. Shown are frequency of peptides, identified from 6 SCX fractions of background K562 proteome (black), or from 6 SAX (dark grey) or SCX (light grey) fractions in labeled, VSV-infected K562, occurring in one or more fractions.

To evaluate whether the SAX and SCX fractionation identified similar peptides, two datasets were formed: one from all 6 SCX fractions and one from all 6 SAX fractions. Duplicate peptides were removed and the overlap between the two datasets was assessed; the resulting diagram is shown in Figure 3.5. The overlap was found to be very small (<10%), suggesting these separation techniques are complementary and allow analyses of distinct sections of the proteome. As these techniques separate based on isoelectric point, the differences must be related to charge generation, and therefore the acidic/basic residues on a peptide. As all tryptic peptides contain only one basic residue (K/R) at the C-terminal end of the

peptide, basic residues were thought to not play a role in the peptide separation. To verify this theory, the average number of missed cleavages was calculated for the SAX and SCX datasets and the values were found to be comparable (0.3 and 0.4 missed cleavages per peptide for SCX and SAX, respectively). The other option is that the differences in identifications between SAX and SCX are due to the presence of acidic groups in the identified peptides; this was investigated by calculating the frequency of acidic residues (D/E). It was found that aspartic acid had a frequency of 10.0% and 6.9% and glutamic acid had a frequency of 15.5% and 7.7% in SAX and SCX datasets, respectively. This gives a total acidic residue frequency of 25.5% in SAX and only 14.5% in SCX, and since the average length of a peptide is greater than 10 amino acids, the peptides identified in the SAX dataset will have at least 1 additional extra acidic residue, on average.

Figure 3.5. Overlap between total unique peptides in SAX and SCX fractions showing high complementarity. Diagram generated by Whitehead Institute Venn diagram generator.[11]

Figeys and colleagues have recently observed this phenomenon from peptides eluted after doing on-column digestion,[2,12] although it could not be generally extended towards direct SCX/SAX fractionation of peptides. A possible explanation is the higher negative charge on acidic peptides confers them higher affinity towards SAX functional groups than their protonated counterparts towards SCX functional groups. In turn, this strong affinity towards SAX material allows them to (1) not be lost during loading steps and (2) undergo better separation than if they were weakly bound. It is possible that the overlap would increase between the two separation techniques if replicates of the samples were analyzed; however, the complementarity of the techniques is clearly demonstrated, if lacking statistical rigor.

IMAC specificity

The specificity of the IMAC resin is best evaluated by comparing the number of phosphopeptides identified versus the number of identified non-phosphorylated peptides in the enriched sample. At an ions score cutoff of 15, 70% (110/157) of identified peptides contain at least one phosphorylation site. Increasing the ions score cutoff to a much more stringent value of 35, and thus filtering out all but the highest scoring results, the proportion of phosphopeptides decreases somewhat to 60% (32/53). This indicates that non-phosphorylated peptides are binding to the IMAC resin, albeit in a manner still allowing substantial enrichment of phosphopeptides. The IMAC-enriched fraction from the labeled, VSV-infected K562 sample provides some insight into the mechanism of non-phosphorylated peptides

binding to the resin. These enriched samples showed phosphopeptide proportions of 89% (361/405) and 100% (44/44) at ions score cutoffs of 15 and 35, respectively. These results are illustrated graphically in Figure 3.6. The only functional difference between the enrichment of the two samples was that the latter contained twice the amount of total peptides compared to the former (2 mg versus 1 mg). As the same volume of resin was used in both samples, this suggests that, at non-stoichiometric amounts of resin to phosphopeptides, non-specific binding occurs. The non-phosphorylated peptides typically contain one or more carboxylic acid groups (in aspartic and glutamic acid residues) that may co-ordinate with the iron group of the resin and thus provide competition to the phosphate groups for these binding sites, as previously discussed. If a sufficient number of phosphopeptides are present, they will competitively bind and occupy all the binding sites of the IMAC resin; however, if insufficient phosphorylated material is present, then non-phosphorylated peptides will occupy the remaining sites. It can therefore be concluded that this resin performs well for phosphopeptide enrichment, with the stipulation that the amount of phosphopeptide material is properly estimated and the amount of IMAC resin is adjusted accordingly.

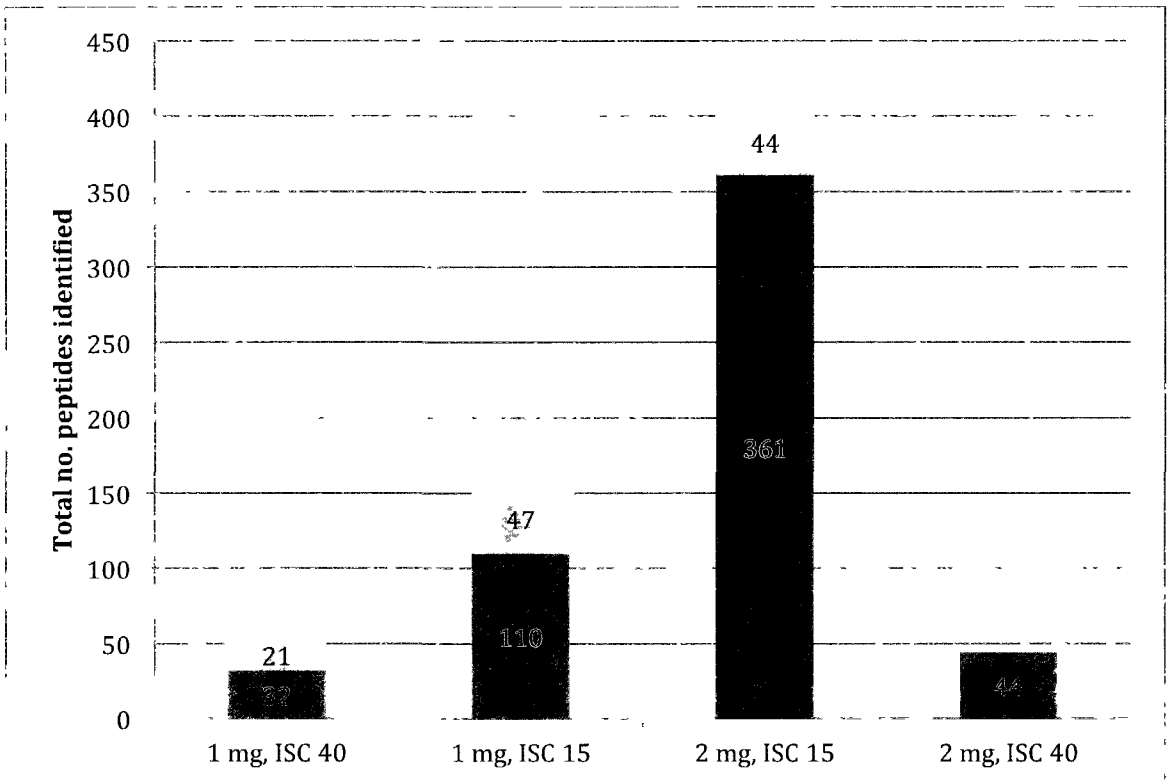


Figure 3.6. Comparison of number of phosphorylated (dark grey) versus non-phosphorylated (light grey) peptides for background (1 mg) and labeled (2 mg) samples. The samples are shown both at a stringent ions score cutoff (ISC) of 40 and a less stringent ISC of 15.

It is important to note that, while IMAC elution fractions showed high proportions of phosphopeptides, the IMAC flow-through fraction from the labeled sample was not entirely devoid of phosphorylated peptides. In fact, after SCX fractionation of the flow-through, a substantial number of phosphopeptides (78 unique) were identified in the low pH (3.5) fraction at an ions score cutoff of 15. The high-scoring peptides in the SCX fraction were predominantly singly phosphorylated, in contrast to the quantified phosphopeptides in the IMAC fraction, which showed many multiply phosphorylated species (~1.7 phosphorylation sites/peptide, on average). It is a previously known phenomenon that IMAC

enriches more strongly for multiply phosphorylated peptides,[13] so this will not be further discussed; it is sufficient to state that no single enrichment technique has been found to enrich for all phosphopeptides because they all show some bias.

IMAC reproducibility

The overlap between phosphopeptides identified at a high confidence level in the K562 background phosphoproteome and the VSV-infected K562 phosphoproteome is shown in Figure 3.7. The overlap is surprisingly small, indicating either poor reproducibility (either from the enrichment or mass spectrometry analysis steps) or a substantial change in the phosphorylation pattern. As the VSV-infected K562 phosphoproteome contains both the infected and control cells, labeled and combined, it is unlikely to see such a substantially poor overlap between the background and labeled phosphoproteomes. The expression of several phosphopeptides was found to be substantially altered (*vide infra*); however these changes are insufficient to explain the poor overlap. To further investigate this phenomenon, a technical replicate of the VSV-infected phosphoproteome, as previous studies have showed that technical replicates can increase the number of phosphopeptide identifications by 30-40%.[14, 15] While the technical replicate yielded many of the same top-scoring phosphopeptides, further investigation showed that of the 92 phosphopeptides identified, 59 (64.1%) of these were uniquely found in the replicate, a substantial improvement in number of phosphopeptide identifications. Nevertheless, this cannot account entirely for the poor overlap found in Figure 3.7; it is likely that the variation in phosphopeptide identification is also

due to the enrichment of samples by IMAC. This has been widely reported in the literature, with small changes in procedure being considered to enrich for complementary parts of the phosphoproteome. Based on these results, it would seem that a large part of this complementarity is in fact due to irreproducibility of IMAC enrichment. It has been stated in protocols that the repetition of entire experiments is recommended when doing quantitative phosphoproteomic analyses of biological samples[3]; while never stated explicitly in the literature, it is likely in part due to the inherent variability of IMAC.

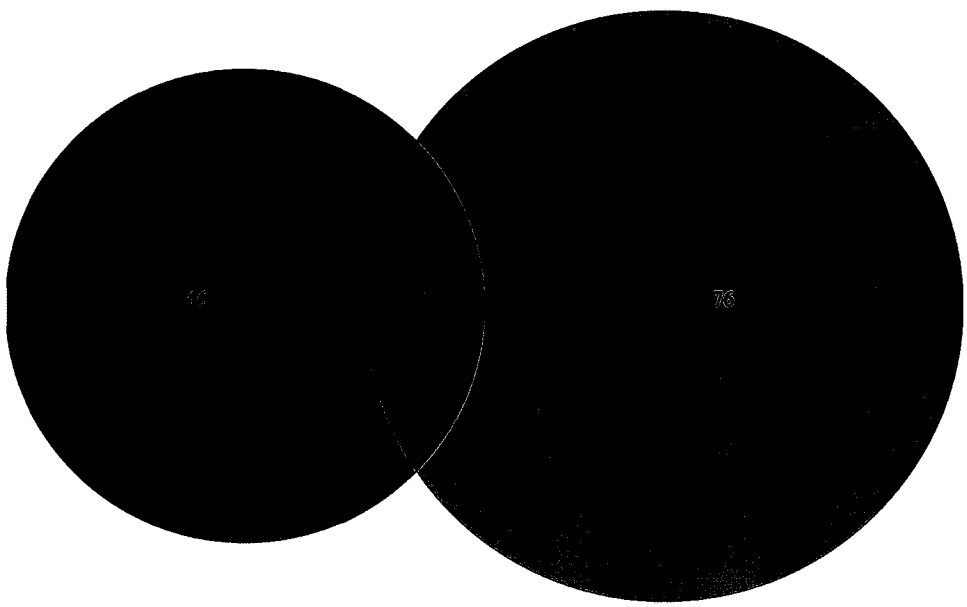


Figure 3.7. Overlap between phosphopeptides found in background K562 phosphoproteome (left, blue) at a Mascot score ≥20 and quantified phosphopeptides from VSV-infected K562 cells (right, red). Diagram generated by Whitehead Institute Venn diagram generator.[11]

3.1.6 Peptide identification and quantitation

Selection of ions score cutoffs

For both proteomic and phosphoproteomic analyses, an ions score cutoff of 20 in Mascot has become the default for filtering out poor spectral matches.[16, 17] This Mascot score was adopted for proteomic samples, but it was observed that relatively good spectra were being filtered out in phosphoproteomic samples. As seen in Figure 3.8, a spectrum matching a phosphopeptide with a neutral loss peak of -98 Da and a good y-ion series was assigned a Mascot score of 16. This meets the criteria for a good match, and should not be excluded from the dataset. For this reason, all spectra with Mascot scores between 15 and 20 were manually inspected for those that were quantified.

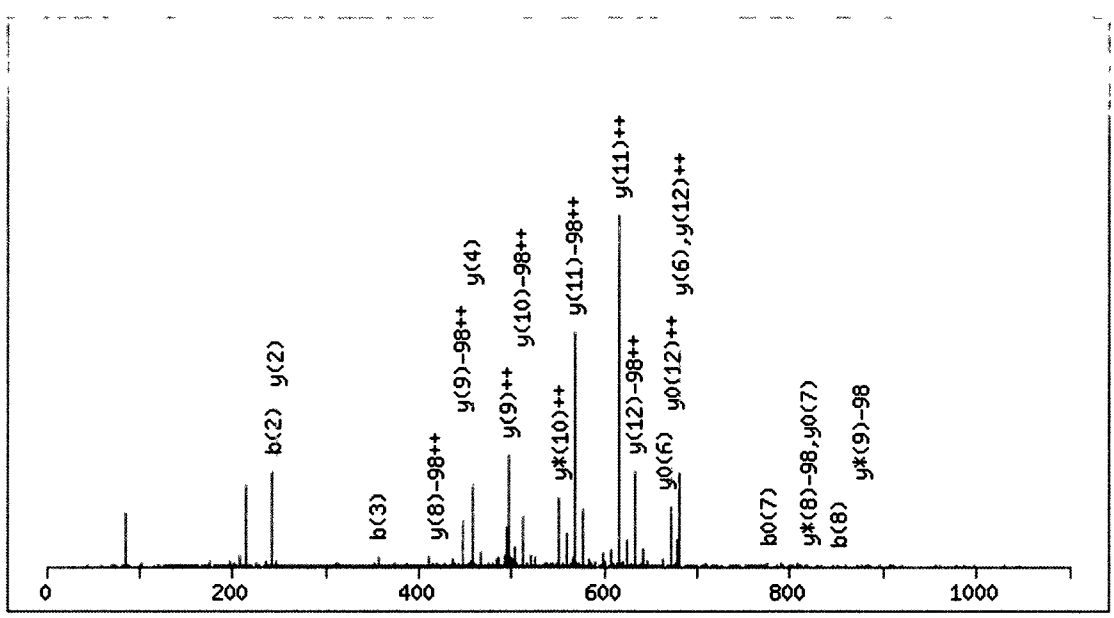


Figure 3.8. MS/MS spectrum identifying the phosphopeptide LELQGPRGpSPNAR. The Mascot ions score of this match was only 16 despite showing many ion matches.

Manual quantitation of peptides

After identification by Mascot, dimethylated peptides are quantitated manually through the Analyst software – the process is shown in Figure 3.9. Mascot lists the elution time of all identified peptides and this elution time can be used to find the peak in the MS spectrum (steps 1-2, Figure 3.9). An extracted ion chromatogram (XIC) of this mass is taken within a user-defined or automatic mass interval (3) and the maximum height of the XIC in the elution interval is taken (4-5). If Mascot has identified the other isotopic peptide of the pair, the process is repeated for the corresponding light/heavy peptide. The ratio of the heights obtained then gives the expression ratio of the peptide from the two states. If Mascot has not identified the other isotopic pair, however, then the m/z difference must be calculated by

$$\Delta\frac{m}{z} = \frac{(n)\times 4\text{Da}}{z} \qquad (3.1)$$

Where n is the number of modification sites in the peptide (lysine residues and N-terminus). If the mass of the heavy isotopic pair is trying to be found, this m/z is added to that of the light-labeled peptide, or subtracted if vice-versa.

