STUDIES OF PROTEIN FUNCTION BY LIQUID CHROMATOGRAPHY-MASS SPECTROMETRY

Gwillym Declan Williams

A dissertation submitted to the Faculty of Graduate Studies in the partial fulfillment of the requirements for the degree of Doctor of Philosophy

Graduate Program in Chemistry
York University
Toronto, Ontario

June 2014

Abstract

Complete genomes of many organisms have been recorded using high throughput nucleic acid sequencing; however the intricate functions of the gene products, namely proteins, remain largely undiscovered. As the most prominent technology for polypeptide sequence determination and quantification, tandem mass spectrometry is used extensively to address questions regarding the behaviours of proteins in the biological context. The capacity of mass spectrometers to measure covalent modifications of amino acids is also of great value in the investigation of biological processes since specific reactive sites within proteins modulate their activity. Integration of liquid chromatography into mass spectrometry platforms has greatly improved their sensitivity and throughput and is of particular benefit in the analysis of highly complex polypeptide mixtures. Studies of the behaviours of individual intracellular proteins in bacterial, plant and metazoan systems employing liquid chromatography-mass spectrometry are described herein. Amino acid phosphorylation, a type of covalent modification which can switch a protein between inactive and active states, was investigated in detail in each system. The phosphorylation of individual amino acid residues and resulting alterations in function were determined in GraR and VraR, two antibiotic resistance factors of the human pathogen *Staphylococcus aureus*. Such information improves understanding of bacterial antibiotic resistance. Three phosphosites of starch branching enzyme IIb which appear to regulate starch biosynthesis in *Zea mays* (corn) as well as kinases for which these sites are putative substrates were identified. The definition of such molecular drivers of starch formation could inform the generation of novel cereal varieties capable of making novel starch

types. Regulation of the highly conserved multifunctional protein Beta-catenin by p38 mitogen activated protein kinase was explored. Gene products with affinity for Beta-catenin, including several not previously known to interact, were identified from *Rattus norvegicus* (rat) smooth muscle cells. Since Beta-catenin is integral to animal growth and repair, mapping its interaction network has broad implications for the advancement of medicine. The specificity of peptide ion mass and peptide fragment ion mass relative to instrumental mass accuracy and the consequence for tandem-mass spectrometry-based quantification are explored mathematically in a separate chapter. *In silico* comparison of peptide ion/product ion mass pairs calculated from the human proteome revealed a range of mass redundancy from high to low simulated mass accuracy. Product ions from a single peptide sequence were found to differ in their tendencies toward mass redundancy however no correlation between size and sequence specificity was apparent, demonstrating accuracy limitations in quantitative tandem mass spectrometry of peptides. This dissertation illustrates research into protein function conducted on three types of mass spectrometers and demonstrates some effects of liquid chromatographic and mass spectrometric performance on proteomic studies.

Acknowledgements

The pursuit of a doctorate entails challenges of many kinds, from financial strain to the pressures of balancing ones personal life with the demands of academic research to the psychological toll of the continuous experimental failure integral to scientific endeavour. The scientific process is one of trial and error. Though disappointment in experimental results is to be expected, it was at times the hardest aspect of a graduate degree in the sciences to bear, and a hardship I was protected from by interactions with my companions. It was the company of my colleagues, friends and family which allowed me to carry on through all my professional trials.

I wish to thank the following people for sharing their expertise in subjects technical, philosophical and professional: Dr. K.W.M. Siu, Dr. A. Hopkinson, Dr. S. Mädler, Dr. L. DeSouza, Dr. U. Verkerk, Dr. O. Krakovska Dr. S. Voisin, Dr. I. Saminathan, Dr. J. Zhao, N. Chevannes-McGregor, Dr. O. Masui, J. VanNostrand, Dr. V. Romanov, Dr. A. Matta, Dr. A. El-Faramawy, Dr. D. Golemi-Kotra, M. Fridman, Dr. M. Emes, Dr. I. Tetlow, Dr. A. Makhmoudova, Dr. J. McDermott, S. Ehyai, Dr. I. Coe, Dr. N. Grañé Boladeras, G. Koyanagi and Dr. V. Blagojevic. I also offer thanks to my friends outside of the workplace and my family for their tremendous support.

Table of contents

List of Figures

List of Tables

List of Abbreviations

ABA	abscisic acid
APC	adenomatous polyposis coli
ATF2	activating transcription factor 2
ATP	adenosine triphosphate
BRG1	brahma-related gene 1
C18	octadecane
CAD	collision-activated dissociation
CBP	cyclic AMP response element binding protein binding protein
CID	collision-induced dissociation
CK1	casein kinase 1
CPK	calcium-dependent protein kinase
CREB	cyclic AMP response element binding protein
DC	direct current
DEAE	diethylaminoethanol
DMSO	dimethyl sulfoxide
DTT	dithiothreitol
ECD	electron capture dissociation
ESI	electrospray ionization
ETD	Electron transfer dissociation
FAB	fast atom bombardment
FDR	false discovery rate
FOXO	forkhead box-O

FT-ICR	Fourier transform ion cyclotron resonance
FWHM	full peak width at half maximum peak height
GraR	glycopeptide resistance associated regulator
GSK	glycogen synthase kinase
HDAC1	histone deacetylase 1
HPLC	high performance liquid chromatography
Hse	homoserine
IgG	immunoglobulin gamma
kDa	kiloDalton
kPa	kiloPascal
LARP	la-related protein
LC	liquid chromatography
LRP	low-density lipoprotein-related receptor
MALDI	matrix assisted laser desorption-ionization
MKK	mitogen-activated protein kinase kinase
MRMhr	multiple reaction monitoring high resolution
MS	mass spectrometry
MS/MS	tandem mass spectrometry
m/z	mass-to-charge
p38	p38 mitogen-activated protein kinase
PBS	phosphate buffered saline
PD	plasma desorption
PDGF	platelet-derived growth factor
PSA	prostate-specific antigen
PTM	post-translational modification
RF	radio-frequency alternating current

RP	reversed phase
RR	response regulator
SBE	starch branching enzyme protein family
SbeIIb	Starch branching enzyme IIb
SCX	strong cation exchange
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SILAC	stable isotope labelling by amino acids in cell culture
SIM	selected ion monitoring
SP	starch phosphorylase
SRM	selected reaction monitoring
SS	starch synthase
Stk1	serine-threonine kinase 1
Stp1	serine-threonine phosphatase 1
SUMO	small ubiquitin-like modifier
s/n	signal-to-noise
TCA	trichloroacetic acid
TCS	Two-component systems
Th	Thompson
TLE	transducin-like enhancer of split
TOF	time-of-flight
VraR	vancomycin resistance associated regulator
VSMC	vascular smooth muscle cells
Wnt	wingless-type mouse mammary tumor virus integration site

Chapter 1

Theoretical and practical aspects of polypeptide analysis by high performance liquid chromatography-mass spectrometry

1.1 Introduction

Proteins are a diverse group of biopolymers whose primary structures are linear amino acid chains connected by amide bonds. Since polypeptides are integral to biology, myriad biological phenomena can be explored through their characterization. Proteins can be distinguished by their physiochemical parameters. Reactivity, isoelectric point and optical properties are generally imprecise protein identifiers due to the functional and structural redundancies of these diverse macromolecules. Each protein has a unique primary structure, or sequence, which reflects the genetic sequence encoding it as well as transcription and translation. Covalent post-translational modifications introduce additional levels of differentiation. Primary structures are definitive protein identifiers.

The first widely accepted technique for the determination of primary polypeptide structure, commonly known as Edman degradation, involved sequential derivatization of the amino-terminal amino acid followed by selective acid-catalyzed cleavage and identification of the free, modified residue (1). Though automation improved throughput, Edman degradation required the analyte in relatively large amounts and at high purity (2). The utility of the technique is also restricted by interference from amino acid modifications and inherent reaction efficiency limits. These shortcomings necessitated a more sensitive and less restrictive protein sequencing technique.

The sensitivity of automated Edman sequencers was matched by early mass spectrometers equipped with polypeptide-compatible ionization sources (3). Though polypeptides are not volatile, techniques by which they could be transitioned from the

solid or liquid phase into the gas phase permitted their analysis by mass spectrometry (MS). Fast atom bombardment (FAB) and plasma desorption (PD) were the first techniques used to ionize native polypeptides for MS but both had practical problems such as low ion yields (3). Electrospray ionization (ESI) and matrix assisted laser desorption-ionization (MALDI) simplified the ionization of labile, non-volatile molecules such as proteins, increasing the applicability of MS to polypeptide analysis (4-6). Tandem mass spectrometry (MS/MS) provided a rapid alternative to time-consuming sequential labelling and cleavage of single amino acids in Edman degradation (7). With MS/MS, sequence information could be derived from the mass spectra of both intact polypeptide ions and fragments produced by their gas-phase dissociation. Complementary MS and MS/MS data for a peptide could be acquired in seconds or less. Furthermore, unlike Edman sequencing, MS analysis was unhindered by modification of N-terminal amino groups. Thioredoxin from *Chromatium vinosum* was the first protein to be completely sequenced by a combination of MS and MS/MS (8). With continual improvement in mass resolution, sensitivity, versatility and throughput, MS ultimately eclipsed Edman degradation as the primary technology for protein sequencing.

Though it has demonstrable advantages over other polypeptide sequencing approaches, MS requires specific sample conditions. Every ionization source is limited by the maximum number of molecules that it can simultaneously ionize. In mixtures, different compounds compete for ionization, reducing the signal of any single species. The capacity of a mass spectrometer to sample ions is limited by physical constraints such as the orifice diameter in ESI instruments. Every mass analyzer operates over a mass-to-charge (m/z) range, beyond which it cannot effectively transmit ions. Ion detectors and mass analyzers also have limits of sensitivity and speed. All mass spectrometers therefore operate within specific ranges of analyte concentration, analyte size and sample complexity.