This quantitation procedure is needed for dimethyl labeling due to the chromatographic shifts associated with the hydrophobicity difference between hydrogen and deuterium. The mechanism(s) governing this isotopic effect is not completely understood, but it is believed to relate to the change in vibrational energy from the increased nuclear mass rather than any potential energy changes,

as explained by the Born-Oppenheimer approximation.[18] In any case, since there are retention time shifts of up to a minute between the light- and heavy-labeled peptides, one cannot simply take the ratios of the peaks in a single MS spectrum and an XIC must be used to find the peak maximums for both light- and heavy-labeled peptides.

Automatic quantitation of peptides

The manual quantitation method is possible for very small datasets, but is far too time-consuming and labor-intensive to be feasible for large datasets that may contain thousands of spectra. Furthermore, the retention time shifts between light- and heavy-labelled peptides means that commercial software packages (e.g. AB Sciex ProteinPilot) are incompatible with dimethyl-labeled data. For these reasons, a custom script in PHP was written to perform the same steps performed in the manual quantitation method automatically. After Analyst file (.wiff) conversion to a parsable data format (mzML) by the open-source file conversion software ProteoWizard, masses from a Mascot search are automatically and sequentially searched using the parameters described in the Experimental section. The data is outputted as a tab-delimited file that can be opened in Microsoft Excel, where the ratios can be sorted. The XIC method was compared against a simple peak maximum method and the averages and standard deviations were found to be comparable. The automated XIC method has, however, been found to over- or underestimate the values for outlier ratios; for this reason, all peptides found to be over- or underexpressed are manually quantitated for verification.

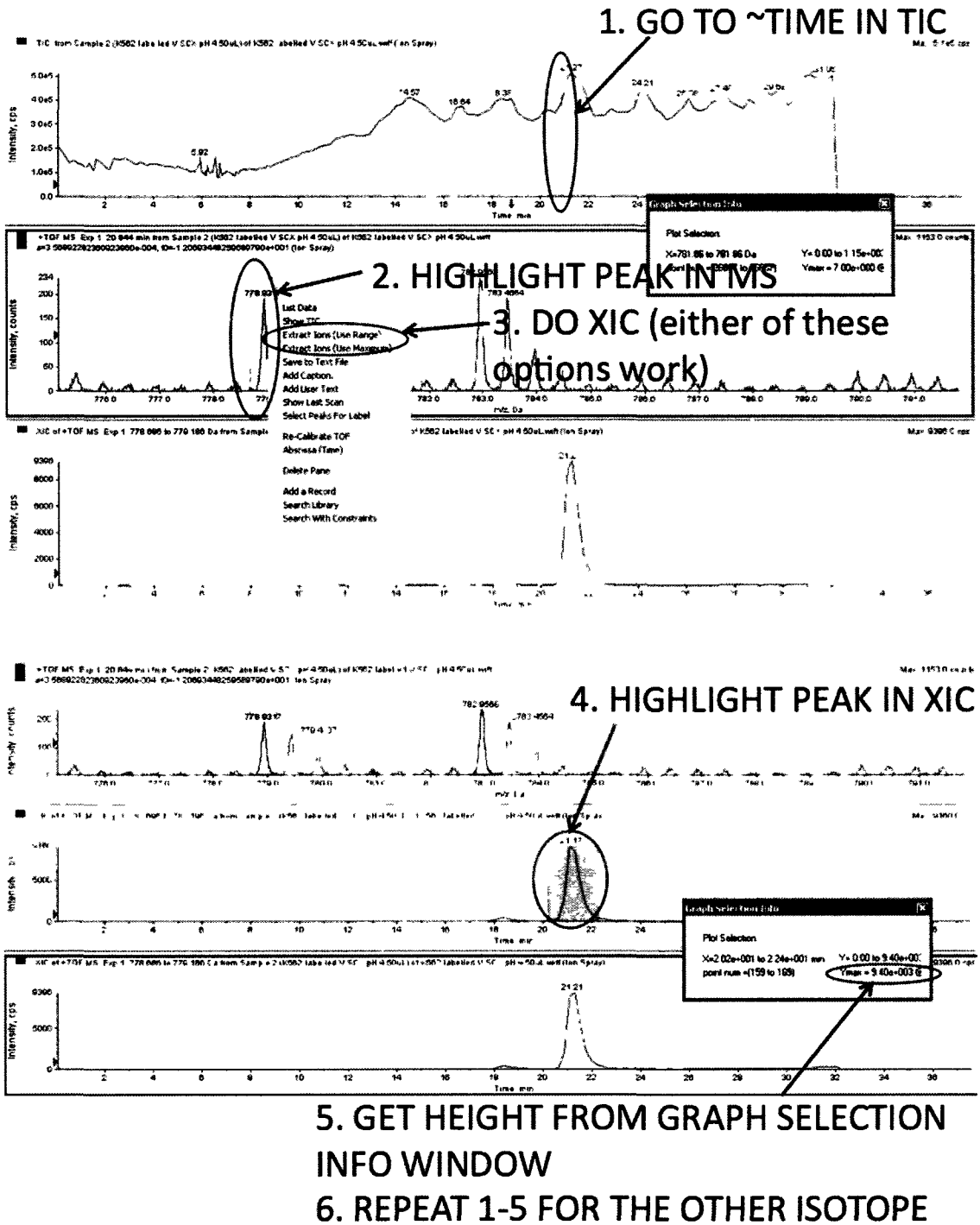


Figure 3.9. Manual quantitation method for dimethyl labeled peptides. Steps are described in the text.

3.2 Qualitative K562 analysis

3.2.1 Background proteomic analysis

A background proteome of untreated K562 control cells was collected by a 2-dimensional separation method consisting of SCX in the first dimension and RPLC in the second dimension. Details are provided in the Introduction and Experimental sections, but briefly: SCX in conjunction with pH elution separates based on isoelectric point of peptides, which is orthogonal to RPLC mechanism of separating by hydrophobicity. This allows a large background K562 proteome to be identified, which is necessary for (1) identifying functionally enriched features of this malignant cell line, (2) providing a background when searching upregulated proteins as the foreground, and (3) validating the qualitative aspects of the method and the separation efficiencies of the 2 dimensions.

Untreated K562 cells were lysed, digested, and fractionated using the previously described methods. A total of 6 SCX fractions (pH 4.5, 5, 5.5, 6, 7, & 800 mM ABC salt) containing a substantial amount of peptides were analyzed by LC-MS/MS using a 90-minute gradient and an information-dependent acquisition mass spectrometry method. The resulting data files were processed and concatenated to give a single, searchable file. This file was compared against an *in silico* digested and fragmented copy of the IPI human database using Mascot software. After filtering peptide matches below a Mascot ions score cutoff of 20, a total of 699 proteins were identified. The gene symbols of these proteins were extracted, compiled, and submitted to the GeneCoDis algorithm to identify functionally enriched properties of

this dataset. The Ensembl human database was used as a background and the gene ontology (GO) biological processes (BP), molecular functions (MF), and cellular components (CC) were all investigated. Of the 699 proteins found, 660 had GO annotations; the remainder could not be searched. The algorithm identified functional enrichment of many GO annotations, particularly for cytosolic, ribosomal, and nuclear cellular components, protein and nucleotide binding molecular functions, and translation and translational elongation biological processes. It is not surprising that mass spectrometry identified primarily proteins in soluble cellular components, as these types of proteins are most amenable to tryptic digestion and mass spectrometry. Similarly, the molecular functions and biological process that are functionally enriched are generally soluble proteins that are involved in the translation of proteins and are of relatively high abundance. The top 8 functionally enriched GO terms are shown schematically in Figure 3.10, as well as all 15,905 other annotations occurring at a frequency of 10 or greater for this set of proteins. A more complete set is listed in Appendix 1. An important aspect of establishing the background is that mass spectrometry generally can only investigate a narrow dynamic range; as such, peptides identified in shotgun proteomics usually belong to abundant proteins rather than non-abundant ones present in low concentrations (e.g. cytokines). This functional enrichment therefore identified a background proteome that is abundant and amenable to mass spectrometry and, without quantitation, does not in fact reveal properties regarding the biological state of the cell.

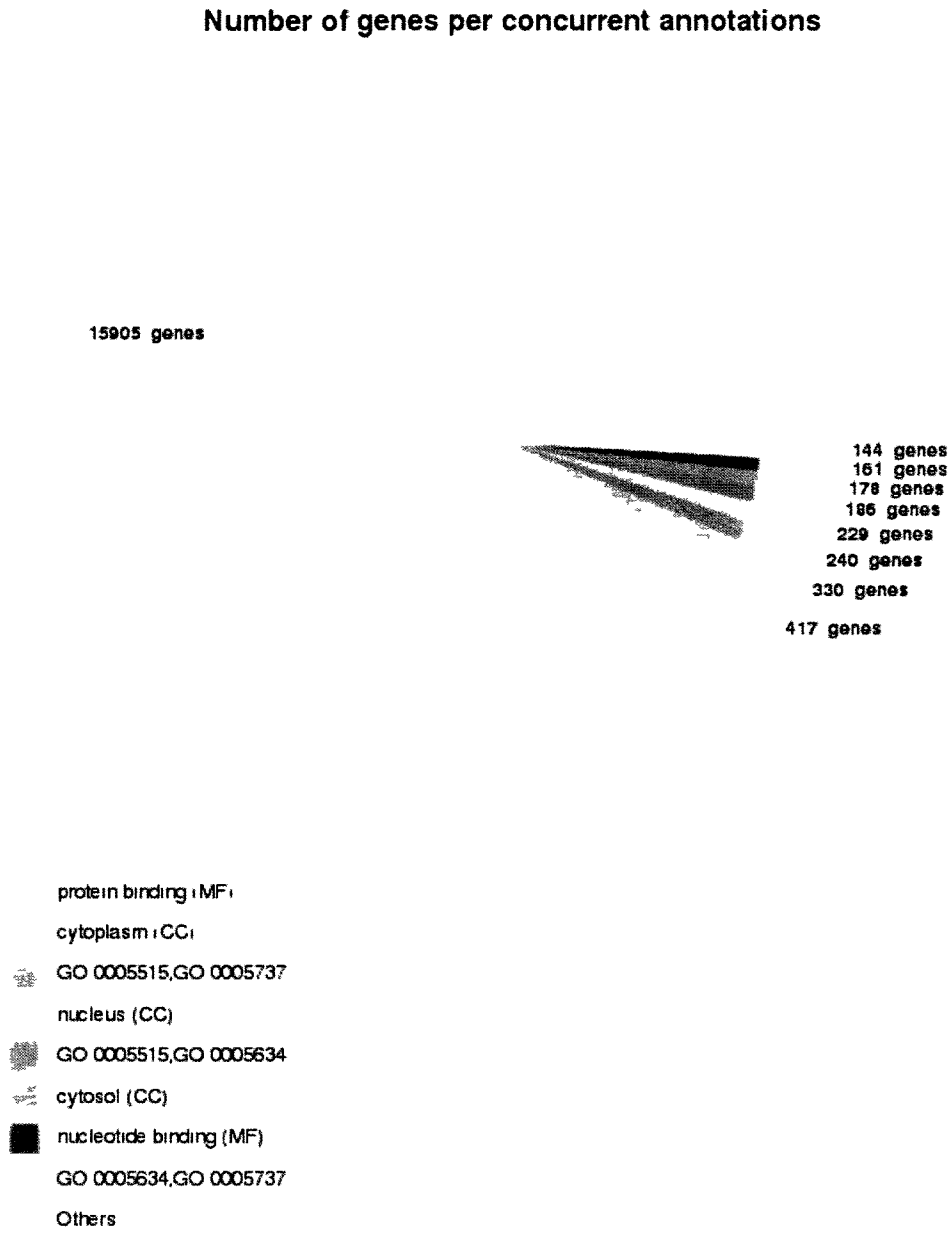


Figure 3.10. Functionally enriched annotations in the proteomic dataset relative to the Ensembl human database. Figure generated by GeneCoDis software. Shorthand GO annotations used: GO:0005515 (protein binding MF), GO:0005737 (cytoplasm CC), GO:0005634 (nucleus CC).

3.2.2 Background phosphoproteomic analysis

A background phosphoproteomic analysis of K562 cells was performed to evaluate the performance and specificity of the IMAC resin (discussed in section 3.1.5) and to identify the major phosphorylation sites and motifs in this cell line.

The 96 phosphopeptides identified in the background phosphoproteome were classified into 3 groups defined by their general motif: proline-directed, acidic, basic, or other. The methodology used for these classifications is explained in the Experimental section. This process yielded 43 (44%) phosphopeptides containing proline-directed motifs, 20 (20%) containing basic motifs, 19 (19%) containing acidic motifs, and 16 (16%) containing other motifs. Note that 2 phosphopeptides contained motifs that had criteria classifying it as both basic and acidic. The genes containing these phosphopeptides were entered into GeneCoDis to search for functional enrichment of GO annotations (biological process, molecular function, cellular component) in each category.[19, 20] It was found that proteins containing proline-directed motifs were functionally enriched in the nucleus (24 genes), basic motifs (7 genes) were in the cytosol, and acidic motifs in the cytoplasm (8 genes) and nucleus (8 genes), consistent with previous reports.[21] Genes from the proline-directed category had significant functional enrichment of the nucleotide binding (17 genes) and RNA binding (11 genes) molecular functions and mRNA processing biological process (8 genes), suggesting that many of the phosphopeptides found in this experiment are involved in gene expression and regulation, particularly those containing proline-directed motifs.

The Motif-X algorithm was then used to find specific motifs in the background phosphoproteome by searching an interval of 6 amino acids around the centered phosphorylated residue.[22] The algorithm reduced the original 96 phosphopeptides to 82 unique sites, followed by a further reduction to 74 sites (5 too close to protein termini, 5 containing no unique mappings). The algorithm identified 3 motifs when searched against the IPI human database: 1 basic (R-X-X-pS), 1 acidic (pS-X-X-E), and 1 proline-directed (pS-P), with 10, 14, and 36 phosphopeptide matches, respectively. These motifs are shown in Figure 3.11, where the motifs are centered around the phosphorylated residue and the frequency of other residues is directly proportional to their size. The aligned sequences corresponding to each discovered motif are listed in Appendix 2. None of these motifs are consensus sequences for a specific kinase by themselves; for example, pS-P is not universally recognized, but P-X-pS-P is the motif recognized by MAP kinase. Similarly, R-X-X-pS is not a specifically recognized motif, but the subclasses R-R-X-pS and R-X-R-X-X-pS are recognized by CK2 and Akt kinase, respectively. It is important to note that these simple motifs are a gross simplification of the 3-dimensional domains recognized by kinases; nevertheless, they are widely reported in the literature and have been a successful model for the phosphorylation of specific sites by kinases.

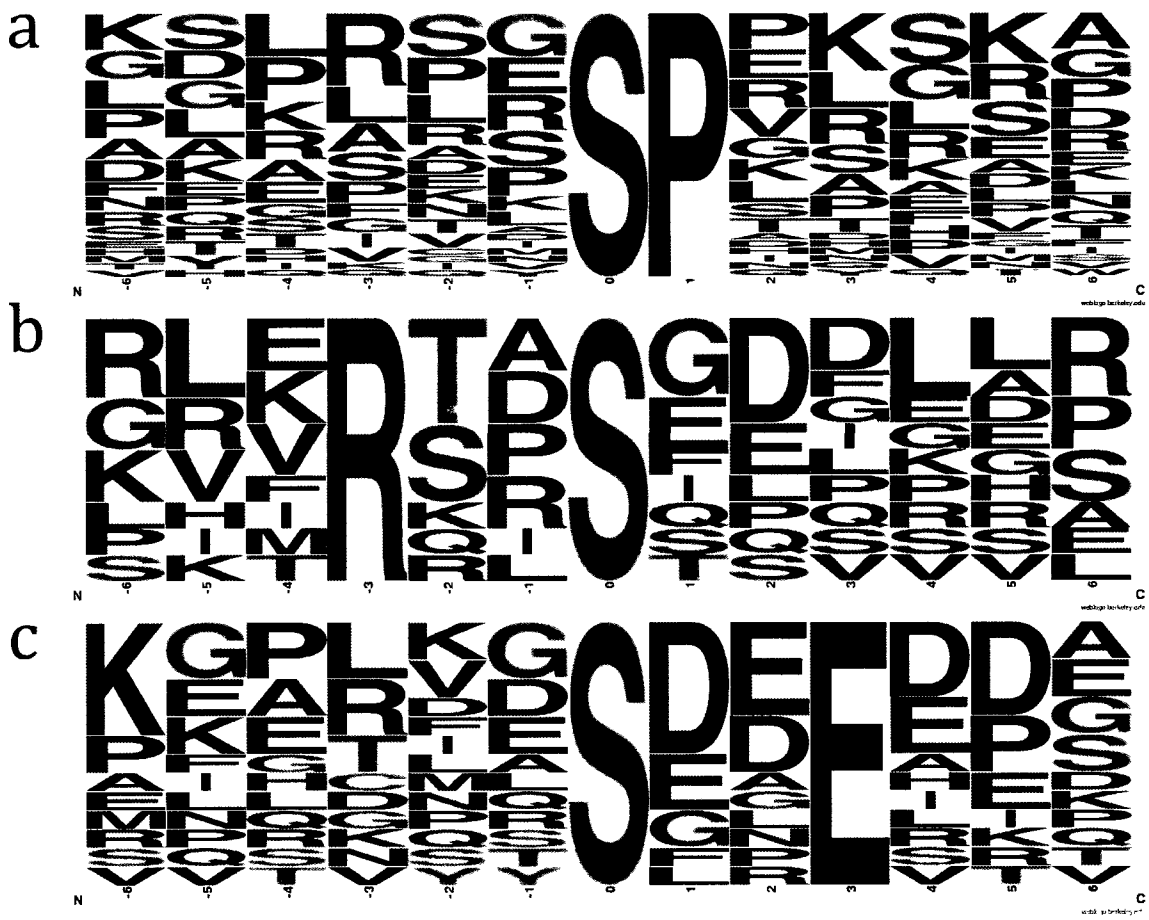


Figure 3.11. Phosphorylation motifs found in the background phosphoproteome using the Motif-X algorithm. Dataset included all phosphopeptides with a Mascot score ≥15. The algorithm identified a proline-directed motif (a), a basic motif (b), and an acidic motif (c).

Effect of phosphorylation on number of missed cleavages

It was observed that phosphoproteomic samples contained far more missed cleavages than their proteomic counterparts. Calculating the number of missed cleavages per peptide for the background K562 proteome showed they contained on average 0.11 missed cleavages/peptide ($n = 2188$) versus 0.84 missed cleavages/peptide ($n = 108$) for the background phosphoproteome or 1.36 missed

cleavages/peptide ($n = 405$) for the labeled phosphoproteome. Despite having a higher trypsin-to-substrate ratio (1:25 versus 1:50), the labeled phosphoproteome actually contained more missed cleavages than the background phosphoproteome, indicating that a lack of protease was not the reason for this phenomenon. This has been previously reported in the literature[23] when Molina and colleagues observed that Lys-C was no more effective than trypsin for the generation of long, ETD-amenable phosphopeptides. They found that phosphopeptides contained an average of 1.1 missed cleavages/peptide, comparable to the number found here. Evidence has shown that trypsin shows poor activity if the cleavage site contains an aspartic acid or glutamic acid immediately N- or C-terminal to it. Of the 91 missed cleavage sites in the background phosphoproteome, 52 (57%) of these have this property; Table 3.2 shows the breakdown of these missed cleavages. It has also been found that phosphorylation sites in proximity to a lysine or arginine will also prevent digestion at this site, presumably due to the acidity of the phosphate group and particularly at R/K-X-pS/pT sequences.[24] The partial digestion of proteins can be problematic in quantitative proteomics, as small differences in tryptic activity can lead to peptides being spread across different numbers of missed cleavages in varying abundances. That the proteomic samples showed nearly full digestion and that the phosphoproteomic samples containing more trypsin did not show a decrease in missed cleavages indicates that the tryptic digestion was efficient. The missed cleavages are at known sites of poor trypsin efficiency and should not impair the quantitation of the phosphopeptides.

Table 3.2. Classification of missed cleavage sites in the background K562 phosphoproteome.

Sequence	Number present	Percentage of total (%)
KD	7	7.7
KE	11	12.1
DK	7	7.7
EK	3	3.3
RD	1	1.1
RE	3	3.3
DR	13	14.3
ER	7	7.7
R/K-X-pS/T	3	3.3
Others	36	39.6
Total	91	100

3.3 Quantitative VSV-infected K562 analysis

3.3.1 Quantitative proteomic analysis

Validation of the strategy

After collecting data from 12 total fractions (6 SCX, 6 SAX), the resulting mass spectra were compiled and processed by the script described in the Experimental section. An evaluation of the inter- and intra-fraction protein variation was done to quantify the precision of the method. Actin is an ideal protein in which to do this, as it is an abundant, highly conserved protein whose main role is structural. As such, it is unlikely to be differentially expressed under varying biological conditions and levels should remain constant in the same cell type. It was found, for all fractions containing high-scoring peptide hits to cytoplasmic actin (gene name: ACTB), that the average intra-fraction protein L/H ratio was 0.873 with a standard deviation of

0.083. This gives a relative standard deviation across all fractions of just under 10% (9.6%). An individual fraction with multiple high-scoring actin hits (SCX pH 3.5) was used to evaluate the variation of the same protein in the same fraction and the numbers were comparable to the inter-fraction results: a L/H average of 0.871 with a standard deviation of 0.093 (RSD = 11%).

These averages differ somewhat from the average L/H of all peptides in a fraction. The L/H among the 12 quantified SAX/SCX fractions was found to be 1.55 (range 1.18-1.97), indicating that there are substantially more upregulated than downregulated proteins. The relative standard deviation on the average L/H for each fraction was ~92% (range: 66-130%). Note that this does not necessarily mean the method is not accurate; it merely gives an idea of the biological protein variation in the two cellular states.

Since there was such a large discrepancy between the L/H ratio for actin and the average L/H of all labeled proteins in the sample, it was difficult to assign a value for those that are significantly up- or downregulated. It was decided that ratios ≤ 0.5 would be considered downregulated and those ≥ 2.5 would be considered upregulated.

Properties of the differentially regulated proteins

After filtering of ratios above 0.5 and below 2.5, the identifications of differentially expressed peptides were manually validated from the MS/MS spectra in Mascot and the ratio manually validated in Analyst. The resulting peptide masses were matched to their peptides in Mascot and gene names noted. Overall, 53

proteins were found to be upregulated in VSV-infected K562 cells versus control K562 cells and 11 proteins were found to be downregulated. A list of protein gene names, their L/H ratios, and the number of peptides used in the quantitation are tabulated in full in Appendix 3. The upregulated and downregulated proteins were submitted to STRING 8.3 (string-db.org) to identify protein-protein interactions and to create a functional network. The network for the overexpressed group of proteins is depicted in Figure 3.12; the downregulated proteins showed no protein-protein interactions and are therefore not shown. GeneCoDis was used to further identify functionally enriched GO annotations in the differentially expressed proteins (using the background K562 proteome as the background).

It has been shown that VSV carries host proteins within its envelope into other cells[25] and recent work by Grdzelishvili and colleagues has successfully characterized this "virion proteome" by mass spectrometry.[26] That work identified a total of 64 proteins, many of which that have been found to be upregulated in the VSV-infected K562 cells, including elongation factor 1 alpha (EEF1A), heat shock protein 90 (HSP90), histone H4 (HIST1H4A), lactate dehydrogenase A (LDHA), Ras-related protein (RAP1B), and heat shock 71 kDa protein (HSPA8). As there is overlap between the upregulated proteome here and the virion proteome, there is the possibility that the upregulation is partially due to the deposition of the virion proteins into the cell. However, if this were the case, all identified proteins in the study by Grdzelishvili would be expected to be upregulated. Cytoplasmic actin was one of the virion proteins found and therefore this protein would be upregulated as well; on the contrary, it was found that actin was not upregulated. Actin was found

to be in a ratio (L/H) of approximately 0.87 and approximately at the desired 1:1 ratio, with little variation. The ratio of actin had adequate precision between the VSV-infected cells and control cells that it was used for standardization, as described above. Furthermore, if these proteins were in sufficient quantities to skew the L/H ratios, one would expect a large amount of viral envelope protein present in the sample as well. A search of the data against the NCBI database in the Virus taxonomy, however, shows no matches, indicating that this phenomenon is not a factor. Finally, the MOI used here (10) means that, on average, 10 virus particles are introduced per cell – a number inadequate to alter the cellular proteome in itself. An alternative theory is that the proteins that are embedded in the viral envelope are those that happen to be upregulated in response to viral infection and thus are more likely to be embedded in the viral envelope. As many proteins found to be upregulated in the VSV-infected K562 cells have been found to be necessary in some role for VSV, those that are of interest will be discussed in detail.

The only functionally enriched cellular component was the ribonucleoprotein complex annotation (7 genes, $p < 0.05$). The 7 proteins comprising this annotation are 4 of the heterogeneous nuclear ribonucleoproteins (hnRNPs), a 70 kDa heat shock protein (HSPA8), ribosomal protein L4 (RPL4), and Lupus La protein (SSB). This complex is responsible for binding pre-mRNA and for the transport from the nucleus to the cytoplasm. The members of the complex are seen in the bottom right corner of Figure 3.12; interestingly, the members are not exactly identical, except for the hnRNPs. It is possible that SSB, RPL4, and HSPA8 associate with the complex but do not directly physically interact with the hnRNPs and are thus not included in

the ribonucleoprotein complex in the network. Note that the SFRS5 and SFRS6 proteins are splicing factors and, while not annotated as members of the ribonucleoprotein complex, they interact with this complex as they are involved in removing introns during the splicing process.

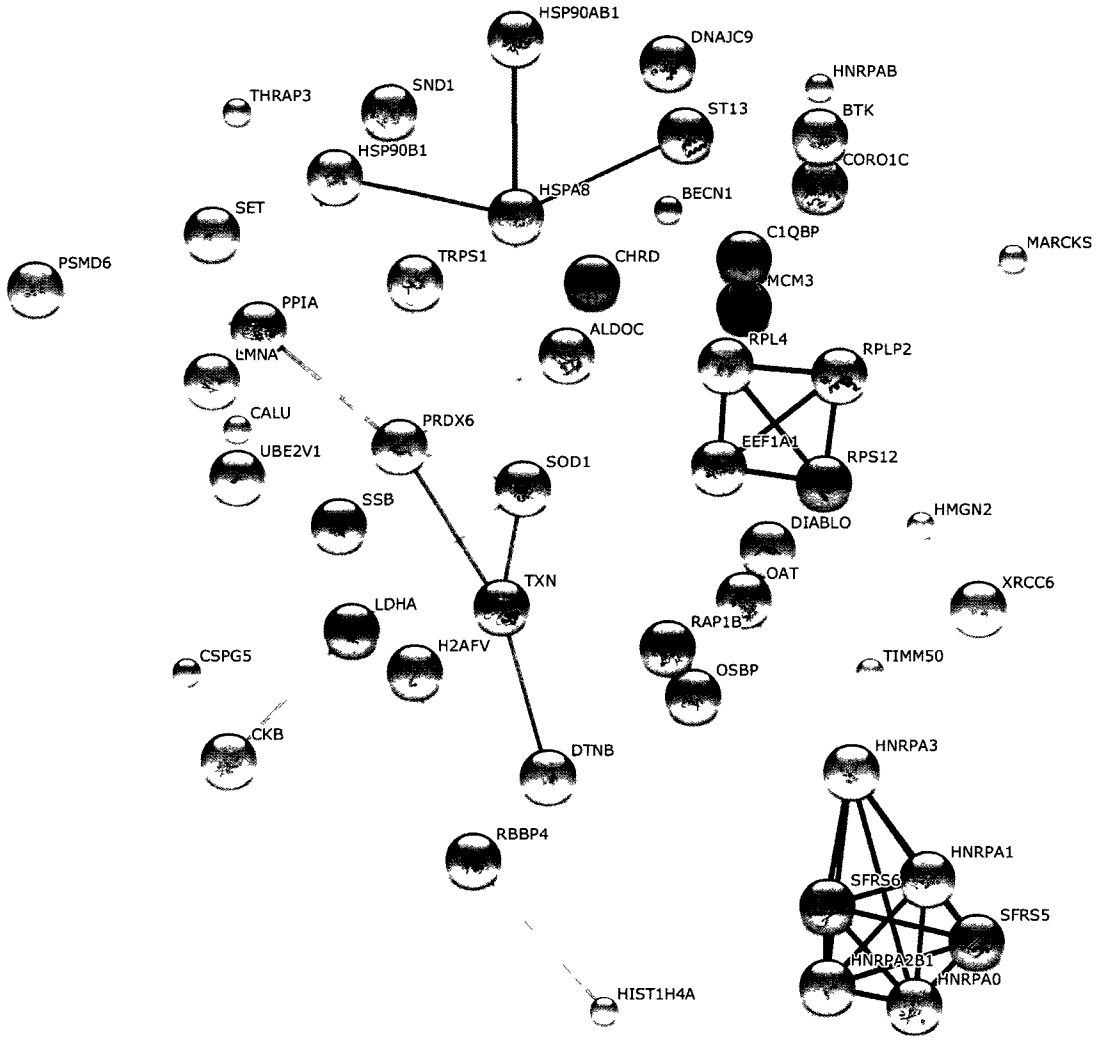


Figure 3.12. Protein-protein interaction network generated by STRING 8.3. Confidence view is shown: line weight is proportional to interaction confidence.

Role of heterogeneous ribonucleoproteins

It was suggested in section 1.5 that VSV protein(s) targets the nuclear pore complex and inhibits nuclear import and export, although there have been studies that have supported[27, 28, 29] and disputed this.[30, 31] It is known that VSV inhibits gene expression via its M protein, but it has also been shown to bind Rae1,[31] a protein involved in nuclear mRNA export but not required for mRNA export in higher eukaryotes[32]; its role is therefore currently unclear. A class of proteins termed hnRNPs (heterogeneous ribonucleoproteins) are nuclear-cytosolic transport factors and play a large role in mRNA export. After VSV infection, it has recently been shown that some of these proteins (hnRNPA1, hnRNPK, hnRNPC1/C2) relocalize to the cytoplasm, while others (hnRNPB1, lamin A/C) do not.[33] That study revealed many other important aspects of VSV function and hnRNP behavior: (1) proteins with nuclear localization signals (NLS) retained their ability to undergo nuclear import after VSV infection, thus nuclear transport is not inhibited; (2) hnRNP relocalization required Rae1; and (3) reduction in hnRNPA1 by silencing slowed the rate of apoptosis in VSV-infected cells, but had no effects on viral replication. That research showed for the first-time the differences in concentration of hnRNPs in a transformed cell line using VSV. They suggest that further studies are needed to show whether this is due to an increased export of some hnRNPs or due to an increase in cytoplasmic retention. The data presented here suggest a third option, however: that the increase in cytoplasmic hnRNP is due to an overall increase in hnRNP expression. The K562 cell line showed a dramatic overall increase in several hnRNP species, including hnRNPA1 (6.2-fold increase), hnRNPA2 (3.2-fold

increase), and hnRNPA2B1 (6.0-fold increase). Overall, a total of 6 hnRNP species were shown to be significantly upregulated in VSV-infected K562 cells versus control K562 cells. Furthermore, another hnRNP, lamin A/C, that has been shown to not localize to the cytoplasm during VSV infection is overexpressed as well, indicating a general trend among many hnRNPs.