Top-down proteomics, the investigation and identification of intact proteins by MS, presents technical and bioinformatics challenges which have yet to be overcome (9). Gas phase-fragmentation and spectral interpretation are complicated by the structural

heterogeneity and size of protein analytes. Many mass spectrometers can resolve and detect ions up to a few thousand Thompsons (Th), however proteins range in mass from tens to hundreds of kiloDaltons (kDa). Consequently, it is impractical to sequence a protein from its molecular ion. Bottom-up proteomics, the assignment of primary protein structure based on sequencing of peptides, is more commonly employed than the top-down approach. Proteins of interest must be digested into peptides prior to MS in bottom-up proteomics workflows. Purified endoproteinases are typically used for protein digestion, yielding peptides whose sequences depend on the enzyme's specific cleavage sites (10). Digestion increases sample complexity, producing many small peptides from a single protein, but confers useful sequence characteristics to each peptide in the form of the protease-specific terminal amino acids. The exclusive specificity of trypsin for cleavage at the C-termini of arginine and lysine residues makes this enzyme particularly useful for bottom-up experiments as the basic side chains of these amino acids readily accept protons from acidic matrices to confer a positive charge to the peptide(11). The broad distribution of lysine and arginine residues throughout most proteomes and the high catalytic efficiency of trypsin also contribute to the popularity of this protease so that MS/MS of tryptic digests has become the standard for protein sequencing. It should be noted that a proline residue immediately C-terminal to Arg or Lys blocks the action of trypsin and that this is accounted for in automated spectral interpretation. Consecutive arginine or lysine residues diminish tryptic activity, creating peptides with so called ragged ends (10).

To be efficiently characterized by MS, peptides in a mixture such as a protease digest must be separated prior to reaching the detector. Separation prior to ionization reduces matrix effects on ionization and decreases the separatory burden on the mass analyzer. A perfect separation would rapidly and completely isolate each analyte while removing non-proteinacious components of the biological matrix, but no technique satisfies this ideal. Similarities in the physiochemical properties of polypeptides as well as other biomolecules impede their partitioning and this has remained a technical challenge in the field of proteomics from its inception. Many techniques are available for protein and

peptide separations. Sample complexity, the number of proteins to be identified, desired sequence coverage of each protein and MS performance characteristics are all considerations in the optimization of a sample preparation procedure.

High performance liquid chromatography (HPLC) provides efficient separation under conditions in which polypeptides are stable and consequently can be applied to proteins or peptides. Though gas chromatography delivers comparable resolution and speed, it is incompatible with polypeptides, which are non-volatile. Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) is convenient for the isolation of small numbers of polypeptides in the kDa mass range but cannot be directly coupled to MS (12, 13). In contrast, HPLC is capable of resolving hundreds or thousands of species, can be adapted to the separation of small or large molecules and is compatible with MS.

Provided the mobile phase promotes ionization, an HPLC pump can be directly coupled to a mass spectrometer via ESI so that compounds spatially focussed on the analytical column are measured as they elute. This instrument configuration, known as on-line LC-MS or LC-ESI-MS, has the capacity to capture extensive sequence information from a single analysis. With continuous separation and analysis, complex mixtures of biomolecules can be rapidly characterized in an automated fashion. The automated process, known as data-dependent acquisition or shotgun sequencing (used interchangeably with bottom-up proteomics), begins with the acquisition of a mass spectrum of the LC eluate to establish the m/z values of peptides eluting at any given time. The peptide ions are then isolated based on their m/z values, fragmented and mass spectra of the resulting product ions are acquired. The cycle is then repeated for the duration of the LC run. Ongoing technological advances in LC-ESI-MS have dramatically increased throughput and data quality so that tens of thousands of peptides can be routinely sequenced from a single sample in minutes or several hours.

Currently, ESI-equipped mass spectrometers surpass the throughput of those with MALDI interfaces for several reasons. MALDI-MS requires off-line sample fractionation and real-time user input to generate high sequence coverage from complex

peptide mixtures, whereas ESI allows continuous automated sampling of a chromatographic eluate. Since MALDI produces mostly singly charged molecular ions, compounds of high molecular weight ionized in this manner tend to exceed upper instrumental m/z limits (6). In contrast, ESI generates singly and multiply charged molecular ions with the degree of ionization depending in part on analyte structure (4). For instance, longer peptides have more ionisable moieties than shorter peptides and therefore tend to obtain higher charge states by ESI. Intramolecular electrostatic repulsion between charged sites such as protonated His, Lys or Arg side chains decreases with increasing peptide length so that larger peptides can accommodate more charges. The tendency of peptides to reach charge states proportional to their length and unique chemistries by ESI compresses their m/z range to match the characteristics of many conventional mass spectrometers.

The following sections of this chapter describe fundamental concepts in liquid chromatography, ESI and MS particularly with respect to their application in polypeptide analysis.

Fundamental principles of chromatography

All chromatographic systems consist of a stationary phase, a mobile phase and some support within which both are contained. As their names suggest, the mobile phase moves through the fixed stationary phase. In essence, chromatography involves a series of liquid-liquid or liquid-solid extractions in which one of the solvents is continuously moved along with any solutes it contains. Analytes are partitioned between the stationary and mobile phases depending on intermolecular interactions between all three, so that every compound has a velocity unique to each chromatographic system.

The column has convenient fluid dynamic properties for a chromatographic system and is used extensively for gas (GC) and liquid chromatography (LC), in which the mobile phases are gas and liquid respectively. The typical LC column is filled with densely packed stationary phase in order to maximize surface area and therefore solute

adsorption. A Gaussian spatial distribution of solute within the column produces symmetrical peaks when eluate concentration is plotted over separation time.

Given the mobile phase flow rate, the dimensions of the stationary phase and the dimensions and retention time (t_R) of a peak, the separatory efficiency of the chromatographic column can be determined. Considering chromatographic separation as a series of extractions, it follows that resolving power is proportional to the number of these events. Each extraction is termed a theoretical plate and the height (H) and number (N) of theoretical plates are common metrics of column performance. The value of N depends on the method by which peak width is measured. For Gaussian peaks the following are equivalent:

$$N = \left(\frac{t_R}{\sigma}\right)^2$$

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

$$= 8\ln 2\left(\frac{t_R}{W_h}\right)^2$$

Where t_R is the time at which the species elutes relative to the beginning of the separation, σ is the standard deviation of t_R, W is the baseline peak width and W_h is the peak width at half height (14). Asymmetry, fronting and tailing are common features of chromatographic peaks causing their shapes to deviate from Gaussian distributions (Figure 1.1). The terms tailing and fronting describe analyte elution beyond the slopes of the front and rear of the peak respectively. Where noise obscures analyte-dependent signal around the base of a peak, its limits cannot be unambiguously measured. In such cases, N cannot be accurately determined from the baseline peak width and W_h or σ are more appropriate. The height of each theoretical plate is described in relation to column length (L) and to standard deviation (σ) of the Gaussian analyte elution profile, which corresponds to peak width.

$$H = \frac{\sigma^2}{L}$$

$$= \frac{L}{N}$$

Since *H* is expressed in units of length, σ must be expressed in relation to mobile phase velocity and the column dimensions with the use of a conversion factor. Resolution increases with column length as longer columns accommodate more stationary phase but decreases with column diameter due to increased lateral diffusion. Long, narrow columns produce high backpressure and therefore require high performance mobile phase pumps.

For evaluating the properties of a stationary phase-mobile phase pair, it is convenient to consider the chromatographic separation a single extraction, as terms used to describe extraction efficiency apply to chromatography. The mass distribution ratio D_m, describes the relative amount of a solute in each phase according to the relationship:

$$D_m = \frac{n_{stationary\ phase}}{n_{mobile\ phase}}$$

The relationship is valid in terms of the absolute amount or concentration of a solute and reflects the equilibrium established between interactions of the solute with the stationary and mobile phases. The corresponding equilibrium constant is termed the partition constant K, where:

$$K = \frac{C_{stationary\ phase}}{C_{mobile\ phase}}$$

and $C_{stationary\ phase}$ and $C_{mobile\ phase}$ are the molarities in each respective phase (14). For any given analyte, K increases with increasing retention time. In practice, the value of K is manipulated to minimize separation time and maximize chromatographic performance by adjusting column temperature and phase composition.

Chromatographic resolution

The degree to which any two analytes are separated by a chromatographic system is expressed as the system's chromatographic resolution, R_s . Assuming Gaussian peak shapes:

$$R_s = \frac{2(t_{R2} - t_{R1})}{w_{b1} + w_{b2}}$$

where t_{R1} and t_{R2} are the retention times of the two analytes and w_{b1} and w_{b2} are the corresponding baseline peak widths (15). Resolution is proportional to peak separation and inversely proportional to peak width. The minimum resolution at which two peaks are resolved, or do not overlap at their baselines, is 1. It is convenient to estimate the area of a Gaussian peak by the extrapolation of straight lines from t_R along the boundaries of either side (Figure 1.1). This simple and commonly employed integration technique most accurately describes the areas of triangular peaks but omits some area, particularly around the base, when applied to other peak shapes. The minimum resolution at which two Gaussian peaks are resolved using such integration must be extended from 1 to at least 1.5. Deviation of GC and LC peak shapes from the ideal Gaussian distribution increases the minimum resolution required for complete separation and necessitates more complex peak limit modeling. For these reasons, the peak integration technique employed must be considered along with the reported chromatographic resolution of two analytes.

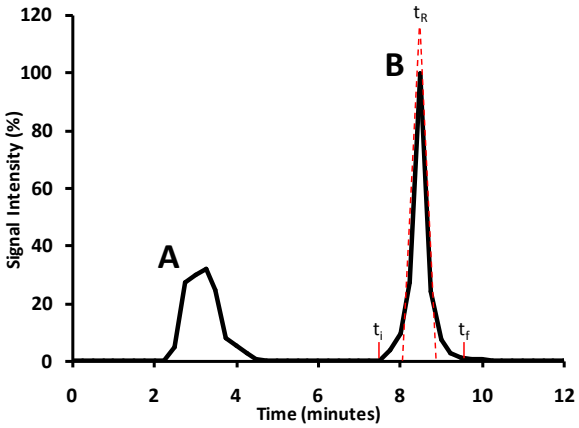


Figure 1.1: Two fully resolved chromatographic peaks with infinite signal-to-noise (s/n). (A) A non-ideal, broad peak with tailing, the area of which is inaccurately described by both normal distribution and triangular distribution. (B) An ideal chromatographic peak having normal distribution. t_i ; the time at which the peak begins, t_f ; the time at which the peak ends, t_R ; the retention time of the peak. Dashed lines indicate the limits of a triangle by which the peak area is approximated. The actual peak area includes all signal underneath the peak between t_i and t_f.