It is not entirely clear why the cell would increase the levels of hnRNPs in response to viral infection, particularly when these increases speed up the apoptotic effects of the virus. It would be interesting if this increase in hnRNPs were responsible for the improved apoptotic effects of VSV on transformed cell lines, however a recent proteomic study on a different virus and non-transformed cell line (dengue fever in endothelial cells) showed an increase in hnRNPs after infection as well.[34] This would suggest that the data presented here showing an increase in hnRNP concentration in VSV-infected K562 cells is perhaps a general antiviral response by the cell rather than being specific to VSV or transformed cell lines. While the increase in hnRNPA1 has been shown to aid apoptosis in VSV-infected malignant cells, there are other viruses (HTLV-1) that replicate more rapidly when there is a deficiency in hnRNPA1,[35] meaning that no general trend can be established. The hnRNPA1 protein has, however, been shown to interact with IκBα and induce transcription by the NF-κB pathway[36] where it may induce transcriptional activation of apoptotic genes. Alternatively, apoptosis may be induced by hnRNPA1 by internal ribosome entry site-initiated translation of the apoptotic peptidase activating factor 1 (apaf-1),[37] followed by release of cytochrome c and activation of caspase-3 dependent apoptosis.[38]

Interestingly, a recent study where an immunoprecipitation of several hnRNPs was performed showed that vimentin is an interactor with hnRNPs as well as associating with a viral protein of dengue fever.[39] Disruption of vimentin complexes resulted in dissociation of this complex and a reduction in viral replication. Vimentin was a protein found to be upregulated in VSV-infected K562 cells as well, but was filtered out of the dataset due to being slightly below the ratio cutoff. Due to the similarities in overexpression of hnRNPs in dengue fever and VSV viruses, it is possible that vimentin plays a role in VSV infection as well and should be further investigated.

Nevertheless, this is the first data showing that the total cellular concentration of many hnRNPs is increased rather than simply undergoing localization changes. Further studies using subcellular fractionation should be done to confirm this phenomenon and identify if this total concentration increase is due to increases in all cellular compartments or if it is compartment-specific.

Role of chaperone proteins

Calnexin is a chaperone protein that was found to be upregulated and is critical for the folding of VSV glycoprotein (G), which is responsible for transporting the virus to the cell surface. Directly after synthesis, BiP, the first molecular chaperone to associate with G protein, folds G protein into monomers.[40] After dissociation from BiP, G protein folds into trimers with the help of the second molecular chaperone, calnexin. Calnexin has been shown to be necessary for the correct folding of G protein and hence for the subsequent, consecutive transport to

the Golgi apparatus and plasma membrane, where they appear on the cell surface.[41,42] The increase in expression of calnexin shown here is therefore likely induced by VSV to aid in its translocation, budding, and infection of new host cells.

The heat shock proteins are another class of molecular chaperones and several members are upregulated; they are grouped at the top of Figure 3.12. While the heat shock proteins were initially believed to only be responsible for polypeptide folding under cellular stress (e.g. heat), one of its members, Hsp90 (HSP90 gene), has also been associated with signal transduction after stimulation by various hormones or growth factors.[43] As such, Hsp90 has been shown to interact with specific "client" proteins (kinases, steroid receptors, transcription factors), unlike Hsp70 (HSP8A), which does not show specificity.[44] Hsp90 has also been found to be overexpressed in malignant cells and, clinically, has been associated with reduced survival rates in breast cancer.[45] The role of Hsp90 in cancer is extensive as many of its client proteins are oncoproteins involved in evasion of apoptosis, non-responsiveness to growth suppression signals, etc.; for this reason, it has become a therapeutic target for treatment of malignant cells.[46]

Hsp90 not only plays a large role in malignant cells, but it also plays a role in the infection ability of negative-strand viruses. It was recently shown that inhibition of Hsp90 or knockdown of expression by siRNA resulted in destabilization of the L protein of VSV.[47] This folding of this protein is responsible for the function of the RNA polymerase and, with Hsp90 activity inhibited, viral growth is slowed by more than 2 orders of magnitude. That Hsp90 is overexpressed in malignant cells and that

it is an important factor in viral growth may be a previously unexplored factor in VSV's ability to grow in tumor cells. Furthermore, the data presented here show that VSV-infected K562 cells show substantial upregulation of Hsp90, with a 3.8- and 7.6-fold increase for HSP90AB1 and HSP90B1, respectively. VSV therefore shows the ability to increase Hsp90 expression, even in a malignant cell line where Hsp90 is already overexpressed. The upregulation of the heat shock proteins is likely due to the general cellular stress applied to the cell when infected, but the increase may also have the secondary, non-intentional role of aiding viral growth.

It is important to note that, while both VSV and Hsp90 inhibitors are both good candidates for anti-cancer treatments, a combinatorial therapy would most likely be less effective than either in itself. Hsp90 inhibitors would block the ability of VSV to replicate and thereby not induce cytolysis in a VSV-induced manner. Furthermore, any VSV-induced increase in Hsp90 would counteract the effectiveness of the Hsp90 inhibitor.

Role of nucleosome proteins

SET is a 39 kDa protein found predominantly in the nucleus but also occurring in the ER as part of the large (270-420 kDa) endoplasmic reticulum-associated complex. This complex has also been found to have HMG2 (HMGN2 gene) as a member and plays two roles: (1) regulating DNA by controlling chromatin structure and (2) induction of T-cell mediated cytolysis after viral infection by granzyme A (GzmA).[48] The first role is more likely to be the one suggested by the observed upregulation of SET and HMG2: VSV blocks transcription in the nucleus as

previously described and a cellular response is triggered. Evidence shows that SET reverses nucleosome assembly, increases DNA accessibility, and results in increased transcription.[49] Therefore the cell, in response to a decrease in transcription by VSV-induced effects, upregulates proteins that are able to increase DNA accessibility and transcription to compensate for this loss. This hypothesis would further be confirmed by subcellular fractionation to confirm that the differential expression of HMG2 and SET in VSV-infected cells is occurring in the nucleus and not the ER.

The second scenario is considered unlikely as it is involves T-cell mediated import of GzmA-containing vesicles into the infected cell. GzmA is a serine protease that induces apoptosis by a novel, non-caspase dependent manner in which DNA is "nicked".[50] It has been postulated that, as HMG2 and SET open up the chromatin structure, the exposed DNA will undergo digestion by apoptotic nucleases.[48] This may also suggest why histone fragments were found to be upregulated after VSV-infection: after removal of histone tails by GzmA, chromatin undergoes a configuration change from a compact to extended state, thus allowing nuclease digestion.[51] The histone fragments would no longer be bound to the chromatin and, instead of being lost after lysate centrifugation in the sample workup step, would appear in the proteome. Furthermore, a peptide originating from granzyme A was identified in the proteome of K562 cells (albeit at a low Mascot score 25), suggesting its presence is possible in this cell line. If true, it would be the first time that transformed cells undergo this mechanism of apoptosis after VSV infection. A large degree of skepticism remains, however, as it has not been previously shown that GzmA can be expressed in this cell line without the presence of cytotoxic T-

lymphocytes or natural killer cells and it seems unlikely this mechanism can be activated in any other manner. Furthermore, as GzmA is a protease, it is not entirely clear why its substrates should be upregulated upon VSV-infection. A related protein, granzyme B, has been shown to be induced by the stimuli tumor necrosis factor alpha and Fas ligand in myeloid leukemia cell lines.[52] Intriguingly, this process was mediated through reactive oxygen species and can be induced by thioredoxin reductase inhibition. Thioredoxin was found to be upregulated (3.4-fold) and would have the same effects as the accumulation of thioredoxin after thioredoxin reductase inhibition. Fas is also involved in one of the apoptotic pathways believed to be involved in VSV-induced apoptosis (*vide infra*) and these data suggest a possible role of the granzyme family of proteins in VSV-induced cell death.

Role of reactive oxygen species (ROS)

It is now common knowledge that reactive oxygen species (ROS) and hypoxic conditions play a role in carcinogenesis. Uncontrolled cell proliferation leads to chronic hypoxia, which triggers many of the common hallmarks of cancer, including apoptotic resistance, altered metabolism, and angiogenesis.[53] ROS are important in malignant cells as they are responsible for apoptotic suppression and cell proliferation through ROS-dependent signaling mechanisms such as tyrosine kinases.[54] There are two strategies for treating cancer within the ROS paradigm: (1) decrease the number of ROS (or increase ROS scavengers) to decrease ROS-dependent signaling and cell proliferation or (2) increase the number of ROS (or

decrease ROS scavengers) to induce apoptosis. The latter involves ROS (or pharmacological agent) destabilizing the voltage dependent anion channel (VDAC), resulting in mitochondrial pore opening.[55] The mitochondria releases pro-apoptotic factors such as cytochrome c, AIF, procaspases, and Smac/Diablo (section 1.5).[56]

Interestingly, in the VSV-infected K562 cells, proteins with ROS-scavenging and anti-oxidant properties were found to be upregulated, including thioredoxin (TXN), superoxide dismutase (SOD1), and peroxiredoxin (PRDX6), although at least one is downregulated (glutathione s-transferase, GSTP1). It would seem that VSV, in fact, is an anti-apoptotic virus in K562 cells, which is known to not be true. One apoptotic protein downstream of this cascade, Diablo, was upregulated, perhaps due to a different pathway than the one described here. However, it is unlikely that these apoptotic/anti-apoptotic factors are functioning in a ROS-independent manner due to the ubiquitous nature of ROS in cancer cells and differential expression of several ROS-scavenging species. In fact, it was shown that the voltage dependent anion channel protein (VDAC1) was downregulated: an effect known to slow transformed cell growth.[57] Alternatively, this could suggest that the voltage channel has been destabilized and the mitochondrial pore opened; however, this is contrary to the evidence that ROS-scavenging species are upregulated, as ROS are required for VDAC destabilization. It is possible, however, that the Bak and Bax proteins from the Bad-mediated apoptotic pathway could destabilize the VDACs as well. This will be discussed in further detail as it was found that a differentially expressed phosphorylation site is an upstream regulator of this pathway (*vide infra*).

Another protein, translocase of inner mitochondrial membrane 50 (TIM50), has been shown to play an important role in regulation of mitochondrial membrane permeabilization. It was found that downregulation of this protein induced membrane permeabilization and release of cytochrome c to the cytoplasm.[58] This protein was found to be upregulated in VSV-infected K562 cells (3.2-fold) and it is possible that it has the anti-apoptotic effect of preventing membrane permeabilization.

Proteins related to reactive oxygen species clearly play some role in VSV-induced effects on leukemic cells as many of these proteins are differentially expressed. It is difficult, however, to define the exact mechanisms governing the observed changes. Future work will have to be performed to corroborate this evidence and to provide data that shows in more detail the effects that VSV infection has on ROS-scavenging proteins and associated apoptotic pathways. Future work will also have to show that the expression changes of these ROS-related proteins are induced only in transformed cells – it is possible that these changes are a general phenomenon in all VSV-infected cells, perhaps in the role of nucleotide synthesis for viral replication.

3.3.2 Quantitative phosphoproteomic analysis

Of the 96 phosphopeptides identified in the double-labeled fraction, 46 showed well-defined peaks suitable for quantitation by MS and adequate MS/MS spectra for high-confidence identifications. The phosphopeptides were quantified

using the XIC-maximum method as described in the Experimental section. The average light/heavy ratio (i.e. ratio of VSV-infected cells versus control) was found to be 1.58 with a large standard deviation of 1.38. It is important to note that small datasets can easily be skewed by outliers due to low sample size. For example, by the removal of the top 3 up- and down-regulated phosphopeptides, the average L/H and standard deviation are reduced to 1.38 and 0.65, respectively. A recent article in which the authors performed dimethyl labeling used a ratio of 2 (or 0.5) to establish whether a protein was significantly up- or downregulated, however phosphorylation was not investigated.[59] The ratios used as cutoffs in the quantitative proteomic analysis were considered (2.5 and 0.5); however, unlike in the proteomic analysis where actin can be used for standardization, there is no standard, unchanged phosphopeptide in the phosphorylated dataset. It was decided to keep the criterion for upregulated phosphopeptides (2.5), but choose a less stringent criterion (0.75) for downregulated phosphopeptides. Phosphopeptides were also pseudo-standardized against their non-phosphorylated counterparts; for example, if a non-phosphorylated peptide was upregulated, it was assumed that a phosphopeptide from the same protein and containing a similar L/H ratio did *not* have an upregulated phosphorylation site.

Four upregulated (eight sites) and four downregulated (nine sites) phosphoproteins met these criteria, as shown in Table 3.3a and 3.3b, respectively. Note that no downregulated phosphopeptides would have met the criterion set for downregulated, non-phosphorylated peptides; this could be an indicator that these phosphopeptides may not, in fact, be downregulated. An investigation into the

function of these proteins shows that several have roles in transcriptional activation or have apoptotic/anti-apoptotic properties, which will be briefly discussed. A functional enrichment shows that BCLAF1 and DBC1 are involved in the induction of apoptosis, though the algorithm has difficulties assigning a statistical weight using such a small dataset as the foreground. Three proteins have annotations in the cytoplasm and nucleus, and two proteins (HDGF and BCLAF1) are involved in DNA binding. An analysis of known physical interactions shows that none of these proteins directly interact with one another; this does not, however, necessarily imply that they cannot belong to any common pathways.

Table 3.3a. Upregulated phosphopeptides in VSV-infected K562 cells.

Gene name	Protein name	Sequence	Ratio
HDGF	Hepatoma-derived growth factor	GNAEGpSpSDEEGKLVIDEPAK	9.32
BCLAF1	Bcl-2-associated transcription factor 1	pYSPpSQNSPIHHIPSR	2.66
HN1	Hematological and neurological expressed 1 protein	RNSSEApSSGDFLDLK	2.62
KIAA1967 or DBC1	Deleted in breast cancer 1	pSVApSNQpSEMEFSSLQDMPK	2.54

Table 3.3b. Downregulated phosphopeptides in VSV-infected K562 cells.

Gene name	Protein name	Sequence	Ratio
SMN1	Survival motor neuron protein 1	GTGQpSDDpSDIWDDTALIK	0.57
MATR3	Matrin-3	RDpSFDDRGPSLNPVLDYDHGSR	0.58
HMGA1	High mobility group protein A1	KQPPVpSPGTALVGSQKEPSEVPpTPK	0.61
		KLEKEEEEGISQEpSpSEEEQ	0.63
LMNB2	Lamin-B2	LKLpSPpSPSSR	0.73

A functional enrichment of the four downregulated proteins show that they are mostly nuclear proteins (three out of four) involved in structural molecular activity (LMNB2, MATR3). The network illustrated in Figure 3.13 shows that these proteins have co-expression and physical interactions with each other and a family of nucleoporin proteins, consistent with them being nuclear proteins that play structural roles in the cell. These changes could be related to the effects of VSV on nuclear import/export, although the exact role of downregulation of phosphorylation on these proteins remains to be clarified.

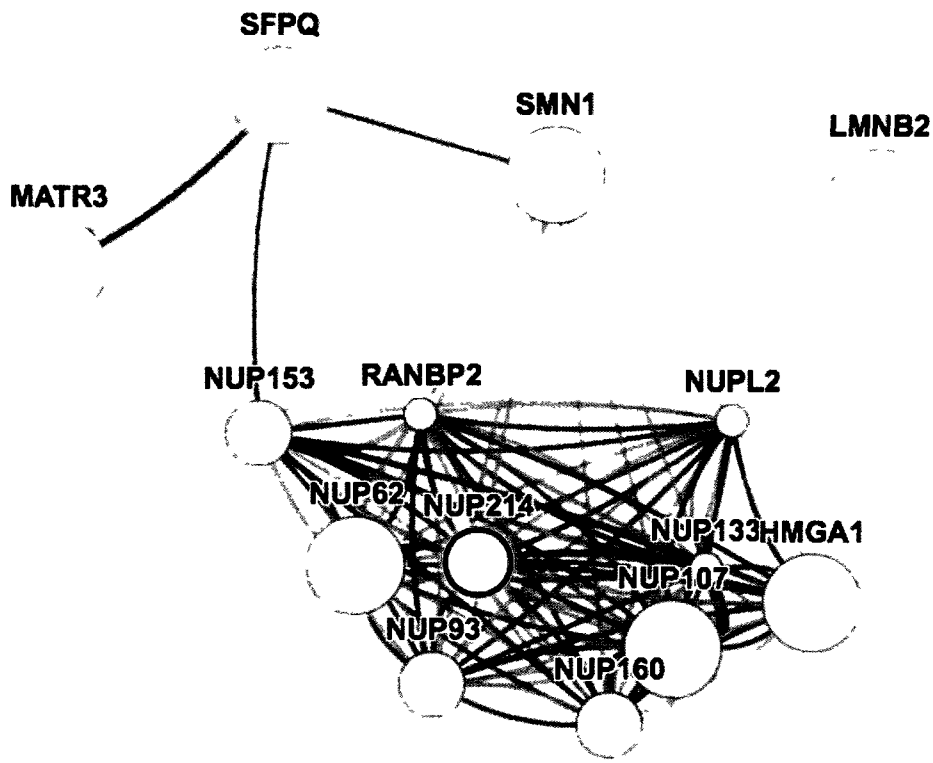


Figure 3.13. Protein-protein interaction network of proteins with downregulated phosphorylation sites. Image generated by GeneMania. Blue lines are physical interactions, green are pathways, purple are co-expression.

Identification of a differentially phosphorylated anti-apoptotic protein

Phosphorylation was found to be upregulated on the S132 and S133 residues of the anti-apoptotic protein hepatoma-derived growth factor (HDGF). HDGF is a heparin-binding growth factor that acts as a mitogen and stimulates growth in multiple cell types, including hepatoma cells, fibroblasts, endothelial, and smooth muscle cells.[60, 61] It is known to be overexpressed in human malignancies such as hepatocellular carcinoma, lung cancer, gastric cancer, melanoma, colorectal cancer, and pancreatic cancer.[62, 63, 64, 65] HDGF is an anti-apoptotic protein and blocking of this protein has been found to induce apoptosis in human colorectal carcinoma cells.[66] Apoptosis has been found to be mediated by both the Bad-mediated intrinsic[67] and the Fas-mediated extrinsic apoptotic pathways,[68] as shown in Figure 3.14. VSV induces oncolysis in cells via apoptotic pathways and it has been found that blocking BCL-2, another antiapoptotic protein, sensitizes primary chronic lymphocytic leukemia cells to these viruses.[69] BCL-2 prevents Bak-Bax oligerimization and disrupts the downstream caspase-dependent apoptotic effects previously described (i.e. through mitochondrial membrane permeabilization, cytochrome c release, caspase activation).

Information regarding the post-translational modifications on this protein remains limited. It has been determined that phosphorylation on S165 is necessary for the cellular secretion of HDGF; loss of this site results in N-terminal truncation of HDGF.[70] The effects of phosphorylation on other residues in this protein are currently not known. The solvated structure of the N-terminal region containing the

PWWP domain of the protein has yet to be determined. Without further structural investigation, the implications of upregulated phosphorylation on S125 and S126 remain theoretical. However, if these sites are involved in a similar biological process as S165, they may be needed for secretion of HDGF and imply a more active state of HDGF. A further examination of the data reveals that the S165 was in fact identified (with good y-series and 98 Da neutral loss), but in a SAX fraction (pH 6) rather than the IMAC fraction. Qualitatively it appears as though this phosphorylation site is upregulated, but unfortunately the phosphopeptide signal has a signal-to-noise ratio that is too low for quantitation.

A possible explanation for the upregulation of these phosphorylation sites could be that the K562 cells are responding to VSV infection by inducing anti-apoptotic proteins to prevent cell death. After 30 minutes of exposure to VSV, cells are only in the initial stages of infection; it is only after 24-48 hours that substantial amount of cell death occurs at 10 MOI.[69] The cells may be attempting to prevent death by upregulating the activity of HDGF in its early response to VSV-infection. After a longer VSV incubation time, cells would begin to undergo apoptosis and it is likely phosphorylation would be strongly downregulated in HDGF, as this process has been shown to occur rapidly and entirely during induced-apoptosis of endothelial cells.[71] This dephosphorylation occurs early on during the apoptotic process – HDGF was completely dephosphorylated after 1 hour while mitochondrial membrane permeabilization was only observed after 3 hours.

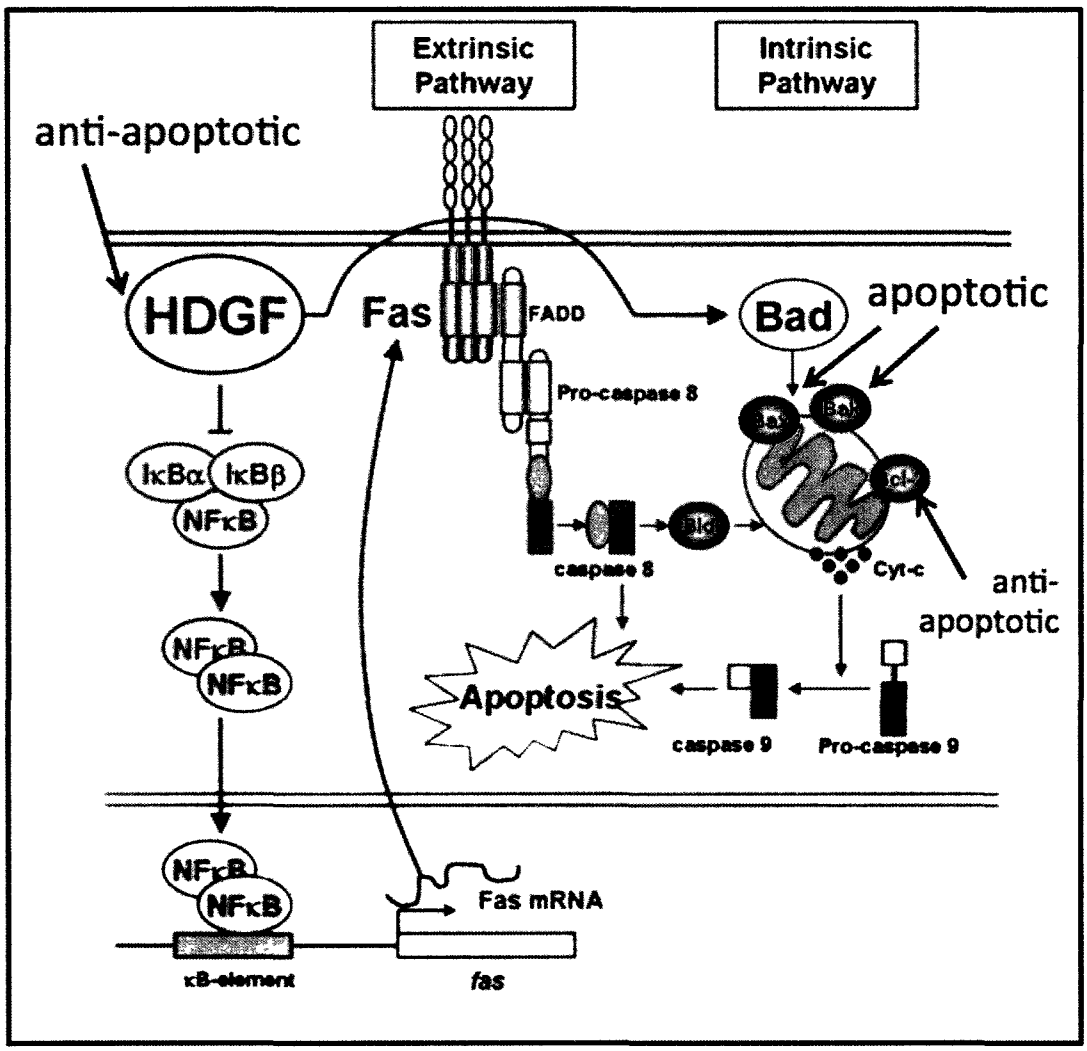


Figure 3.14. Apoptotic pathways controlled by HDGF. Adapted from Tsang *et al.*[68]

As BCL-2 was a successful target to increase apoptosis during infection of malignant cells by VSV, it may be possible that HDGF could be a successful target as well. Since HDGF is an upstream regulator of BCL-2 in the intrinsic apoptotic pathway, inhibiting it may have the same apoptotic effect as blocking HDGF. Furthermore, blocking HDGF may also increase apoptosis via the extrinsic pathway: without HDGF, IκBα is phosphorylated and degraded, releasing NFκB and resulting

in transcriptional activation of apoptotic genes. This is a clear advantage over BCL-2 as a target; BCL-2 inhibition is not in itself sufficient to induce apoptosis in malignant cell lines, indicating that cells have other anti-apoptotic proteins present. The role of HDGF is complicated somewhat by the pro-apoptotic nature of the protein; a gene silencing study of HDGF showed that HDGF is required for the release of pro-apoptotic factors Smac/Diablo from the mitochondria upon induced apoptosis by a stimulus.[72] This perhaps provides some reasoning to the upregulation of both HDGF phosphorylation and Diablo proteins in this work.

Evidence has shown that the BCL-2 protein family of can also regulate transcription, in addition to the role shown in Figure 3.14.[73] BCL-2 associated transcription factor (BCLAF1) has been shown to interact with the antiapoptotic BCL-2 and BCL-xL, but not the pro-apoptotic Bax; an overexpression of BCLAF1 is known to induce apoptosis.[74] It is believed that BCL-2 and BCL-xL typically bind BCALF1 and it is only during overexpression of BCLAF1 that it is translocated to the nucleus and thereby induces transcriptional activation of apoptotic genes. It is not clear what role phosphorylation has on the apoptotic properties of this protein; it has been found that BCLAF1 directly interacts with a protein kinase C (PKC) family member, induces TP53 mediated apoptosis, and is speculated as phosphorylation substrate of PKC.[75] A search of the phosphopeptide by the NetPhosK algorithm, however, did not show either upregulated phosphorylation site of BCLAF1 as a possible substrate of PKC. It can only be concluded that BCLAF1 and HDGF phosphorylation are playing a role in the apoptotic dynamics following VSV-infection; the exact mechanisms remain to be fully understood.

3.4 Conclusions

A novel proteomic and phosphoproteomic strategy for quantitative analyses of cellular samples has been developed. This workflow incorporates cellular lysis and proteolytic digestion, dimethyl labeling, and orthogonal separation of peptides by SCX/SAX and enrichment of phosphopeptides by IMAC for analysis by tandem mass spectrometry on a hybrid quadrupole-time-of-flight instrument. The strategy was successfully applied to characterize changes in the proteome and phosphoproteome of a chronic myelogenous leukemia cell line after infection by an oncolytic virus for 30 minutes. Results were compared against the background, non-labeled K562 proteome and phosphoproteome in which functional enrichment of gene ontology terms and phosphorylation motif discovery had been performed. Changes in the proteome and phosphoproteome of VSV-infected cell were identified by a custom in-house computational program derived from a manual XIC-based method. In the VSV-infected K562 cells, 54 proteins and 8 phosphorylation sites were found to be upregulated and 11 proteins and 9 phosphorylation sites were found to be downregulated compared to K562 control cells. Bioinformatics approaches were used to find the properties of these differentially regulated proteins and phosphorylation sites. It was found that groups of proteins with specific functionalities were differentially regulated, including molecular chaperones, ribonucleoproteins, nucleosome proteins and enzymes controlling the presence of reactive oxygen species. Several differentially regulated phosphorylation sites have been identified on proteins regulating the apoptotic pathways induced by VSV. Implications and potential explanations for differentially

regulated species have been discussed, as well as the performance characteristics of the quantitative method. Comparisons of the separation efficiency have been made and SAX has been found to have better separation efficiency than SCX for tryptic peptides as evidenced by fewer peptides appearing redundantly in multiple fractions. The effects of dimethyl labeling of N-termini and lysine residues on peptides have been evaluated and it was found that the labeling causes an altered peptide elution distribution compared to their non-dimethylated counterparts. Future work is discussed in the following chapter.

References

1. Xu, F., Wang, W. H., Tan, Y. J. & Bruening, M. L. Facile trypsin immobilization in polymeric membranes for rapid, efficient protein digestion. *Anal Chem* **82**, 10045-10051 (2010).

2. Zhou, H., Hou, W., Lambert, J. P., Tian, R. & Figeys, D. Analysis of low-abundance proteins using the proteomic reactor with pH fractionation. *Talanta* **80**, 1526-1531 (2010).

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. Burke, T. W., Mant, C. T., Black, J. A. & Hodges, R. S. Strong cation-exchange high-performance liquid chromatography of peptides. Effect of non-specific hydrophobic interactions and linearization of peptide retention behaviour. *J Chromatogr* **476**, 377-389 (1989).

5. Goodlett, D. R. et al. Differential stable isotope labeling of peptides for quantitation and de novo sequence derivation. *Rapid Commun Mass Spectrom* **15**, 1214-1221 (2001).

6. Salomon, A. R. et al. Profiling of tyrosine phosphorylation pathways in human cells using mass spectrometry. *Proc Natl Acad Sci U S A* **100**, 443-448 (2003).

7. Ndassa, Y. M., Orsi, C., Marto, J. A., Chen, S. & Ross, M. M. Improved immobilized metal affinity chromatography for large-scale phosphoproteomics applications. *J Proteome Res* **5**, 2789-2799 (2006).

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

9. Baxter, E. W. & Reitz, A. B. in *Organic Reactions* (John Wiley & Sons, Inc., Hoboken, NJ, USA, 2004).

10. Nie, S. et al. Comprehensive profiling of phosphopeptides based on anion exchange followed by flow-through enrichment with titanium dioxide (AFET). *J Proteome Res* **9**, 4585-4594 (2010).

11. Whitehead Institute for Biomedical Research. Venn Diagram Generator. *http://jura.wi.mit.edu/bioc/tools/venn.php*

12. Zhou, H., Hou, W., Lambert, J. P. & Figeys, D. New ammunition for the proteomic reactor: strong anion exchange beads and multiple enzymes enhance protein identification and sequence coverage. *Anal Bioanal Chem* **397**, 3421-3430 (2010).

13. Mann, M. et al. Analysis of protein phosphorylation using mass spectrometry: deciphering the phosphoproteome. *Trends Biotechnol* **20**, 261-268 (2002).

14. Bakalarski, C. E., Haas, W., Dephoure, N. E. & Gygi, S. P. The effects of mass accuracy, data acquisition speed, and search algorithm choice on peptide identification rates in phosphoproteomics. *Anal Bioanal Chem* **389**, 1409-1419 (2007).

15. Zhai, B., Villen, J., Beausoleil, S. A., Mintseris, J. & Gygi, S. P. Phosphoproteome analysis of Drosophila melanogaster embryos. *J Proteome Res* **7**, 1675-1682 (2008).

16. Boersema, P. J. et al. In-depth qualitative and quantitative profiling of tyrosine phosphorylation using a combination of phosphopeptide immunoaffinity purification and stable isotope dimethyl labeling. *Mol Cell Proteomics* **9**, 84-99 (2010).

17. Olsen, J. V. et al. Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* **127**, 635-648 (2006).

18. Turowski, M. et al. Deuterium isotope effects on hydrophobic interactions: the importance of dispersion interactions in the hydrophobic phase. *J Am Chem Soc* **125**, 13836-13849 (2003).

19. Nogales-Cadenas, R. et al. GeneCodis: interpreting gene lists through enrichment analysis and integration of diverse biological information. *Nucleic Acids Res* **37**, W317-22 (2009).

20. 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).

21. Villen, J., Beausoleil, S. A., Gerber, S. A. & Gygi, S. P. Large-scale phosphorylation analysis of mouse liver. *Proc Natl Acad Sci U S A* **104**, 1488-1493 (2007).

22. Schwartz, D. & Gygi, S. P. An iterative statistical approach to the identification of protein phosphorylation motifs from large-scale data sets. *Nat Biotechnol* **23**, 1391-1398 (2005).

23. Molina, H., Horn, D. M., Tang, N., Mathivanan, S. & Pandey, A. Global proteomic profiling of phosphopeptides using electron transfer dissociation tandem mass spectrometry. *Proc Natl Acad Sci U S A* **104**, 2199-2204 (2007).