Peak broadening is caused by analyte diffusion in both stationary and mobile phases. The surfaces of the packing material and the inner column wall create turbulent flow, further broadening analyte distribution. Both diffusion and turbulence are affected by the mobile phase flow rate so that peak widths are highest at extremes of high and low flow. Broad peaks of low maximum intensity also result when localized saturation of the stationary phase increases analyte diffusion as a result of column overloading. These factors restrict chromatographic separations to ranges of analyte concentrations, resolution and speed (16).

Because analytes are concentrated on the column, sensitivity in GC-MS and LC-MS experiments depends on the quality of chromatographic separation. Though the two peaks shown in Figure 1.1 have comparable areas, peak A would not be detected if the sensitivity of the detection system decreased by 40%. Quantitative MS based on measurement of chromatographic peak area is most accurate where peak areas are uniform with high s/n. Optimal separation upstream of the ion source therefore provides optimal MS performance. Current state-of-the art HPLC systems produce optimal baseline peak widths on the order of several seconds over separations of several minutes to hours.

Polypeptide separation can be achieved using a number of stationary phases, including those with ion-exchange, size exclusion, ligand affinity, polar and non-polar properties. Packing materials commonly used for polypeptide HPLC consist of porous spheres several micrometers in diameter which have high surface area-to-volume ratios, are stable

under pressures of tens of KiloPascals (KPa), and can be functionalized with a variety of compounds.

Reversed phase chromatography

Reversed phase (RP) liquid chromatography, in which the stationary phase is hydrophobic, affects efficient peptide separation under ESI-compatible conditions and has been widely adopted for LC-MS sequencing. Aromatic and aliphatic hydrocarbons both bind peptides and octadecane (C18) is the most widely used in proteomics. The performance of RP packing materials depends on analyte size, for instance long aliphatic hydrocarbons tend to retain proteins so that C18 resin provides good peptide separation but poor protein elution.

Polypeptides adsorbed on RP can be separated by binary gradients over which the organic solvent content of the mobile phase is varied. Acetonitrile and methanol are volatile, water-miscible solvents capable of solvating hydrophobic compounds and are therefore common binary mobile phase modifiers. In general, gradients of increasing organic content elute peptides in order of increasing hydrophobicity and size, with the majority of peptides eluting in less than 40% acetonitrile. Since their molecular masses are lower than those of peptides, water and small organic molecules do not overlap peptide mass spectra. This simplifies data interpretation and the automated selection of precursor ions for MS/MS.

Ion-pairing agents boost LC separation as well as ionization. Formic acid, acetic acid and trifluoroacetic acid (TFA) are used for the generation of protonated peptides while ammonium formate, ammonium acetate and ammonium bicarbonate are used for peptide anion production (17). Each has advantages and disadvantages, for instance TFA has been shown to improve LC separation while diminishing ESI signal intensity and stability relative to formic acid and acetic acid (18, 19). Formic acid was the modifier used in the studies described here due to its overall LC-MS performance and low toxicity relative to TFA.

Ion exchange chromatography

Ion exchange chromatography exploits the association between solvated analyte ions and ions fixed in the stationary phase. Peptides are retained on ion exchange resins because their amines are readily protonated and their carboxyl groups are readily deprotonated. Many types of ion exchange chromatography are used for polypeptide separation. Weak anion exchange (WAX) stationary phases such as those based on diethylaminoethanol (DEAE) are commonly used in intact protein purification. Strong cation exchange (SCX) is well suited to peptides since at low pH, peptides are stable, cationic and interact with deprotonated strong acids. Sulfonic acids make versatile SCX stationary phases due to their low pKa values. With gradients of increasing ionic strength, peptides are eluted from SCX resins according to their charge states as they are displaced from the anionic sites by other cations present in the mobile phase. The technique suffers from low resolution because the distribution of peptide charge states tends to be narrow.

Though RP provides superior peptide separation than ion exchange, the two are complementary, providing higher sequence coverage when used in series than separately (20). The high salt content of ion exchange mobile phase impedes gas phase analyte ionization so desalting is required prior to MS. This can be accomplished by fractionating the SCX eluate and desalting the fractions separately or by using on-line RP as in the multidimensional protein identification technology (MudPIT) workflow (21). Biological specimens such as blood serum contain more polypeptides than can be separated by reversed phase chromatography alone so two-dimensional LC is often used to expand sequence coverage from samples of such complexity.

Metal ion and metal oxide affinity chromatography

Protein phosphorylation is a prominent post-translational modification but the physical properties and low abundance of phosphoaminoacids often necessitate specialized sample preparations for their observation. Phosphate bonds with serine, threonine and tyrosine side chains are relatively stable, tolerating drying, SDS-PAGE and a wide pH range without dissociation and can therefore be purified (22). Interactions between phosphates

and metal cations are exploited in immobilized metal affinity chromatography (IMAC), in which the stationary phase contains chelated metal ions such as Fe(III) (23). While histidine-containing sequences can be purified using nickel and copper columns, the affinity of phosphopeptides for metal ions differs, with trivalent gallium reported to provide the lowest binding of non-phosphorylated peptides (24). Metal oxide affinity chromatography (MOAC), in which the stationary phase contains stable metal oxides such as ZiO_2 and TiO_2, is a more selective variant of IMAC though both techniques tend to retain peptides containing aspartate and glutamate (25, 26). One approach to improving phosphopeptide specificity is the derivatization of carboxyl groups (27). Inclusion of aliphatic hydroxy acids in the MOAC mobile phase was also reported to enhance selectivity for phosphopeptides by decreasing interactions between acidic peptides and the stationary phase without interfering with LC-MS analysis (28).

Nanoflow HPLC

Sensitivity, which can be measured as the lower limit of detection of a mass spectrometer for a given analyte, is critical to the characterization of compounds present in low abundance or whose concentration is low relative to matrix components. Nanoflow LC is defined by mobile phase flow rates in the hundreds of nanoliters per minute range and by column bores in the range of tens to hundreds of micrometers, which impart higher sensitivity compared to conventional microliter per minute HPLC. The sensitivity improvement is due to confinement of analytes in smaller mobile and stationary phase volumes, a concentrating effect. To restrict post-column analyte diffusion and provide stable electrospray, nanoflow LC systems are coupled to ESI emitters several micrometers in internal diameter. Nano-electrospray ionization enhances sensitivity further because droplet desolvation and ionization increase inversely with emitter bore. The passage of ions from source to mass spectrometer is several orders of magnitude more efficient in nano-ESI compared to ESI (29, 30). Limited sample loading capacity is an inherent disadvantage of nanoflow HPLC so the technique is best for analytes present in low concentration ranges or where high sample volumes are unavailable.

The manipulation of sub-microliter fluid volumes requires automation and in nanoflow HPLC, the trap column serves as an interface between microliter and sub-microliter scale liquid handling systems. Trap columns are many times shorter and wider than the analytical columns with which they are paired, a geometry which accommodates both microliter per minute flow from a sample loading pump and low flow from the pump supplying the separatory gradient. The trap and analytical column packing material are matched so they respond in the same way to the mobile phase. A switching valve separates the two elements. In sample loading mode, the valve brings a sample syringe or a fixed volume loop connected to a loading pump in line with the trap. Polypeptides are passed over the trap in solution and bind to it, while other soluble material is carried away with an excess of solvent. The valve position is then changed, bringing the trap in line with the downstream column and the polypeptides are eluted from both with mobile phase. Trap unloading can be accomplished either by reversing the flow or by keeping its direction constant, the latter allowing the trap to act as a particulate filter. The geometry of the trap and its size relative to the analytical column and the analytical gradient are important to system performance. Excessive trap volume can lead to delayed retention times, peak broadening, low sensitivity and column overloading. Considering that a typical nanoflow column with an internal diameter of 75 µm and a length of 150 mm has a volume of 0.66 µl and is operated at flow rates from 200 to 300 nl per minute, a trap 1 mm long 1 mm in diameter would more than double the volume of the chromatographic bed and would take from 2.6 to 3.9 minutes to flush with mobile phase. An RP trap can be used for sample desalting during loading; however excessive mobile phase volumes diffuse or wash out the analyte, decreasing sensitivity. Desalting the sample in advance of trapping is advisable in light of anecdotal evidence of excessive salt formation around the ESI orifice during analysis of SCX fractions.

Operating principles of mass spectrometers

All mass spectrometers contain one or more ion source, mass analyzer and ion detector. Ion sources transition neutral analytes to the gas phase while imparting them with charge. Mass analyzers separate ions on the basis of their m/z. Detectors convert ion related

events, such as the discharge of an ion onto a metalized surface, into electronic signals for further processing into interpretable forms. Mass spectrometry technologies of relevance to modern protein analysis are discussed in the following sections.

Electrospray ionization

During ESI, charged droplets are continuously generated from the end of a liquid filled capillary and accelerated toward the orifice of a mass spectrometer by an electrostatic potential. As they travel, the unstable droplets collapse into successively smaller particles, ultimately releasing ions. The process transpires in milliseconds to microseconds depending on initial droplet size (31). In contrast to the similar atmospheric pressure chemical ionization process, electrospray is known as a soft ionization technique because it preserves the integrity of labile species such as peptides. A few key aspects of the ESI mechanism are described here.

An electric field between the electrospray needle and the orifice draws the eluting liquid into a point or several points from which droplets are ejected. The liquid surface from which droplets are ejected has a conical shape known as a Taylor cone. It follows that the charge of the orifice must be opposite that of the droplets and that the emitter must have no net ionic charge. Accumulation of counterions within the emitter would counteract the ESI voltage, reducing ion transmission. A steady state is maintained at the emitter because counterions of those ejected confined to the tip by the electric field undergo reduction or oxidation. The Taylor cone can be sustained indefinitely if the flow rate of liquid from the capillary is matched to the ESI voltage and if the redox chemistry of the solution allows the balancing of charge.