24. Schlosser, A., Pipkorn, R., Bossemeyer, D. & Lehmann, W. D. Analysis of Protein Phosphorylation by a Combination of Elastase Digestion and Neutral Loss Tandem Mass Spectrometry. *Anal Chem* **73**, 170-176 (2001).

25. Cantin, R., Methot, S. & Tremblay, M. J. Plunder and stowaways: incorporation of cellular proteins by enveloped viruses. *J Virol* **79**, 6577-6587 (2005).

26. Moerdyk-Schauwecker, M., Hwang, S. I. & Grdzelishvili, V. Z. Analysis of virion associated host proteins in vesicular stomatitis virus using a proteomics approach. *Virol J* **6**, 166 (2009).

27. Her, L. S., Lund, E. & Dahlberg, J. E. Inhibition of Ran guanosine triphosphatase-dependent nuclear transport by the matrix protein of vesicular stomatitis virus. *Science* **276**, 1845-1848 (1997).

28. von Kobbe, C. et al. Vesicular stomatitis virus matrix protein inhibits host cell gene expression by targeting the nucleoporin Nup98. *Mol Cell* **6**, 1243-1252 (2000).

29. Petersen, J. M., Her, L. S. & Dahlberg, J. E. Multiple vesiculoviral matrix proteins inhibit both nuclear export and import. *Proc Natl Acad Sci U S A* **98**, 8590-8595 (2001).

30. Belov, G. A. et al. Early alteration of nucleocytoplasmic traffic induced by some RNA viruses. *Virology* **275**, 244-248 (2000).

31. Faria, P. A. et al. VSV disrupts the Rae1/mrnp41 mRNA nuclear export pathway. *Mol Cell* **17**, 93-102 (2005).

32. Sitterlin, D. Characterization of the Drosophila Rae1 protein as a G1 phase regulator of the cell cycle. *Gene* **326**, 107-116 (2004).

33. Pettit Kneller, E. L., Connor, J. H. & Lyles, D. S. hnRNPs Relocalize to the cytoplasm following infection with vesicular stomatitis virus. *J Virol* **83**, 770-780 (2009).

34. Kanlaya, R., Pattanakitsakul, S. N., Sinchaikul, S., Chen, S. T. & Thongboonkerd, V. Alterations in actin cytoskeletal assembly and junctional protein complexes in human endothelial cells induced by dengue virus infection and mimicry of leukocyte transendothelial migration. *J Proteome Res* **8**, 2551-2562 (2009).

35. Kress, E., Baydoun, H. H., Bex, F., Gazzolo, L. & Duc Dodon, M. Critical role of hnRNP A1 in HTLV-1 replication in human transformed T lymphocytes. *Retrovirology* **2**, 8 (2005).

36. Hay, D. C., Kemp, G. D., Dargemont, C. & Hay, R. T. Interaction between hnRNPA1 and IkappaBalpha is required for maximal activation of NF-kappaB-dependent transcription. *Mol Cell Biol* **21**, 3482-3490 (2001).

37. Cammas, A. et al. Cytoplasmic relocalization of heterogeneous nuclear ribonucleoprotein A1 controls translation initiation of specific mRNAs. *Mol Biol Cell* **18**, 5048-5059 (2007).

38. Zou, H., Henzel, W. J., Liu, X., Lutschg, A. & Wang, X. Apaf-1, a human protein homologous to C. elegans CED-4, participates in cytochrome c-dependent activation of caspase-3. *Cell* **90**, 405-413 (1997).

39. Kanlaya, R., Pattanakitsakul, S. N., Sinchaikul, S., Chen, S. T. & Thongboonkerd, V. Vimentin interacts with heterogeneous nuclear ribonucleoproteins and dengue nonstructural protein 1 and is important for viral replication and release. *Mol Biosyst* **6**, 795-806 (2010).

40. Field, B. N. et al. *Fields Virology (2-Volume Set)* (Lippincott Williams & Wilkins, 1996).

41. Bergmann, J. E., Tokuyasu, K. T. & Singer, S. J. Passage of an integral membrane protein, the vesicular stomatitis virus glycoprotein, through the Golgi apparatus en route to the plasma membrane. *Proc Natl Acad Sci U S A* **78**, 1746-1750 (1981).

42. Hammond, C. & Helenius, A. Folding of VSV G protein: sequential interaction with BiP and calnexin. *Science* **266**, 456-458 (1994).

43. Pratt, W. B. The hsp90-based chaperone system: involvement in signal transduction from a variety of hormone and growth factor receptors. *Proc Soc Exp Biol Med* **217**, 420-434 (1998).

44. Terasawa, K., Minami, M. & Minami, Y. Constantly updated knowledge of Hsp90. *J Biochem* **137**, 443-447 (2005).

45. Pick, E. et al. High HSP90 expression is associated with decreased survival in breast cancer. *Cancer Res* **67**, 2932-2937 (2007).

46. Workman, P., Burrows, F., Neckers, L. & Rosen, N. Drugging the cancer chaperone HSP90: combinatorial therapeutic exploitation of oncogene addiction and tumor stress. *Ann N Y Acad Sci* **1113**, 202-216 (2007).

47. Connor, J. H., McKenzie, M. O., Parks, G. D. & Lyles, D. S. Antiviral activity and RNA polymerase degradation following Hsp90 inhibition in a range of negative strand viruses. *Virology* **362**, 109-119 (2007).

48. Fan, Z., Beresford, P. J., Zhang, D. & Lieberman, J. HMG2 interacts with the nucleosome assembly protein SET and is a target of the cytotoxic T-lymphocyte protease granzyme A. *Mol Cell Biol* **22**, 2810-2820 (2002).

49. Matsumoto, K. et al. Stimulation of DNA transcription by the replication factor from the adenovirus genome in a chromatin-like structure. *J Biol Chem* **270**, 9645-9650 (1995).

50. Beresford, P. J. et al. Granzyme A activates an endoplasmic reticulum-associated caspase-independent nuclease to induce single-stranded DNA nicks. *J Biol Chem* **276**, 43285-43293 (2001).

51. Weintraub, H. & Van Lente, F. Dissection of chromosome structure with trypsin and nucleases. *Proc Natl Acad Sci U S A* **71**, 4249-4253 (1974).

52. Guilloton, F., Jean, C., de Thonel, A., Laurent, G. & Quillet-Mary, A. Granzyme B induction signalling pathway in acute myeloid leukemia cell lines stimulated by tumor necrosis factor alpha and Fas ligand. *Cell Signal* **19**, 1132-1140 (2007).

53. Fruehauf, J. P. & Meyskens, F. L. J. Reactive oxygen species: a breath of life or death? *Clin Cancer Res* **13**, 789-794 (2007).

54. Bae, Y. S. et al. Epidermal growth factor (EGF)-induced generation of hydrogen peroxide. Role in EGF receptor-mediated tyrosine phosphorylation. *J Biol Chem* **272**, 217-221 (1997).

55. Madesh, M. & Hajnoczky, G. VDAC-dependent permeabilization of the outer mitochondrial membrane by superoxide induces rapid and massive cytochrome c release. *J Cell Biol* **155**, 1003-1015 (2001).

56. Bauer, M. K., Schubert, A., Rocks, O. & Grimm, S. Adenine nucleotide translocase-1, a component of the permeability transition pore, can dominantly induce apoptosis. *J Cell Biol* **147**, 1493-1502 (1999).

57. Koren, I., Raviv, Z. & Shoshan-Barmatz, V. Downregulation of voltage-dependent anion channel-1 expression by RNA interference prevents cancer cell growth in vivo. *Cancer Biol Ther* **9**, 1046-1052 (2010).

58. Guo, Y. et al. Tim50, a component of the mitochondrial translocator, regulates mitochondrial integrity and cell death. *J Biol Chem* **279**, 24813-24825 (2004).

59. Wang, F. et al. A fully automated system with online sample loading, isotope dimethyl labeling and multidimensional separation for high-throughput quantitative proteome analysis. *Anal Chem* **82**, 3007-3015 (2010).

60. Everett, A. D., Lobe, D. R., Matsumura, M. E., Nakamura, H. & McNamara, C. A. Hepatoma-derived growth factor stimulates smooth muscle cell growth and is expressed in vascular development. *J Clin Invest* **105**, 567-575 (2000).

61. Oliver, J. A. & Al-Awqati, Q. An endothelial growth factor involved in rat renal development. *J Clin Invest* **102**, 1208-1219 (1998).

62. Iwasaki, T. et al. Hepatoma-derived growth factor as a prognostic marker in completely resected non-small-cell lung cancer. *Oncol Rep* **13**, 1075-1080 (2005).

63. Hu, T. H. et al. Expression of hepatoma-derived growth factor in hepatocellular carcinoma. *Cancer* **98**, 1444-1456 (2003).

64. Yamamoto, S. et al. Expression of hepatoma-derived growth factor is correlated with lymph node metastasis and prognosis of gastric carcinoma. *Clin Cancer Res* **12**, 117-122 (2006).

65. Ren, H. et al. Expression of hepatoma-derived growth factor is a strong prognostic predictor for patients with early-stage non-small-cell lung cancer. *J Clin Oncol* **22**, 3230-3237 (2004).

66. Liao, F., Dong, W. & Fan, L. Apoptosis of human colorectal carcinoma cells is induced by blocking hepatoma-derived growth factor. *Med Oncol* **27**, 1219-1226 (2010).

67. Tsang, T. Y. et al. Downregulation of hepatoma-derived growth factor activates the Bad-mediated apoptotic pathway in human cancer cells. *Apoptosis* **13**, 1135-1147 (2008).

68. Tsang, T. Y. et al. Mechanistic study on growth suppression and apoptosis induction by targeting hepatoma-derived growth factor in human hepatocellular carcinoma HepG2 cells. *Cell Physiol Biochem* **24**, 253-262 (2009).

69. Tumilasci, V. F. et al. Targeting the apoptotic pathway with BCL-2 inhibitors sensitizes primary chronic lymphocytic leukemia cells to vesicular stomatitis virus-induced oncolysis. *J Virol* **82**, 8487-8499 (2008).

70. Thakar, K. et al. Secretion of hepatoma-derived growth factor is regulated by N-terminal processing. *Biol Chem* **391**, 1401-1410 (2010).

71. Clermont, F. et al. HDGF is dephosphorylated during the early steps of endothelial cell apoptosis in a caspase-dependent way. *J Cell Biochem* **104**, 1161-1171 (2008).

72. Machuy, N. et al. A global approach combining proteome analysis and phenotypic screening with RNA interference yields novel apoptosis regulators. *Mol Cell Proteomics* **4**, 44-55 (2005).

73. Linette, G. P., Li, Y., Roth, K. & Korsmeyer, S. J. Cross talk between cell death and cell cycle progression: BCL-2 regulates NFAT-mediated activation. *Proc Natl Acad Sci U S A* **93**, 9545-9552 (1996).

74. Kasof, G. M., Goyal, L. & White, E. Btf, a novel death-promoting transcriptional repressor that interacts with Bcl-2-related proteins. *Mol Cell Biol* **19**, 4390-4404 (1999).

75. Liu, H., Lu, Z. G., Miki, Y. & Yoshida, K. Protein kinase C delta induces transcription of the TP53 tumor suppressor gene by controlling death-promoting factor Btf in the apoptotic response to DNA damage. *Mol Cell Biol* **27**, 8480-8491 (2007).

CHAPTER 4

Limitations and future directions

Future work based on this project falls into 3 distinct categories: experimental improvements, data analysis improvements, and hypothesis confirmation. Improvements to the experimental procedures would allow one to "dig deeper" into the proteome and phosphoproteome, increasing dataset sizes, identifying other potential regulated pathways, and coming closer to a full understanding of the molecular mechanisms of VSV infection and induced apoptosis. As the results in this work are preliminary and, to this author's knowledge, the first and only attempt at characterizing VSV-induced effects in a cancer cell line by shotgun proteomics, there is substantial room for improvement. The addition of replicate analyses typically greatly increases the number of identified peptides, a phenomenon that has been discussed in this work. Thus, a simple repetition of the experimental protocol may yield different identifications of differentially regulated proteins and phosphorylation sites. An increase in starting material will also greatly increase the number and confidence of phosphorylation site identification; a recent protocol has suggested the minimum amount of starting material for phosphoproteomics is 5 mg of protein, with a recommended starting amount of 15 mg,[1] in contrast to the 1-2 mg starting material used in this work. Furthermore, recent advances in mass spectrometry have allowed higher resolution, sensitivity, and instrument speed than the hybrid quadrupole time-of-flight mass spectrometer used in this work. For example, a recent quantitative phosphoproteomic study using dimethyl labeling, IMAC, and 2D fractionation prior to analysis by an LTQ-Orbitrap

mass spectrometer yielded >2,800 unique phosphorylation sites, despite starting with 1 mg of total protein content (compared with the 100 phosphorylation sites identified in this work).[2] An increase of an order of magnitude in identifications is substantial, especially when dealing with species that are of low abundance. With an increase in starting material in conjunction with an Orbitrap mass spectrometer, studies have shown that yet another order of magnitude in identified phosphorylation sites can be gained (i.e. >10,000 phosphorylation sites).[3,4] A corresponding increase in separation is needed when dealing with an increased amount of starting material from large proteomes; while the developments in this area are beyond the scope of this discussion, modern techniques could increase the orthogonality and resolution for peptide fractionation methods. Finally, a true nano-flow HPLC system would increase sensitivity and exploit the advantages of a true microfluidic system.

Further work is needed to elucidate the mechanisms of VSV-infection and induced apoptosis. A 30-minute infection with this virus and at this MOI is sufficient for preliminary studies, but by selection of multiple time points, pathways induced during viral progression can be more easily assessed. Multiple time points would provide insight into the time-elapsed progression of signaling cascades and whether different pathways are activated during different times during viral infection and replication. It may also be of use to infect various transformed cell lines to verify the commonality between apoptotic pathways induced by VSV as the virus has been shown to induce apoptosis in many cell lines.

With larger amounts of identified peptides and phosphorylation sites, an increase in automation is also necessary. Large datasets become increasingly difficult to manually validate or quantitate identifications and thus developments in computational proteomics allow better treatment of these datasets. A very early version of an automated quantitation algorithm has been developed for dimethyl-labeled peptides, but it still is not optimized; optimization should reduce the standard deviation of quantified peptides/phosphopeptides and increase confidence in those that are differentially expressed. Accuracy may also be improved by selecting a different quantitation type described in section 1.3.2. SILAC is perhaps the best choice as isotopically-labeled amino acids get directly incorporated into proteins *in vivo*, thereby eliminating variation introduced by sample handling steps or by chemical labeling.

While the data shown here suggest possible pathways in which VSV-induced infection and cell death may occur, it cannot be considered conclusive. Since the advent of mass spectrometry-based proteomics, there has been a certain degree of ambivalence felt by biologists and mass spectrometrists towards one another due to the disconnect between large-scale mass spectrometric analyses and traditional biochemical techniques. For instance, initial reports of biomarkers discovered by mass spectrometry-based serum proteomics have since been dismissed due to chemical artifacts, poor experimental design, and biochemical improbability.[5] Therefore, the discussion of regulated pathways in this work must not be taken as absolute conclusions; rather, they provide hypotheses that must be validated by biochemical techniques. Further evidence can be obtained by immunoprecipitation

of proteins of interest: a complex proteome is thereby reduced to a single protein (or several, depending on the specificity of the antibody used). The immunoprecipitated samples can be again analyzed by mass spectrometry, or a traditional biochemical technique can be used such as a Western blot. Protein expression can also be validated by mRNA microarrays, in which protein expression is indirectly tested through the analysis of mRNA produced prior to translation. This technique does not work for quantitative phosphorylation analyses, and mass spectrometry remains the most powerful technique for quantitative phosphoproteomics. None of these techniques, however, tell us the biological effects of a protein or phosphorylation on a protein. For these reasons, true validation of function does not occur until a gene knockdown or knockout study is done for a protein or site mutations are done on phosphorylated residues.

Mass spectrometry-based shotgun proteomics is a very useful tool for generating large amounts of data and elucidating protein networks and signaling pathways. It can provide valuable hypotheses for further biochemical studies. While the technique does suffer from some drawbacks, technical advancements in mass spectrometry, chromatographic resins, and statistical tools allow researchers to probe deeper and deeper into the proteome and phosphoproteome to decipher the complex molecular mechanisms of the cell. It is hoped that this fundamental knowledge can suggest new therapeutic targets and eventually have an impact on human health and disease treatment.

References

1. Villen, J. & Gygi, S. P. The SCX/IMAC enrichment approach for global phosphorylation analysis by mass spectrometry. *Nat Protoc* **3**, 1630-1638 (2008).

2. Wu, C. J., Chen, Y. W., Tai, J. H. & Chen, S. H. Quantitative phosphoproteomics studies using stable isotope dimethyl labeling coupled with IMAC-HILIC-nanoLC-MS/MS for estrogen-induced transcriptional regulation. *J Proteome Res* **10**, 1088-1097 (2011).

3. Zhai, B., Villen, J., Beausoleil, S. A., Mintseris, J. & Gygi, S. P. Phosphoproteome analysis of Drosophila melanogaster embryos. *J Proteome Res* **7**, 1675-1682 (2008).

4. Villen, J., Beausoleil, S. A., Gerber, S. A. & Gygi, S. P. Large-scale phosphorylation analysis of mouse liver. *Proc Natl Acad Sci U S A* **104**, 1488-1493 (2007).

5. Diamandis, E. P. Analysis of Serum Proteomic Patterns for Early Cancer Diagnosis: Drawing Attention to Potential Problems. *J Natl Cancer Inst* **96**, 353-356 (2004).

APPENDICES

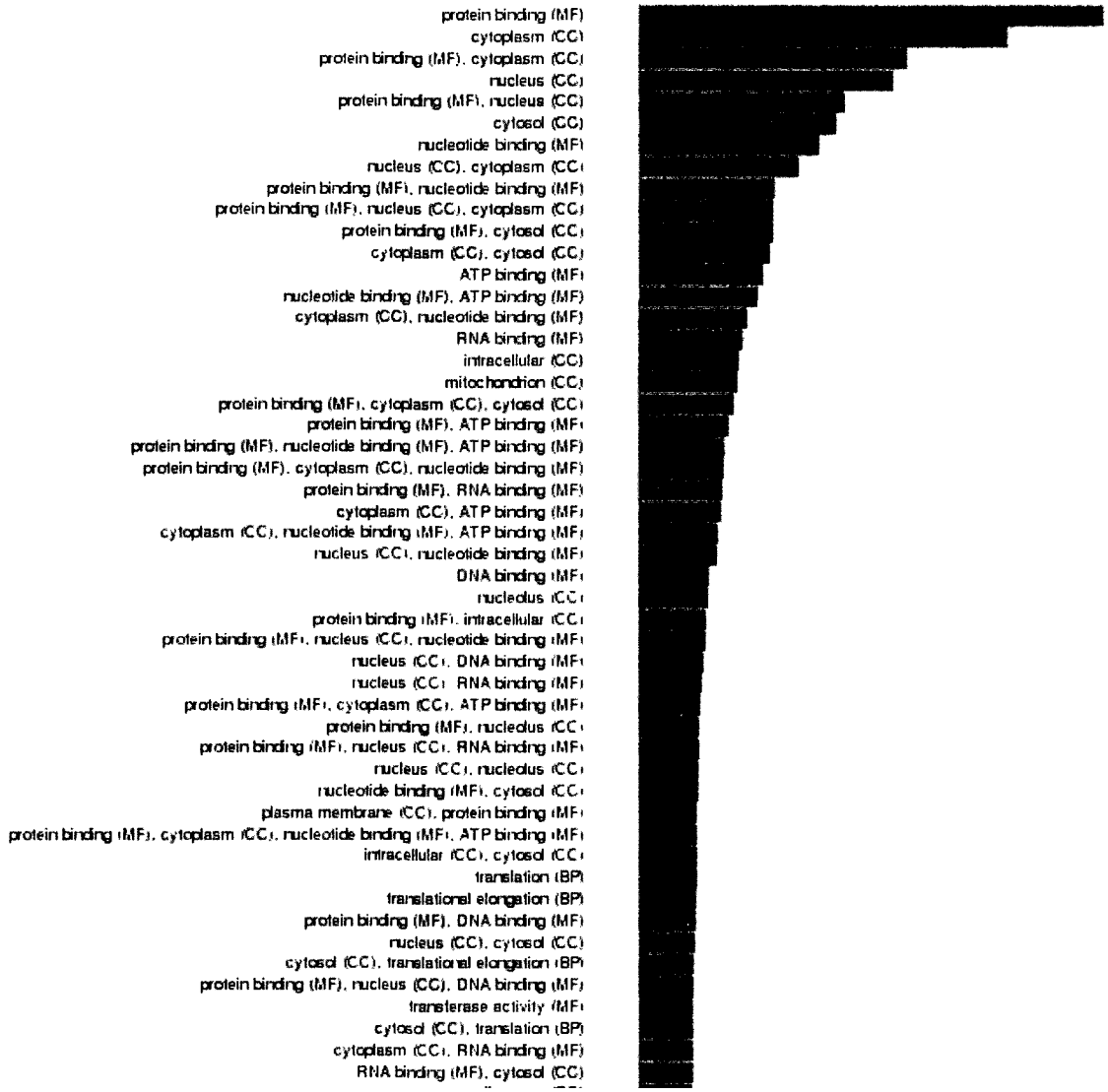
Appendix 1. Exhaustive list of functionally enriched gene annotations in the background proteome.

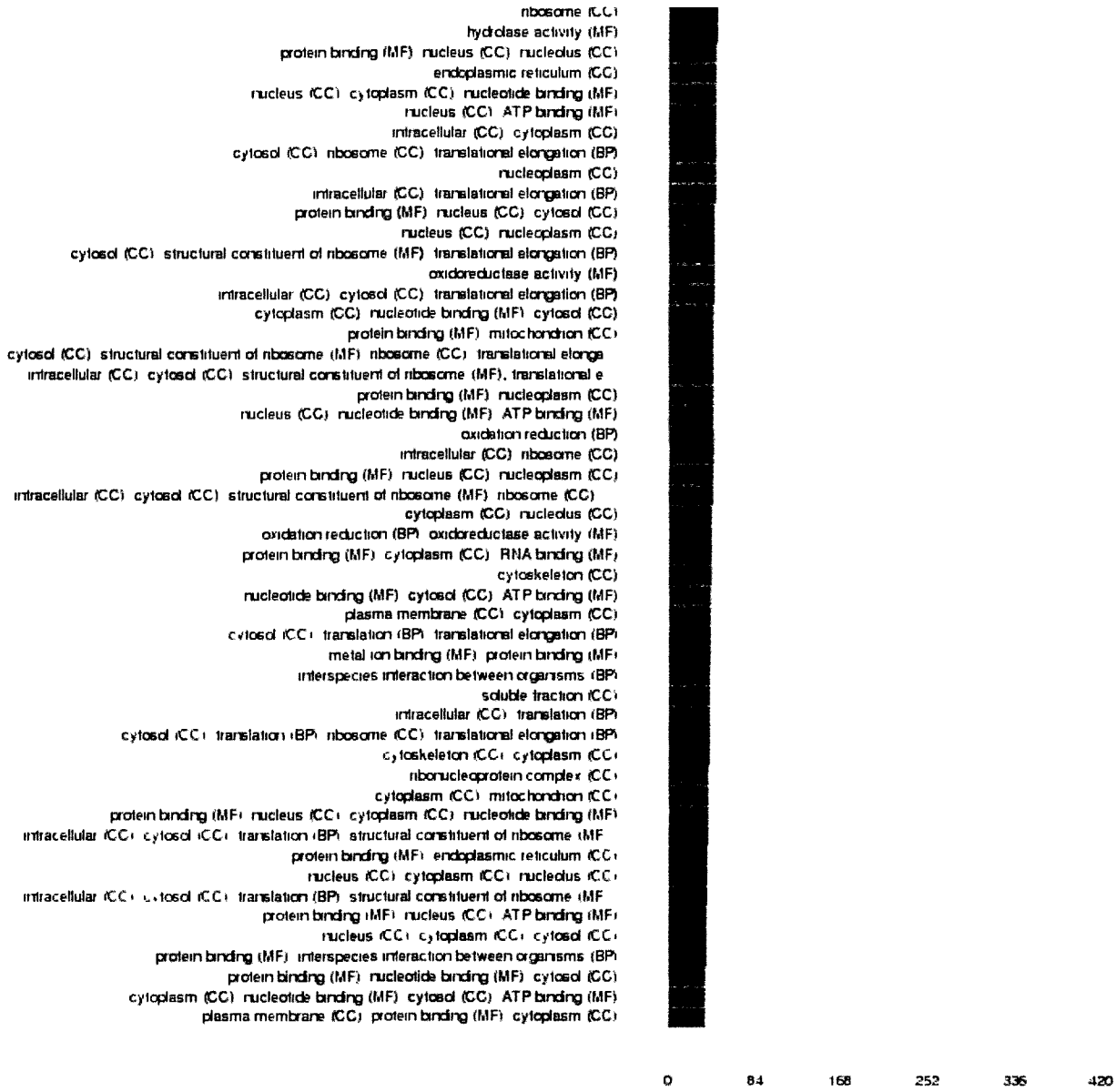
Appendix 2. Centered, 13-amino acid peptide sequences and their corresponding phosphorylation motifs.

Appendix 3. Complete list of upregulated and downregulated proteins found in VSV-infected K562 cells relative to control K562 cells.

Appendix 4. Exhaustive list of proteins identified in background proteome by gene name.

Appendix 1


protein binding (MF)
cytoplasm (CC)
protein binding (MF), cytoplasm (CC)
nucleus (CC)
protein binding (MF), nucleus (CC)
cytosol (CC)
nucleotide binding (MF)
nucleus (CC), cytoplasm (CC)
protein binding (MF), nucleotide binding (MF)
protein binding (MF), nucleus (CC), cytoplasm (CC)
protein binding (MF), cytosol (CC)
cytoplasm (CC), cytosol (CC)
ATP binding (MF)
nucleotide binding (MF), ATP binding (MF)
cytoplasm (CC), nucleotide binding (MF)
RNA binding (MF)
intracellular (CC)
mitochondrion (CC)
protein binding (MF), cytoplasm (CC), cytosol (CC)
protein binding (MF), ATP binding (MF)
protein binding (MF), nucleotide binding (MF), ATP binding (MF)
protein binding (MF), cytoplasm (CC), nucleotide binding (MF)
protein binding (MF), RNA binding (MF)
cytoplasm (CC), ATP binding (MF)
cytoplasm (CC), nucleotide binding (MF), ATP binding (MF)
nucleus (CC), nucleotide binding (MF)
DNA binding (MF)
nucleolus (CC)
protein binding (MF), intracellular (CC)
protein binding (MF), nucleus (CC), nucleotide binding (MF)
nucleus (CC), DNA binding (MF)
nucleus (CC), RNA binding (MF)
protein binding (MF), cytoplasm (CC), ATP binding (MF)
protein binding (MF), nucleolus (CC)
protein binding (MF), nucleus (CC), RNA binding (MF)
nucleus (CC), nucleolus (CC)
nucleotide binding (MF), cytosol (CC)
plasma membrane (CC), protein binding (MF)
protein binding (MF), cytoplasm (CC), nucleotide binding (MF), ATP binding (MF)
intracellular (CC), cytosol (CC)
translation (BP)
translational elongation (BP)
protein binding (MF), DNA binding (MF)
nucleus (CC), cytosol (CC)
cytosol (CC), translational elongation (BP)
protein binding (MF), nucleus (CC), DNA binding (MF)
transferase activity (MF)
cytosol (CC), translation (BP)
cytoplasm (CC), RNA binding (MF)
RNA binding (MF), cytosol (CC)


ribosome (CC)
hydrolase activity (MF)
protein binding (MF) nucleus (CC) nucleolus (CC)
endoplasmic reticulum (CC)
nucleus (CC) cytoplasm (CC) nucleotide binding (MF)
nucleus (CC) ATP binding (MF)
intracellular (CC) cytoplasm (CC)
cytosol (CC) ribosome (CC) translational elongation (BP)
nucleoplasm (CC)
intracellular (CC) translational elongation (BP)
protein binding (MF) nucleus (CC) cytosol (CC)
nucleus (CC) nucleoplasm (CC)
cytosol (CC) structural constituent of ribosome (MF) translational elongation (BP)
oxidoreductase activity (MF)
intracellular (CC) cytosol (CC) translational elongation (BP)
cytoplasm (CC) nucleotide binding (MF) cytosol (CC)
protein binding (MF) mitochondrion (CC)
cytosol (CC) structural constituent of ribosome (MF) ribosome (CC) translational elonga
intracellular (CC) cytosol (CC) structural constituent of ribosome (MF), translational e
protein binding (MF) nucleoplasm (CC)
nucleus (CC) nucleotide binding (MF) ATP binding (MF)
oxidation reduction (BP)
intracellular (CC) ribosome (CC)
protein binding (MF) nucleus (CC) nucleoplasm (CC)
intracellular (CC) cytosol (CC) structural constituent of ribosome (MF) ribosome (CC)
cytoplasm (CC) nucleolus (CC)
oxidation reduction (BP) oxidoreductase activity (MF)
protein binding (MF) cytoplasm (CC) RNA binding (MF)
cytoskeleton (CC)
nucleotide binding (MF) cytosol (CC) ATP binding (MF)
plasma membrane (CC) cytoplasm (CC)
cytosol (CC) translation (BP) translational elongation (BP)
metal ion binding (MF) protein binding (MF)
interspecies interaction between organisms (BP)
soluble fraction (CC)
intracellular (CC) translation (BP)
cytosol (CC) translation (BP) ribosome (CC) translational elongation (BP)
cytoskeleton (CC) cytoplasm (CC)
ribonucleoprotein complex (CC)
cytoplasm (CC) mitochondrion (CC)
protein binding (MF) nucleus (CC) cytoplasm (CC) nucleotide binding (MF)
intracellular (CC) cytosol (CC) translation (BP) structural constituent of ribosome (MF
protein binding (MF) endoplasmic reticulum (CC)
nucleus (CC) cytoplasm (CC) nucleolus (CC)
intracellular (CC) cytosol (CC) translation (BP) structural constituent of ribosome (MF
protein binding (MF) nucleus (CC) ATP binding (MF)
nucleus (CC) cytoplasm (CC) cytosol (CC)
protein binding (MF) interspecies interaction between organisms (BP)
protein binding (MF) nucleotide binding (MF) cytosol (CC)
cytoplasm (CC) nucleotide binding (MF) cytosol (CC) ATP binding (MF)
plasma membrane (CC) protein binding (MF) cytoplasm (CC)
0
84
168
252
336
420

Appendix 2

Motifs		
Proline-directed (pS-P)	**Acidic (pS-X-X-E)**	**Basic (R-X-X-pS)**
LDLISESPIKGRA	KIEDVGSDEEDDS	GLMRTASELLLDR
KATVTPSPVKGKG	KGAGDGSDEEVDG	LRIRTPSGEFLER
KKSLSGSPLKVKN	KKTKFASDDEHDE	RLVRSRSSDIVSS
KDLFDYSPPLHKN	KLSVPTSDEEDEV	RHFRRDSFDDRGP
LAPAKESPRKGAA	KQPLLLSEDEEDT	RLERQISQDVKLE
AGSRSSSPGKLLG	RGLLYDSDEEDEE	KIERSASEPSLHR
VPEFPLSPPKKKD	KEELMSSDLEETA	KKTRTDSTSDGRP
GLARLSSPVLHRL	EEPTVYSDEEEPK	GVKRSRSGEGEVS
DEILNRSPRNRKP	PGPNIESGNEDDD	SRKRKLSGDQPAA
GDLLEDSPKRPKE	VKATNESEDEIPQ	PVVRTPSIQPSLL
LSPRERSPALKSP	AVQRQGSGRESPS	
AKKVVVSPTKKVA	PFRRSDSFPERRA	
GGLRPESPESLTS	MNGRKQSLGELIG	
DRERASSPPDRID	SPHCKRSEAEAKP	
PLDRCRSPGMLEP		
GQPRQGSPVAAGA		
KSKAPGSPLSSEG		
NLLSGKSPRKSAE		
RTAQVPSPPRGKI		
SGGRLGSPKPERQ		
NDPLTSSPGRSSR		
PKPKIESPKLERT		
TSERAPSPSSRMG		
PPARKASPPSGLW		
FAKALESPERPFL		
ATRPSPSPERSST		
PYQSRGSPHYFSP		
SEKSPRSPQLSDF		
EGHSNSSPRHSEA		
MLLLSRSPEAQPK		
RSPALKSPLQSVV		
HRRIDISPSTFRK		
LQGPRGSPNARSQ		
KHTGPNSPDTAND		
FSRPSMSPTPLDR		
DYRGRLSPVPVPR		