In order to be effectively transferred to the mass analyzer, ions exiting the Taylor cone must be desolvated. Once droplets leave the emitter, they are concentrated by solvent evaporation. Ions within the shrinking volume arrange themselves at its surface by mutual repulsion. Where charge repulsion overcomes the cohesive forces holding a droplet together, it ruptures and ejects smaller charged droplets in what is known as

Coulombic explosion. A droplet collapses due to this effect at its Rayleigh limit. The relationship between charge and stability is approximated by the equation:

$$Q_{droplet} = 8\pi(\varepsilon_0 \gamma R^3)^{1/2}$$

Where $Q_{droplet}$ is the droplet charge, ε_0 is electrical permittivity, γ is the surface tension and R is the droplet radius (32). Since the radius term is cubed, desolvation contributes disproportionately to this relationship. Volatile solvents and conditions promoting evaporation provide strong ion current, for instance ion yields are significantly improved by heating (33) and passing dry gas over the source. Though the Rayleigh model is relatively simple, there is much empirical evidence of Coulombic explosion occurring near or at the Raleigh limit (34). The electric field of an ESI source may destabilize the droplet below the Rayleigh limit by drawing charges to the side closest to the orifice, so electrospray voltage affects ionization (35).

Though Coulombic explosion is known to be repeated by droplets of decreasing size during ESI, the progression from solvated ions to molecular ions is incompletely described. Two theories regarding the final stage of the process have been proposed. These are the ion evaporation model and the charged residue model. In the ion evaporation model, multiple ions contained in a single droplet separate from its surface as desolvated species (36). In the charged residue model, Coulombic explosion leads to the formation of droplets containing single analyte molecules, from which all solvent evaporates (37). There is evidence that both occur and that the route by which an analyte ion enters the gas phase depends on its size.

Electrospray ionization produces a dense ion beam, however transit from atmospheric pressure to the vacuum of the mass spectrometer causes particles passing through the orifice to separate by adiabatic expansion. If the pressure difference between the faces of the orifice is 2.05 or higher, a condition met by virtually all atmospheric pressure ionization sources, the gas forms a supersonic free jet (38). Expansion continues until the gas reaches ambient pressure by which point its temperature has decreased to a fraction of

its initial value, so the separation and stability of the ions increase with lower front end pressure. Within the free jet the gas achieves supersonic speeds, while collisions along its boundary restrict the gas speed to subsonic range. Here the fluid on the high pressure face of the orifice defines the speed of sound. Jet dimensions but not jet shape change in relation to orifice diameter. Three problems in mass spectrometry result from free jet expansion; radial ion movement results in ion loss, cold species form gas-phase clusters and cold ions do not readily fragment in MS/MS. All of these issues can be mitigated in the front end.

Energy to break apart adducts and promote intramolecular dissociation can be supplied by heating the source or inducing gas-phase collisions. Various front end design features are used to focus the diffuse ion beam in order to decrease ion loss. Skimmers are conical collimators placed between the orifice and the high vacuum region which deflect a large proportion of neutral species while accommodating the ion beam and controlling the shape of the free jet. A skimmer positioned close to the orifice restricts expansion and keeps those neutral species at the axis of the jet moving at supersonic speeds, where they concentrate ions by collision (39). A non-linear electric field between the orifice and the skimmer also directs ion trajectories. The field strength is highest at the point of the skimmer and decreases with distance from the orifice. Discharge of ions near the skimmer tip is counteracted by the dense stream of supersonic neutrals formed here, which sweep them past the high field region. Farther away from the orifice, the neutral density decreases and ions are steered mainly by the electric field. The influence of the applied potential on an ion's trajectory depends on its m/z. In one model, transmission of singly charged ions with masses between 105 and 609 Da beyond the entrance of the skimmer was twice that of myoglobin with a charge of 9 (m/z 1883), but total ion loss along the entire skimmer wall was comparable for all ions(40). The remaining gas-phase ions are further drawn toward the mass analyzer by an electrostatic potential between the skimmer and an element immediately downstream of it, such as a mass analyzer.

Aside from decreasing gas load on front end vacuum pumps and collimating the ion beam, a skimmer can be used to control ion temperature and therefore ion chemistry.

High orifice-skimmer field regions accelerate the ions through the slower neutrals, causing heating sufficient for declustering or even in-source fragmentation (41). Orifice-skimmer potential can be adjusted to suit an application, for example high voltage is used to dissociate clusters where strong intermolecular attraction diminishes the strength of the molecular ion signal while low voltage is used to stabilize labile species (42, 43). Front end ion optic voltages are often referred to as declustering potentials for these reasons.

One inherent limitation of a skimmer is that it omits a portion of the ion beam from the mass analyzer. Sensitivity gains resulted from replacement of a skimmer with other elements such as a quadrupole in the AB Sciex QJet™ and a stacked ring ion guide in the Thermo S-lens™(44). These systems allow the free jet to expand completely and supply electric fields which confine the ions at sufficient distance from the elements to minimize discharge. It is important to note that the source is not the only area limiting instrument sensitivity. Every mass analyzer has a fixed volume which restricts the number of ions it can simultaneously contain before Coulombic repulsion drives them out of containment and detectors have lower limits of response dependent on electronic noise.

Mass analyzers

Once they reach the low pressure region of the mass spectrometer, ions can be guided by mass analyzers and ultimately detected. The structure of the instrument beyond the source depends on many factors. Scan speed, sensitivity, resolution, mass accuracy, dynamic range, stability and cost all vary with mass analyzer design. Mass spectrometers used extensively for polypeptide analysis incorporate mass analyzers in many different configurations, each with advantages and drawbacks. The forms employed in the following chapters are discussed below.

Quadrupoles

The versatility and simplicity of quadrupoles account for their widespread use. A quadrupole consists of four parallel rods with identical dimensions forming the four corners of a theoretical square and separated by a vacuum. Given round rods, the ideal

distance between opposing corners of the square is 1.148 of the rod diameter so that the electrostatic field they generate is hyperbolic along the axis of the rod set (45). Opposite rods in the quadrupole are electrically connected. Though the field created by elliptical rods is better for ion transmission, cylindrical rods are also suitable. Other multipoles, such as hexapoles and octapoles are also used in mass spectrometers. These differ from quadrupoles in the field shapes they generate and only permit transmission of broad m/z ranges.

The Mathieu equation, shown below, describes ion confinement in a quadrupole as a function of radio-frequency alternating current (RF) and direct current (DC) voltages applied to the rod set (45).

$$\frac{d^2u}{d\xi^2} + (a_u - 2\,q_u \cos 2\,\xi)u = 0$$

Where:

$$a_u = \frac{8eU}{mr^2\Omega} \quad \text{and} \quad q_u = \frac{4eV}{mr^2\Omega}$$

Here u is the position along the cross-sectional quadrupole plane, U is DC voltage, V is RF voltage from zero to peak, r is the distance between electrode surfaces, Ω is the RF frequency in radians per second, ξ is equal to one half the product of time and RF frequency and m is ion mass. The a_u term describes the DC voltage space and the q_u term describes the RF voltage space so that finite combinations of values for these terms give solutions corresponding to stable ion trajectories. Conversely, certain RF voltage and DC voltage combinations will cause an ion to discharge on a rod or pass between rods perpendicular to the ion path of the mass analyzer.

Combined DC voltage and RF voltage are applied to the four rods with a 180 degree phase shift between adjacent rods. The resulting field creates mass discrimination by confining ions with specific m/z values to the space between the rods along the geometric center of the quadrupole rod set. At a given DC voltage (amplitude) and RF voltage

(amplitude), only a narrow m/z range is transmitted. In other words, the ions are pulled and repelled by alternating rod pairs to an extent that they are either held between them or are lost. To scan a mass range, DC voltage and RF voltage are simultaneously ramped. Quadrupoles are driven by a resonant inductor/capacitor circuit at constant RF frequency. Increasing RF frequency tends to decrease mass range with concomitant improvements in resolution and sensitivity, while requiring higher RF voltage (46). Ions transit the axis of the rod set through a combination of their initial kinetic energy and acceleration due to any potential gradient applied between elements located at opposing ends of the quadrupole.

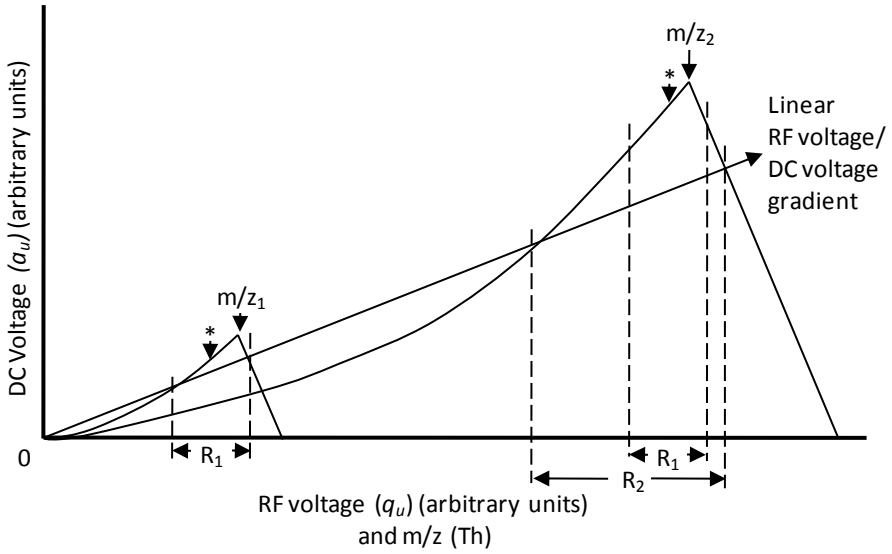


Figure 1.2: Mathieu stability diagrams for two theoretical ions in a quadrupole. Actual m/z values of each ion, the maxima of each stability region, are denoted as m/z_1 and m/z_2 respectively. A linear RF voltage/DC voltage gradient (m/z scan line) starting from the origin is shown as a diagonal arrow. Limits of observed m/z ranges for each ion are defined by the points at which the RF voltage/DC voltage gradient line transects each stability region. R_1: Theoretical resolution 1, R_2: theoretical resolution 2. Average m/z values for each of the ions, assuming both are observed at R_1, are indicated by asterisks. Peak geometries shown are approximations.

Linear RF voltage/DC voltage scans produce spectra in which resolution decreases with increasing m/z (46). Figure 1.2 shows the stability diagram for two theoretical ions in relation to RF and DC voltages at constant RF frequency. Ramping RF voltage/DC voltage linearly from the origin produces high resolution (R_1) in the low m/z region and low resolution (R_2) at higher m/z. The two peaks shown in Figure 1.2 will only have equivalent resolution (R_1) if the DC voltage is increased proportionally to RF voltage or if the intercept of the straight scan line is decreased and the slope is increased, however a linear scan calibrated for constant resolution at two m/z values may have discontinuous resolution for others (47).