Appendix 3

Upregulated proteins in VSV-infected K562 cells

Gene name	Protein name	Average L/H	No. of peptides
CHRD	Chordin	12.23	1
BTK	Bruton agammaglobulinemia tyrosine kinase	12.03	1
MARCKS	Myristoylated alanine-rich protein kinase C substrate	9.29	1
OSBP	Oxysterol binding protein	9.07	1
BECN1	Beclin 1	8.74	1
HSP90B1	Heat shock protein 90kDa beta (Grp94), member 1	7.63	1
HNRNPA3	Heterogeneous nuclear ribonucleoprotein A3	6.75	1
SFRS5	Serine/arginine-rich splicing factor 5	6.56	2
HNRNPA1	Heterogeneous nuclear ribonucleoprotein A1	6.17	3
SSB	Sjogren syndrome antigen B	6.09	2
HNRNPA2B1	Heterogeneous nuclear ribonucleoprotein A2/B1	5.99	2
HMGN2	High-mobility group nucleosomal binding domain 2	5.18	2
H2AFV	H2A histone family, member V	5.17	1
HNRNPAB	Heterogeneous nuclear ribonucleoprotein A/B	4.59	1
HSPA8	Heat shock 70kDa protein 8	4.58	4
SFRS13A	Serine/arginine-rich splicing factor 13 A	4.56	1
PSMD6	Proteasome (prosome, macropain) 26S subunit, non-ATPase, 6	4.47	1
SND1	Staphylococcal nuclease and tudor domain containing 1	4.43	2
DTNB	Dystrobrevin, beta	4.41	1
HIST1H4A	Histone cluster 1, H4a	4.28	1
C1QBP	Complement component 1, q subcomponent binding protein	4.26	2
CALU	Calumenin	4.11	3
ST13	Suppression of tumorigenicity 13 (colon carcinoma) (Hsp70 interacting protein)	4.07	1

Gene name	Protein name	Average L/H	No. of peptides
CSPG5	Chondroitin sulfate proteoglycan 5 (neuroglycan C)	3.91	1
EEF1A	Eukaryotic translation elongation factor 1 alpha 1	3.90	3
LDHA	Lactate dehydrogenase A	3.90	4
RPLP2	Ribosomal protein, large, P2	3.85	1
HSP90AB1	Heat shock protein 90kDa alpha (cytosolic), class B member 1	3.81	8
RPL4	Ribosomal protein, large, P2	3.71	2
RAP1B	RAP1B, member of RAS oncogene family	3.70	1
PRDX6	Peroxiredoxin 6	3.69	3
DNAJC9	DnaJ (Hsp40) homolog, subfamily C, member 9	3.62	1
XRCC6	X-ray repair complementing defective repair in Chinese hamster cells 6	3.60	2
SET	SET nuclear oncogene	3.58	3
HISTH2AA	Histone cluster 1, H2aa	3.49	1
RBBP4	Retinoblastoma binding protein 4	3.46	1
RPS12	Ribosomal protein S12	3.40	2
TXN	Thioredoxin	3.34	1
TRPS1	Zinc finger transcription factor	3.25	1
SOD1	Superoxide dismutase 1, soluble	3.23	1
HNRNPA2	Heterogeneous nuclear ribonucleoprotein A2	3.20	1
TIM50	Translocase of inner mitochondrial membrane 50 homolog	3.19	1
CKB	Creatine kinase, brain	3.06	1
OAT	Ornithine aminotransferase	3.00	1
THRAP3	Thyroid hormone receptor associated protein 3	2.97	1
UBE2V1	Ubiquitin-conjugating enzyme E2 variant 1	2.97	2
SFRS6	Serine/arginine-rich splicing factor 6	2.87	1
HNRPA0	Heterogeneous nuclear ribonucleoprotein A0	2.76	1
MCM3	Minichromosome maintenance complex component 3	2.75	1
LMNA	Lamin A/C	2.68	6
DIABLO	Diablo protein	2.61	1

Gene name	Protein name	Average L/H	No. of peptides
ALDOC	Aldolase C, fructose-bisphosphate	2.58	3
CORO1C	Coronin, actin binding protein, 1C	2.56	1

Downregulated proteins in VSV-infected K562 cells

Gene name	Protein name	Average L/H	No. of peptides
GGCT	Gamma-glutamylcyclotransferase	0.50	1
PRPH	Peripherin	0.49	1
TLN1	Talin 1	0.46	1
NONO	Non-POU domain containing, octamer-binding	0.46	1
PSMC2	Proteasome (prosome, macropain) 26S subunit, ATPase, 2	0.41	1
PABPC1	Poly(A) binding protein, cytoplasmic 1	0.35	1
KIAA1683	KIAA1683	0.35	1
VDAC1	Voltage-dependent anion channel 1	0.34	1
H2AFV	H2A histone family, member V	0.27	1
ZNF268	Zinc finger protein 268	0.17	1
GSTP1	Glutathione S-transferase pi 1	0.14	1

Appendix 4

HSP90AA1	TUBB6	PHGDH	RAB10	PSMA5	IGF2BP2	SHMT2
HSPA8	UBA1	PA2G4	CANX	RPS18	SNORA41	HADH
HSP90AB1	GDI2	HSPB1	ILK2	PAICS	ATP5J2	UBE2M
ENO1	PPIB	HBG1	GSTO1	RPL26L1	RPS28	RAB11B
TUBB	STIP1	KHSRP	MAGEB2	RPL22	RPS15A	MCM3
TUBB2C	PDIA4	CFL1	LRPPRC	TCP1	CALU	RPS7
HSPD1	PARK7	PFN1	HIST1H1C	LARS	NARS	USP5
HSPA5	LDHB	RPS16	HNRNPK	MCM2	RBBP4	HIST1H1B
EEF2	SSB	ACADM	PSMD11	ADH5	RPL8	KARS
TUBB2A	TKT	HADHB	PRDX3	SET	HNRPDL	FASN
CCT8	CCT2	IGF2BP1	NASP	HADHA	UBE2V1	SARNP
PDIA3	PGK1	PARP1	RPS13	HNRNPAB	PRDX5	LGALS1
HSPA9	ALDH1A2	GNB2L1	YWHAB	MDH1	CACYBP	IARS2
LCP1	ALDOA	YWHAH	NUDC	GLOD4	IGLV214	PRDX4
ANXA1	YARS	TUT1	AHCY	RPS14	CYCS	DECR1
HSPA1B	CCT5	ANXA5	FAH	NAP1L1	SYNCRIP	LONP1
GAPDH	HSPE1	PDIA6	HNRNPD	TRIM28	DCTPP1	RUVBL1
HSPA1A	XRCC6	EEF1D	SFPQ	SLC25A5	RPL23A	TROVE2
MDH2	GANAB	RPS3	GART	CS	PABPC1	ESYT1
ACTB	EIF4A1	G6PD	RPS6	RPL10A	ANP32B	EIF3H
PRDX6	EZR	LOC100294459	TMPO	PRKDC	EIF5A2	PSMC2
ATP5B	ATP5A1	LRRC59	ALB	VIM	MTHFD1	SNORD84
EEF1A1	ATIC	ASNS	RANBP1	DUT	ACAT1	EIF5
TUBA1B	CCT7	EPRS	HBZ	CLIC1	HSD17B10	BCAP31
PPIA	FUBP1	RPL18	HNRNPU	NPM1	MIR1181	RAP1B
VCP	WARS	GSTP1	SNAP29	ETFA	PDCD5	PSME3
TPI1	YWHAG	KPNB1	SERPINH1	FH	NANS	ABCF2
ANXA2	VARS	FSCN1	RPL5	TFAM	ILF3	ARPC4
PRDX1	HBE1	RAN	TARS	PFDN2	RPS3A	SNX3
PKM2	HSP90B1	RPL7	GOT1	CTSD	RPL36	UROD
ACTN4	HYOU1	YWHAZ	HPRT1	GLUD1	TIMM44	SNRPB
SND1	CKB	LDHA	AK2	SARS	ETF1	RPL27
TAGLN2	YWHAE	TRAP1	RPL4	HIST1H4E	CAND1	PPA1
AARS	PRDX2	RPL13	LAP3	UBE2V2	RPS9	PTBP1
TUBA4A	RPS19	MSN	GFPT1	HSPA4	RPS25	HMGB2
PGD	CCT3	SUB1	HIST2H2BF	ETFB	COX5A	RUVBL2
CCT6A	HBG2	ACLY	RPL6	GPI	YBX1	RBM39
ACTA1	TUFM	NAP1L4	APRT	PDCD6IP	HARS	LASP1
NCL	P4HB	PRKCSH	ATP5O	ERP29	PCBP1	HMGB1
VCL	SNORA62	NME1NME2	RPS2	HNRNPH3	ACO2	PCK2

CSTB	CSTB	PSMD14	DNAJB1	MCM6	RPL12
MCM7	MCM7	UCHL5	PSMD2	LOC728825	CAP1
MYL12B	MYL12B	RAB5C	NAMPT	PSAT1	RPL23
SLC9A3R1	SLC9A3R1	PDHB	PSMA3	SFRS1	FGA
ATP5D	ATP5D	MAPRE1	FTH1	PGM2	C14orf93
SNORD38B	SNORD38B	HBA2;HBA1	HP1BP3	IARS	HINT1
LOC653881	LOC653881	CSE1L	PGRMC1	NUDT5	SNRPA
OLA1	OLA1	UBE2I	PSMB6	SRM	CTSB
RALA	RALA	PCNA	RNH1	NMT1	DNAJC9
NAA15	NAA15	RPS21	SERPINA1	ACP1	SSBP1
SF3B3	SF3B3	AIFM1	TPM1	WDR1	ARFGAP1
PSMA4	PSMA4	CPOX	KRT9	KRT18	MOBKL1B
HNRNPA2B1	HNRNPA2B1	RPLP2	SEP11	PRKCB	SAR1B
TCEA1	TCEA1	FKBP2	SOD2	PSMC5	USP24
RPL30	RPL30	PTGES3	MARS	BLMH	SERPINB6
DLD	DLD	PSMC3	RPS10	TXNDC12	GARS
RPL27A	RPL27A	ALDOC	INTS3	HMOX2	APEH
PSMA1	PSMA1	ACTR2	PSMD5	ERP44	LMAN1
SNORD24	SNORD24	RAD23B	CLNS1A	LUC7L2	GCN1L1
PCBP2	PCBP2	HNRNPM	XPO5	DYNC1H1	FKBP3
EIF3J	EIF3J	EIF4B	MARCKSL1	LYPLA1	DNM2
CDK1	CDK1	BZW2	SNRPD1	SEC61B	RPS15
RARS	RARS	CRKL	MIF	NPEPPS	CMPK1
TLN1	TLN1	PEBP1	SERPINB9	CKAP5	SBDS
CALR	CALR	DRG1	CNDP2	C11orf80	ILF2
PSMB1	PSMB1	TNPO1	PSMA7	RPS23	THRAP3
RPL19	RPL19	ZC3H18	UBE2N	RPL24	CAPG
EIF4E	EIF4E	LSM3	RCC2	DARS2	CACNA1C
ST13	ST13	UBE2K	SUGT1	UNC13D	EIF2S2
XPO1	XPO1	PSMA6	RPL11	ATP5H	OTUB1
ARHGDIA	ARHGDIA	EIF4G2	SLC25A3	CALB1	VASP
GOT2	GOT2	RPA2	HNRNPF	TGM2	SCP2
TALDO1	TALDO1	SNRNP70	MYL4	NACAP1	PCNP
RFC4	RFC4	UBE2L3	PDXP	XKR3	RPL9
PPA2	PPA2	FDXR	LETM1	ZFYVE20	LMNA
CBS	CBS	PSMD4	ARF1	FEN1	RPL17
CRYZ	CRYZ	APEX1	PAFAH1B3	HK1	HDGF
SRPRB	SRPRB	EIF2S1	ARMC1	DENR	DSTN
PPP2R1A	PPP2R1A	RPL38	GSPT1	HP	CAPN1
IDH1	IDH1	DDX1	CAPZA1	TTN	CHMP4B

HPCA	C9orf64	S100A6	CPT1A	EWSR1	C12orf28
BAT3	PSMD6	ZNF101	TXN	FTL	MSH6
SNRPE	RPN1	ZNF259	FLJ45684	VWA3A	LOC729860
SRRM1	IPO7	SDF4	LOC100290023	MSH2	ERCC6L
EIF3L	SFRS3	SAMSN1	GLRX3	SENP2	
DDX42	A2M	HUS1B	DDB1	IGHD	
UBA2	ARL8B	TMEM109	SERBP1	PPIF	
PPP1CC	TPD52L2	EIF2B1	ASPDH	NCSTN	
STAT1	C1QBP	KDM6A	HM13	LOC100290973	
BTF3L4	IMPDH2	RPS5	FKBP5	CEP110	
SFRS2	HEXB	DOCK6	FANCB	LMAN2	
CISD2	FAM129A	LOC100049716	NSFL1C	STAT5A	
DYNC1I2	CUTA	XRCC5	INMT	REXO2	
RLTPR	GMPPA	USP25	C10orf41	TMX1	
EMD	DHX9	LOC100293209	RPLP1	TTC3	
PSMC4	NONO	SEP9	IFIT2	UBXN4	
G3BP1	PRKAR1A	RPS4X	DSTYK	PSMD3	
NAA50	COTL1	BCLAF1	KRT80	DRAP1	
RPS20	C14orf166	PROX1	AFAP1	AFAP1L1	
RPLP0	AK1	EIF4H	H2AFV	STXBP2	
SERPINB1	GNAL	COMT	VAPB	FBXW5	
PRPS1	TOP1	WHSC1L1	PPM1F	TXNDC17	
GSR	MTERFD2	EIF6	ADRM1	GMPS	
CARS	KCNH5	GLUL	TCEA3	BRPF1	
CAPZB	RNASET2	HMBS	MPP1	CAT	
GLO1	PNP	SRPR	NOD2	PNPT1	
ARPC3	ACAT2	WDR41	WNT7A	FKBP4	
ADCY8	BCAT1	KIAA1383	CDK11A	ZNF638	
DNAJC7	MAT2A	LOC100288655	PGAM2	AGPAT3	
NDUFA5	SLC25A29	CAPNS1	CUL7	KBTBD2	
HMGA1	CORO1C	DDX21	ZBTB9	LOC619207	
CMBL	ASNA1	ANLN	ZYX	CLEC4M	
PAFAH1B2	SPTAN1	RPL15	ARHGAP28	AKR7A2	
STMN1	PIP5K1C	RPA3	ZNF233	DHX15	
OXSR1	KLHL20	TIPRL	C4A	VAT1	
NAA10	UBB	HSF2BP	SRP9L1	SMARCAL1	
RPL35A	GRSF1	ATP1A2	CDYL	CRHR1	
RCC1	PICALM	UFM1	LSM4	CAMSAP1L1	
KRT19	LSM6	MFSD8	DNAJA1	EFHA1	
CCDC141	UAP1	STIM1	DBI	FNBP4	