Due to their asymmetrical stabilizing behaviour, quadrupoles are inherently biased toward lower apparent m/z values. The linear voltage gradient allows disproportionate transmission of the lower m/z portion of each stability region, resulting in shifts of the peak maxima within the resulting mass spectrum toward the average transmitted m/z value (47). Signal processing after ion detection is required to correct for effects of this fundamental limitation of quadrupoles on m/z accuracy. Furthermore, the shapes of the stability regions imply that for a quadrupole, sensitivity can only be improved at the cost of resolution, by effectively lowering the slope of the RF voltage/DC voltage gradient and broadening the transmitted m/z range.

Time-of-flight analyzers

High mass resolution is necessary for studies of isobaric mixtures not separated before the ion source or studies requiring the determination of the isotopic distributions of large (tens of kDa) highly charged ions for which mass separation is compressed by the charge. The distinction of ions of the same nominal m/z value exceeding 300 Thomson (Th) and differing in mass by 0.01 Th requires resolution greater than 30,000 full peak width at half maximum peak height (FWHM), a condition met by three types of mass analyzers. To date, the Fourier transform ion cyclotron resonance (FT-ICR) mass analyzer is unsurpassed in resolution with scan times comparable to other instruments. Second in

mass resolution is the orbital trap (orbitrap) followed by the time-of-flight (TOF) analyzer (48). TOFs provide high resolution based on the familiar Newtonian equation:

$$e_k = ½\ mv^2$$

which relates the kinetic energy (e_k) of a particle to its mass (m) and velocity (v). The time-of-flight system consists of a field-free region under vacuum with analyte-accelerating electronics at its entrance and a detector at the terminus of the ion path. Analytes are initially accelerated at constant DC voltage and travel in the field-free region (time-of-flight tube) at velocities inversely proportional to the square roots of their m/z values. The TOF scan cycle is unique in that it is discontinuous, with each discrete ion packet being accelerated, separated and detected in succession (49). The time required for a scan depends on the drift time corresponding to the upper mass limit. For this reason, TOFs require a pulsing supply of ions such as a MALDI source or a linear ion trap, which accumulates and periodically releases ions produced by a continuous source.

In TOF-MS, mass resolution ($m/\Delta m$) and time resolution (Δt) are related according to the following equation:

$$m/\Delta m = t/2\Delta t$$

Time resolution can be considered time domain FWHM. It follows that the mass resolution of such systems depends on the length of the ion path, with separation of different m/z values increasing over distance. Resolution is also affected by the three-dimensional distribution and velocity distribution of the analyte ion population so many instruments incorporate ion optics for shaping the ion beam. Ions enter the TOF with a range of initial velocities. High acceleration voltage compensates for the velocity distribution at this stage, decreasing the range of flight times of a population of ions having the same m/z. Analyte spatial distribution is particularly important along the axis in which the ions are accelerated. For hybrid Q-TOF instruments in which the TOF entrance is coupled to a quadrupole, the two are typically oriented orthogonally so that ions spread along the length of the quadrupole are accelerated from the same plane in the

TOF drift axis. Some TOF analyzers include one or more ion mirrors which generate electrostatic fields to reverse ion trajectories. Placing an ion mirror, known as a reflectron, opposite the accelerator in the drift tube increases flight time and therefore TOF resolution without increasing instrument size. Such a design produces parabolic ion trajectories so the detector is mounted at the same end of the flight tube as the accelerator. The distance ions penetrate the ion mirror increases with their kinetic energy and this will decrease resolution if not compensated by an electrostatic field gradient (49). Ion mirrors therefore increase resolution both by extending drift time and beam focusing. The static electric field gradient of the reflectron can be given linear, quadratic or curved slopes to optimize focusing (50).

Orbitraps

Many detectors employed in mass spectrometry are electron multipliers similar to those used in spectroscopy, for instance quadrupole instruments use channel electron multipliers and Qq-TOF instruments use multichannel plates, which are arrays of miniature electron multipliers. Upon contact with ions, these detectors release many more electrons and the resulting current is measured as signal. The orbitrap is both a mass analyzer and a detector which uses induced image current rather than ion discharge to generate a signal. The orbitrap consists of two electrodes separated by a vacuum. The outer electrode is a tube segmented at its center into two halves, which surrounds the inner rod shaped electrode. Both have parabolic cross sections. An electrostatic field is applied to the trap to maintain analytes in stable trajectories with rotational and axial components. Axial rotation frequency, that is the rate of oscillation along the length of the trap, is independent of ion energy and position but proportional to $(m/z)^{-1/2}$. Ionic charge induces current on the outer electrode which is converted to m/z by means of Fourier transform (48). Since their recent commercialization, Orbitraps capable of high resolution, high sensitivity and wide dynamic range have been widely adopted in proteomics (51). Orbitraps commonly provide resolution in excess of 100,000 FWHM, significantly higher than the 30,000 FWHM range capability of Qq-TOFs; however their resolution is proportional to scan time so that a scan at maximum resolution can require

several seconds to produce. This has the potential to impact sequence coverage or quantification so resolution and scan speed should be balanced in consideration of the specific application.

1.2 Tandem mass spectrometry of peptides

Sequencing from MS/MS data takes advantage of the patterns in which polypeptide ions fragment. With respect to dissociation, the amides of the peptide backbone are for the most part comparable in their reactivity and are more reactive than amino acid side chains. For these reasons, backbone fragmentation dominates, forming various products containing parts of the original sequence from which the order of amino acids can be ascertained. Peptide ion fragmentation has been extensively studied from intramolecular to proteomic scales, in part to establish rules for spectral interpretation. While the amino acid sequence of a peptide impacts how it fragments, relative abundances of product ions in MS/MS spectra cannot be entirely predicted. Established gas-phase peptide bond dissociation mechanisms define favoured and disfavoured products. The chemistry and universal trends of peptide cation fragmentation facilitating spectral decryption are discussed below.

Gas phase protonated ions can be dissociated via reduction (gain of electrons), collision with neutral species or infrared irradiation. Collision-activated dissociation (CAD), also known as collision-induced dissociation (CID), is the most commonly used technique for peptide MS/MS and involves passing the analyte through inert gas. Collisions between analyte ions and gas molecules supply the energy required for bond disruption. A collision cell includes a mass analyzer, a collision gas supply and a barrier to isolate the gas. Amide carbon-nitrogen bonds of the peptide backbone are the most susceptible to disruption by CAD. The resulting amino terminus-containing fragment ions and carboxyl terminus-containing fragment ions are termed *b*-ions and *y*-ions respectively according to the nomenclature developed by Roepstorff, Fohlmann and Biemann (Figure 1.3) (52).

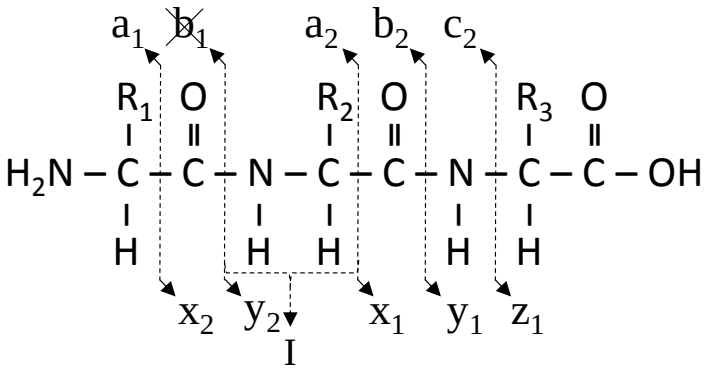


Figure 1.3: Possible gas phase fragments of a theoretical tripeptide labelled according to the conventional Roepstorff-Fohlmann-Biemann nomenclature. The bonds broken in the formation of each product are indicated by dashed lines. Arrows indicate the name of each product ion and its elemental composition, not including the charge-imparting species. The structures of the product ions are not reflected. I: an iminium ion consisting of the overlapping structures of the y_2 and a_2 ions. The b_1 ion is crossed out to indicate that it does not occur in tandem mass spectra of peptides.

The CAD of protonated peptides is charge directed, requiring protonation of the amide undergoing dissociation. In low-energy peptide ions, protons mainly reside on the side chains of the most basic amino acid residues and to some extent on the amino-terminal amine. Collision supplies energy to destabilize the ion, mobilizing charges along the peptide backbone. When the proton localizes at an amide, it can direct carbon-nitrogen bond dissociation through the formation of an intermediate with a protonated five-membered ring. Either the amide nitrogen or oxygen atom can accept protons and cleavage mechanisms requiring protonation at either atom have been proposed. The carbonyl oxygen is more basic than the nitrogen and the carbon atom of a protonated carbonyl is susceptible to nucleophilic attack by the carbonyl carbon N-terminal to it (53). The resulting ring may undergo 1,1-elimination in which one product retains the proton from the ring (54). Though the amide nitrogen has lower proton affinity, the proposed intermediate following nitrogen protonation has a significantly lower energy barrier to dissociation. Here the intermediate is an oxazolone formed through nucleophilic attack

on the carbonyl carbon adjacent to the protonated nitrogen by the nearest N-terminal oxygen. In the theoretical transition state between the oxazolone and its fragments the proton is shared between nitrogen atoms of the residues where cleavage takes place, whereas in the transition state of the protonated-oxygen directed fragmentation the proton is confined to the ring (55). The tendency of the charge to reside on the product having higher proton affinity has been presented as evidence for the nitrogen protonation-directed fragmentation model, in which the proton is more equally shared by the transition state fragments. Conversely the rarity of dissociation at C-terminal amides of proline residues is given as support for the oxygen protonation-directed model (53). Proton transfer from amide oxygen to the amide nitrogen has a low energy barrier in the context of CAD experiments so the movement of a proton from amino acid side chain to amide oxygen to nitrogen is feasible under typical CAD conditions.

Single collision conditions are invaluable for CAD-based thermodynamic studies but result in limited fragmentation. Numerous collisions are generally required to maximize product ion signal and diversity. Multiple collision conditions are invariably used in proteomics experiments. Comparable spectra can be obtained over a range of collision cell pressure and kinetic energy. This is fortunate considering that collision conditions vary with instrument design and that acquisition settings are not standardized. Where optimal collision conditions are unknown, precursors are often accelerated through the collision cell using a voltage ramp to vary their collision energy, which tends to produce more product ions.

The more basic an amino acid, the more collision energy is required to mobilize a proton associated with it. The side chain of arginine has the highest gas phase basicity of any amino acid, followed by lysine and histidine (56). Consequently, singly charged tryptic peptides with C-terminal arginine residues require more energy to fragment than those with C-terminal lysine residues while multiply protonated species require less collision energy to induce dissociation than their singly charged counterparts.

A *y*-ion has the equivalent structure of an intact peptide ion and fragments primarily at the amide C-N bond during CAD (55). The lowest energy *b*-ion structure includes a protonated oxazolone ring at the cleavage site, formed from the C-terminal amino acid and the amino acid N-terminal to it (57). The absence of b_1 ions from the tandem mass spectra of peptides is evidence that interaction of adjacent carbonyl groups is required for the unimolecular reaction to occur. Subjecting *b*-ions and *y*-ions to CAD causes them to fragment similarly to peptide ions. A *b*-ion can undergo further backbone cleavage to form a smaller *b*-ion or the oxazolone can lose carbon monoxide to form an *a*-ion.

In addition to backbone cleavage, internal fragmentation occurs, particularly at aromatic amino acids, proline, histidine, leucine and isoleucine. The resulting iminium ions are single amino acid hybrids of *y*-ions and *a*-ions (Figure 1.3) (58). Iminium ions are conventionally denoted by the single letter abbreviation of the amino acid from which they are derived. Iminium ion information is often omitted by MS/MS scan windows biased toward the collection of ions with higher masses. The only iminium ion with a mass greater than the lysine y_1 ion at 147 Th is that of tryptophan at 159 Th, so that the W ion is often present in proteomic data.

Fragment ionization depends on the retention of charge from the precursor on a product, thus fragments with high proton affinity such as *y*-ions from tryptic peptides are the most abundant. The competition of fragmentation products for protonation is illustrated by the increasing intensity of *b*-ions relative to the corresponding *y*-ions in a series of peptides of increasing N-terminal proton affinity (59). Multiply protonated peptides yield singly and multiply charged products upon dissociation. Precursor charge state can influence sequence coverage. Increasing the number of charges on a peptide increases proton mobility and therefore the accessibility of a proton to amino acids of decreasing basicity.

Of the biologically occurring amino acids, proline alone has a secondary amine at its amino terminus which forms a five-membered ring with the side chain. The high basicity of the secondary amine and to some extent ring strain destabilize the proline amide.

Consequently, C-N bond dissociation at proline amides is energetically favoured and the corresponding product ions tend to be highly abundant (60, 61, 62).

Cleavage of aspartic acid and glutamic acid C-terminal amide C-N bonds is favoured where proton mobility is restricted by the side chain of arginine or by chemical modification (63, 64). This is particularly evident in MALDI MS/MS data, in which tryptic peptides are singly charged. With the single proton associating with the C-terminal arginine, *y*-ions ending at residues C-terminal to acidic amino acids form prominent peaks. The proline effect and aspartic acid effect can be considered synergistic. Where proline is C-terminal to aspartate or glutamate, cleavage between the two is enhanced (65). The proline effect is also enhanced by N-terminal histidine in the presence of mobile protons (53).

Reduction activated dissociation

Electron transfer dissociation (ETD) and electron capture dissociation (ECD) are related techniques with two advantages over CAD. First, they tend to produce more evenly distributed product ion spectra because radical driven bond dissociation is less dependent on primary sequence. Second, they selectively cleave the peptide backbone, leaving labile modifications such as phosphates intact. Fragmentation between alpha carbons and amide carbons produces a-ions and x-ions analogous to the *b/y*-ion series while the c/z-ion series results from dissociation of alpha carbon and amide nitrogen atoms (Figure 1.2). These alternative fragmentation products are absent from CAD MS/MS spectra with the exception of occasional a-ions. In contrast, ETD and ECD mainly produce *c*-ions and *z*-ions from relatively unstable protonated peptide ion radicals (66, 67). Complementary *b/y*-ions and *a/x*-ions also result from electron transfer (68).

Both ECD and ETD cause bond dissociation by the addition of an unpaired electron to the analyte, the former employing an electron beam and the later an electron donating reagent ion. ECD is only efficient in FT-ICR instruments, the most expensive type of mass spectrometer. Conversely, more economical quadrupole ion traps can be equipped with ETD. CAD spectra are information rich so ETD is not necessary for sequencing

unmodified peptides; however the technique has demonstrable advantages for the characterization of peptide modification. For example, reporting of amino acid oxidation by ETD is charge state independent but is inaccurate at some charge states with CAD since peptides with the same sequence but oxidized at different sites produce ambiguous CAD MS/MS spectra (69).

Charge reduction can be a problematic consequence of electron transfer. Singly protonated hypervalent radicals have no net charge and therefore cannot be detected, while doubly protonated radicals exhibit low fragmentation efficiencies without supplemental collisional activation (70). These techniques are therefore best suited to analytes with three or more positive charges.

Polypeptide quantification

Compound identification and quantification are the parallel motives of analytical chemistry. Some biological inquiries can be satisfied using polypeptide identification alone, while others require knowledge of the influence of an independent variable on the concentrations of specific analytes. For instance, human immunodeficiency virus infection testing is based on the presence or absence of viral antigens while the levels of Prostate-specific Antigen and Cancer Antigen 125 in blood are diagnostic of prostate cancer and ovarian cancer respectively. Because LC-MS is versatile and sensitive, the platform is widely used in protein quantification. The amount of a compound in a sample can be inferred from LC-MS data by several methods but all operate on the assumption that an analyte-specific signal is proportional to its concentration. Every quantitative experiment follows one of two sample preparation paradigms, label-free quantification in which native species are measured or labelled quantification in which isotopic tagging is performed prior to analysis and the relative abundances of heavy and light species are measured. Heavy isotopes can be incorporated directly into a protein sample *in vivo* as in the case of stable isotope labelling by amino acids in cell culture (SILAC), or by derivatization of intact proteins or digests thereof (71-73). Alternately the sample can be spiked with synthetic heavy isotope labelled standards. Label-free quantification is

inexpensive but error prone. Assuming the species being quantified have consistent response factors, isotope labelling gives excellent precision and accuracy. Also, if the concentration of the standard is known, the absolute amount of the analyte can be ascertained.

Many LC-MS operating modes can supply quantitative information. The most general approach to quantification is spectral counting, a technique based on data-dependent acquisition which relates the number of spectra assigned to a protein to its concentration, (74). While broadly applicable, the accuracy of spectral counting is constrained by the variable ionization efficiency among peptides, ion suppression, trypsin cleavage site distribution, and precursor ion sampling bias. Protein homology also leads to error in spectral counting measurements, since peptides which could be derived from any of several different proteins must be excluded from the count or assigned to a specific protein somewhat arbitrarily. Like other techniques, the response factor as a function of spectral counts, varies from protein to protein (75).

Targeted analysis increases sensitivity by increasing the amount of acquisition time dedicated to detecting ions of interest. Most targeted strategies make use of ion chromatograms. In selected ion monitoring (SIM), the mass spectrometer is programmed to collect data for masses corresponding to known peptides and areas of chromatographic peaks associated with each precursor ion are subsequently integrated. Isobaric interference and retention time shifts are common in quantification regimes based on precursor ion m/z (76). Incorporating fragmentation into the assay increases its specificity where precursor mass and retention time are questionable identifiers. This is the basis of selected reaction monitoring (SRM), also called multiple reaction monitoring (MRM). In the SRM scan cycle, the mass spectrometer is programmed to select a precursor ion of specific m/z, fragment it, then select a limited number of its product ions, detecting them in series (77). A single precursor ion/product ion pair is known as a transition and transitions are defined in advance of the SRM experiment to target species of known identity. SRM can be multiplexed for the quantification of hundreds of analytes on sensitive and rapid scanning triple quadrupoles with large ranges of linear

dynamic response. These instrument features translate to high throughput and linearity over a wide analyte concentration range. The simultaneous appearance of signals for multiple transitions, which can be overlaid to form a single chromatographic peak, is the basis of specificity in SRM.

Until recently, acquisition schemes were divided into those providing reliable quantification and those sampling the highest number of peptides in a sample, attributes which have been merged with the advent of workflows such as SWATH MS (78). These operate on the concept that multiple precursor ions and their products can be simultaneously manipulated within the same mass analyzer and the resulting chimera spectra deconvoluted. SWATH MS divides the entire precursor ion mass range into overlapping isolation windows tens of Th units wide, subjects the analytes in each window to CID then collects the resulting product ion spectra. The merged data is searched for peptide sequences then mined for quantitative data from recognized analytes. By replacing mass selection with powerful post-acquisition data processing, the process greatly reduces scan time limitations. High resolution is obligatory for spectral deconvolution, so SWATH MS was developed for Qq-TOF instruments. SWATH MS is better suited to the exploration of unknowns than the quantification of specific molecules since it is less sensitive than SRM (79).

1.3 Interpretation of MS/MS data

Machine learning based product ion prediction

Large scale MS/MS datasets have been mined for fragmentation trends since they became available, however features of product ion spectra remain unpredictable (80). In one example, a training set of MS/MS ion trap spectra from 319,578 unique peptides was used to develop sequence-driven peak ranking models for peptides with charge states of 1 to 3. The models were incorporated into an algorithm for predicting relative fragment ion intensity (81). Table 1.1 demonstrates the correlation between results from this program and experimentally derived MS/MS product ions identified by a sequence database

search. The majority of observed fragments were predicted to be among the most intense; however the predicted order of intensity was inaccurate. For instance, four of the six major products of AEFVEVTK were ranked in the top six while only eight of all forty observed ions matched their predicted rank. The difference was most pronounced where neutral losses were prominent, as with YNGVFQECCQAEDK. It is worth noting that the instruments on which the training data and the data shown here were acquired had comparable resolution and mass range and therefore comparable quality. The performance of this algorithm points to the strengths and weaknesses of a predictive approach to spectral interpretation or SRM transition design. Spectrum derived fragmentation models revealed aggregate sequence effects but their failure to correctly rank product ions illustrates the importance of sequence-naive peak scoring systems in spectral interpretation. In fact, combined peak ranking and sequence database searching significantly increased the number of spectra identified with high confidence (82).

Protein/Peptide Sequence	Precursor m/z (Th)	Precursor charge state	Product ion m/z (Th)	Observed product ions	Predicted Intensity Rank
***Bos taurus* Serum Albumin**					
K.AEFVEVTK .L	462.13	2+	**722.41**	**y6**	**1**
			575.34	y5	3
			476.27	y4	4
			347.23	y3	7
			173.09	a2	NA
			201.09	b2	2
K.HLVDEPQNLIK.Q	653.95	2+	956.50	y8	2
			841.48	y7	6
			251.15	b2	4
			712.44	y6	3
			332.18	a6+2	NA
			465.25	b4	7
R.KVPQVSTPTLVEVSR.S	820.97	2+	706.89	y13+2	<25
			900.51	y8	1
			1088.59	y10	2
			756.43	y14+2	<25
			1150.68	b11	11
K.YNGVFQECCQAEDK.G	874.49	2+	**1314.51**	**y10**	**1**
			910.34	y7	4
			1039.38	y8	2
			1150.41	y*9	<25
			462.22	y4	6
			1167.44	y9	3
			893.31	y*7	<25
			1470.60	**y12**	**8**
			750.31	y6	5
K.CCAADDKEACFAVEGPK.L	643.66	3+	741.28	b*13+2	<25
			804.3 or 804.4	b*7 or y15+2	<25
			776.27	a*7	NA
			747.40	y7	8
			768.85	y14+2	<25
***Saccharomyces cerevisiae* Alcohol Dehydrogenase**					
K.SISIVGSYVGNR.A	626.47	2+	**752.37**	**y7**	**1**
			851.44	**y8**	**2**
			1051.55	y10	7
			445.25	y4	6
			608.32	**y5**	**5**
			695.35	y6	9
K.VVGLSTLPEIYEK.M	724.45	2+	**778.40**	**y6**	**1**
			1079.56	**y9**	**2**
			1249.67	y11	8
			439.22	y3	7
			891.48	y7	4
			341.3 or 341.2	a4 or y5+2	NA
			992.53	y8	5

Table 1.1: A comparison of observed and predicted product ion intensity. Tryptic and semi-tryptic peptides from *B. Taurus*-derived serum albumin and *S. cerevisiae*-derived alcohol dehydrogenase were separated by nanoflow HPLC on a 17 cm long, 75

micrometer diameter C18 column then fragmented by CAD on an AB Sciex QTRAP4000^{TM} triple quadrupole linear ion trap. The product ions from each peptide ion are listed from highest to lowest observed peak height. All product ions were identified by Mascot with a fragment mass tolerance of ± 0.6 Da. Relative intensities of up to 25 product ions from each peptide were estimated using the Pepnovo Predict Fragmentation algorithm (81). The precursor ion charges defined in the prediction input were those observed in the experimental data set. Product ions considered by the algorithm included *b*-ions, *y*-ions with and without water and ammonia loss. All cysteine residues were considered carbamidomethylated. Ions for which the predicted and observed relative intensity were in agreement are shown in bold. Tryptic cleavage sites are denoted with periods. Water or ammonia loss is denoted by an asterisk.

Automated sequence assignment from MS/MS data

Modern instruments require milliseconds or seconds to acquire a mass spectrum, thus the rate at which sequence information is collected far exceeds the time required for manual peptide identification. The quantity and variable quality of peptide MS/MS data necessitates bioinformatics solutions to spectral interpretation. Automated algorithms are used to differentiate the spectra of peptides from those of non-peptides, to match ion m/z values to amino acid masses, to match identified peptides to the proteins from which they are derived and to measure the statistical significance of sequence assignments. Many programs providing these functions are available. Each has strengths and weaknesses so that searching a single dataset using different programs can yield different sequence information (83).

The so-called heuristic algorithms, including SEQUEST (Thermo) and X!Tandem, count the number of ions common to experimental and theoretical tandem mass spectra while probabilistic algorithms like Mascot (Matrix Science) consider the likelihood that a set of ions is randomly matched to a peptide accounting for factors such as sequence overlap. Both types of programs rely on extensive protein libraries created primarily by *in silico* translation of genomic data to produce theoretical tandem mass spectra. Sequence can

also be assigned by matching experimental spectra to reference spectra or by *de novo* sequencing, in which experimental data are compared to theoretical product ions generated using known amino acid masses and fragmentation rules (84, 85). De novo sequencing overcomes the main disadvantage to the other two methods, namely that those polypeptides absent from the database cannot be recognized. Spectral library searching outperforms sequence library searching in the number of peptides correctly assigned, though the latter remains the dominant technique. The superior results of spectral libraries have been attributed to the incorporation of accurate product ion intensity and unpredictable product ions into the search (86).

In general, sequencing algorithms compare the masses of precursor ions and their product ions to some reference then score each match. Several qualities of the raw data are critical to the speed and accuracy of this process. Where the charge state of an ion is unknown, its mass is ambiguous. The accurate assignment of charge state therefore reduces the number of possible identities of any ion, decreasing the complexity of the search and improving the confidence of the result. The isotopic distribution of a population of otherwise identical ions is reflected in their isotopic envelope, within which the peaks are one Dalton apart in mass and differ in abundance in a pattern coinciding with their molecular formula. Charge is therefore determined by the difference in m/z of peaks in the isotopic envelope. One reason high resolution is advantageous for protein sequencing is that charge state can be unambiguously determined from isotopic envelopes. Some degree of error in mass measurement is inevitable. Search algorithms account for this error with the use of tolerances, mass ranges over which experimental and theoretical masses are considered equivalent. The differences between actual and theoretical masses serve as metrics for confidence in the match. It follows that mass accuracy determines search complexity as well as the degree to which experimental and theoretical masses match. Highly accurate data can be treated with lower mass tolerances, reducing the number of matching theoretical ions and increasing the confidence of each match. Finally, confidence in an identification increases with the number of fragments matching the hypothetical sequence so the number of product ions

at high s/n is important. The fraction of all ions in a mass spectrum matching the hypothetical sequence is another indicator of accuracy so that confidence is affected by noise. Fragmentation conditions, instrument sensitivity and mass selectivity determine these variables. Excessive collision energy degrades product ion signal while wide precursor transmission range increases noise by allowing multiple species into the mass analyzer, creating chimera spectra (87).

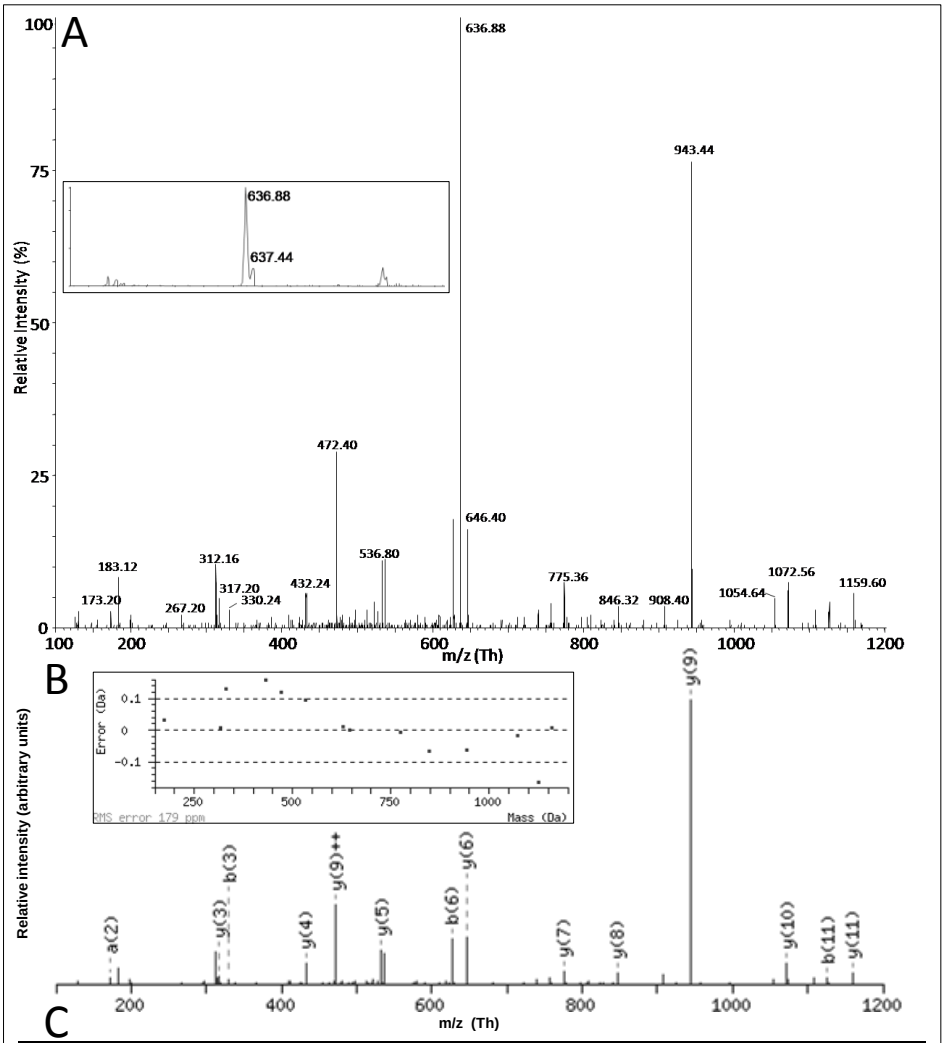
Regardless of the method by which the product ion spectrum is matched to a peptide sequence, identification requires determination of the precursor ion mass. Mass may be estimated where the precursor ion scan lacks baseline resolution of the isotopic envelope. Fragment ions of fully tryptic peptides are mostly singly charged, so observation of their isotopic envelopes is less important.

All search algorithms modify MS/MS spectra to facilitate their manipulation. Filtering out peaks below an s/n threshold may take place at this stage to reduce the size of the dataset. The simplest approach to processing is to maintain all mass spectral information, for what is known as uninterpreted searching, which is used by algorithms such as Mascot (88). Uninterpreted searching permits the identification of low intensity product ions and allows probability-based scoring of the peptide identification at the cost of increased computational burden.

The search space, the total number of solutions to a query, can be reduced by cursory peptide sequence tag searches to generate from a whole proteomic database a smaller set of hypothetical identities. A peptide sequence tag consists of several product ions in an MS/MS spectrum representing a short continuous series of residues. Since a tag is specific to the peptide ion from which it originated and will appear a limited number of times within a database, confining possible spectral identities to those containing the tag sequence is an effective means of simplifying the search (89). The use of tags also circumvents error caused by amino acid modifications. In contrast, searches for candidate sequences based on precursor mass, such as those performed by SEQUEST, must either accommodate all possible modifications or exclude modified peptides (90). The database

search engines ProteinPilot (AB Sciex) and Spectrum Mill (Agilent) are examples of programs employing peptide sequence tags (89, 91). These two software packages demonstrate the versatility of the tag approach since ProteinPilot uses uninterpreted data while Spectrum Mill uses both uninterpreted and processed data.

Figure 1.4A demonstrates that MS/MS spectra can contain numerous ions to which no portion of a peptide sequence can be assigned, including many low intensity ions distributed over the entire m/z range. Peaks at 183.12, 312.16, 536.80, 908.40 Th do not match the predicted ion series despite their relatively high intensity compared to the known *b*-and *y*-ions. These may be noise or peptide fragments undefined by the fragmentation model used in the algorithm. The Mascot algorithm groups peaks into fragmentation series then determines the probability that all ions in a series simultaneous occurrence is random. Only groups of fragments exceeding this confidence contribute to the peptide score (Figure 1.3C). This excludes the majority of spectral noise from consideration. The inset of figure 1.4B shows the distribution of mass accuracy over the mass range for the identified product ions. Restricting the product ion mass tolerance to $\pm$0.05 Da in this case would result in the rejection of seven fragments, the most abundant y_9 ion included. Expanding the mass tolerance beyond $\pm$1.3 Da would add incorrect identifications without affecting the number of true identifications, thus mass tolerance should be set to the practically achievable limits of the mass spectrometer to maximize reliable sequence coverage. Spectral noise can randomly match the masses of known product ions, leading to false positive identifications. This problem is most pronounced when peaks of low signal intensity are included in the search.



#	a	a^{++}	b	b^{++}	Seq.	y	y^{++}	y*	y*$^{++}$	#
1	86.10	43.55	114.09	57.55	L					12
2	173.13	87.07	201.12	101.07	S	*1159.58*	580.30	1142.56	571.78	11
3	302.17	151.59	330.17	165.59	E	*1072.55*	536.78	1055.53	528.27	10
4	399.22	200.12	427.22	214.11	P	*943.51*	472.26	926.48	463.75	9
5	470.26	235.63	498.26	249.63	A	*846.46*	423.73	829.43	415.22	8
6	599.30	300.16	627.30	314.15	E	*775.42*	388.21	758.39	379.70	7
7	712.39	356.70	740.38	370.69	L	*646.38*	323.69	629.35	315.18	6
8	813.44	407.22	841.43	421.22	T	*533.29*	267.15	516.27	258.64	5
9	928.46	464.73	956.46	478.73	D	*432.25*	216.63	415.22	208.11	4
10	999.50	500.25	1027.49	514.25	A	*317.22*	159.11	300.19	150.60	3
11	1098.57	549.79	1126.56	563.79	V	246.18	123.59	229.15	115.08	2
12					K	147.11	74.06	130.09	65.55	1

Figure 1.4: Mascot search result identifying the fully tryptic peptide LSEPAELTDAVK from a tryptic digest of purified endogenous human prostate-specific antigen (PSA). The MS/MS spectrum was obtained from 100 fmol of a PSA digest during an LC-MS experiment. The protein was identified by the algorithm from the single spectrum shown

above. Mascot search parameters were as follows. Peptide mass tolerance: ±1.2 Da, Fragment mass tolerance: ±0.6 Da, Enzyme: Trypsin, Protein database: SwissProt October 2010 (521,016 sequences; 183,900,292 residues). Mascot statistics for the sequence assignment from this spectrum were as follows. Expect: 3.3e-05, Ions Score: 71, Fragment ion matches: 14/88, calculated peptide mass: 1271.6609 Da. Peptide mass deviation: -0.0354 Da. A: The original MS/MS spectrum. Inset of A: The precursor ion isotopic envelope showing the monoisotopic and first heavy isotopic peaks having a mass difference of 0.56 Th. B: The MS/MS spectrum shown in the Mascot result. As shown in the figure, a 4 Th window centered on the precursor ion m/z is omitted from the search. B inset: the mass error of each identified product ion in Daltons. C: The Mascot generated table of theoretical product ions showing the ions used in scoring. Ions shown in bold numbering match theoretical product ion masses but cannot be distinguished from random peaks. Ions considered to be derived from the identified peptide sequence are shown in large bold numbering. Ions included in the scoring are shown in large italicized bold numbering.

Sequence assignments from MS/MS spectra vary considerably in their reliability. Statistical evaluations of search results aid in their interpretation. Numerous tests fulfil this requirement. One of the most popular is false discovery rate estimation, which entails finding the distribution of scores for a mixed set of true and false positive results as well as the distribution of scores for a set of false positive results then determining the overlap of these score populations to establish limits of confidence. False positive sets are generated by searching the data against a randomized or reversed version of the same protein library used to find true positives under the assumption that this will result exclusively in incorrect assignments. The mean score of the true positive distribution is higher than the mean of the distribution generated from the decoy library search so only true positives exceed a certain threshold value (92). Cutoff scores based on perceived false positive to true positive ratios can be used as filter criteria to balance sensitivity with specificity, which are defined in the following equations (83).

$$algorithm\ sensitivity = \frac{true\ positive}{true\ positive + false\ negative}$$

$$algorithm\ specificity = \frac{true\ negative}{true\ negaitve + false\ positive}$$

Here a true positive is a correct sequence assignment with a score above the cutoff, a true negative is an incorrect assignment with a score below the cutoff, a false positive is an incorrect assignment exceeding the cutoff and a false negative is a correct assignment beneath the cutoff. A cutoff score at 99% confidence corresponds to the value at which one in one hundred sequence assignments is false and is clearly superior in specificity than a cutoff of 95%, however the latter may be more sensitive, giving more true positives. Increase in content sometimes makes relaxing cutoff stringency desirable.

There is no universal randomization method for decoy database creation, and the estimated false discovery rate may vary with the method used (93). One alternative is to search random MS/MS spectra against a proteomic library. Compared to reversed database searching, this was shown to improve protein identification by differentiating the peptide-to-protein ratios of authentic and false positive sequence assignments (94). One metric of accuracy that avoids false discovery rate estimation altogether is the expectation value, the probability that a given score would occur by chance based on the size of the search space and the score distribution of all hypothetical sequences matching a given spectrum. Since it is independent of the means by which a score is calculated, the expectation value can be applied to results from any spectral matching function (95). Mascot and X!Tandem both calculate expectation values but in different ways. For each spectrum, X!Tandem plots the number of hypotheses per score and extrapolates to the value of the presumed true positive to find its frequency, while Mascot directly relates the number of hypotheses per score to a user defined confidence interval then reports the difference between the highest score and that corresponding to the limit of the confidence

interval (83, 96). Interestingly, the two programs differ in their sensitivity to search space (97).

Search algorithms are integral to proteomics workflows and each has its own nuances. The direct comparison of datasets from different groups is complicated by the use of different of spectral interpretation software under different settings. The lack of a universal standard can in part be explained by the variety of solutions to polypeptide inference. Fortunately the most widely accepted algorithms tend to give complementary rather than contradictory outputs.

1.4 Special considerations in polypeptide analysis

Proteins are large, reactive molecules naturally forming in intricate heterogeneous mixtures and as such many obstacles to their analysis continue to confound elucidation. These technical issues remain the focus of much study while biological discoveries continue to be made from LC-MS data in spite of them. Some details of proteomic analysis of are discussed below.

Influence of the sample matrix on LC-MS

Diverse sample composition can profoundly affect the success of polypeptide LC-MS. Analytes and contaminants interact during analysis to alter separation and sensitivity in both chromatography and mass spectrometry. HPLC systems are susceptible to shifts in retention time and peak shape caused by column overloading. Stationary phase fouling by non-polypeptides, exemplified by the detrimental effect of plasma lipids on blood analysis, is another common problem degrading overall performance (98).

Interference also occurs at the ESI source, resulting in ion suppression or ion enhancement known as the matrix effect (99). This decreases dynamic range and sensitivity for certain compounds as manifested in the detrimental effect of increased sample loading amount on the number of peptide identifications from a single sample (100). Two sources of the matrix effect have been identified; competition between

sample components for charged species such as protons and disruption of droplet formation or collapse (101). Detergents and non-volatile salts suppress ESI by competing with analytes for ionization and forming deposits on ion optics. MALDI is also impaired by detergents so the problem may not be circumvented by changing sources (102). Peptides compete for ionization in a sequence and concentration dependent manner. Over the course of an LC-MS run, the effect is most pronounced where coelution is highest, usually at high organic solvent concentration (103-105). Even heavy isotope labelled internal standards, which control for variation in ionization efficiency, cause matrix effects at high concentrations (106). The problem varies among highly similar matrices such as plasma or urine from different individuals (107, 108).

The elimination of all non-proteinacious components of a sample prior to LC-MS is highly desirable. Dialysis and reversed phase extraction both remove water soluble inorganic compounds with minimal polypeptide loss. The removal of detergents is more problematic because of their limited solubility and high affinity for reversed phase materials. Solid phase extraction and liquid phase extraction techniques for the reduction of detergent content exist but are not entirely effective (109-113). Proteins of no relevance to an investigation can cause severe matrix effects, for instance the high abundance proteins of blood hamper detection of less concentrated biomarkers. Being composed of the same amino acids, analytes of interest and proteinacious contaminants are not easily isolated. Selective removal of either the analyte or abundant matrix proteins is possible by affinity purification. It should be noted that the elimination of abundant species from a complex proteome leaves behind a great deal of the matrix which may still hinder LC-MS. Figure 1.5 shows the protein distribution of human blood serum before and after removal of the most concentrated proteins. Untreated sera run in lanes two and three have wide bands characteristic of the most abundant species while the treated serum in lane four does not; however most bands are common to all three lanes. A Bradford protein assay standardized against bovine gamma globulin indicated the protein content of the treated serum to be approximately 20% that of the untreated serum.

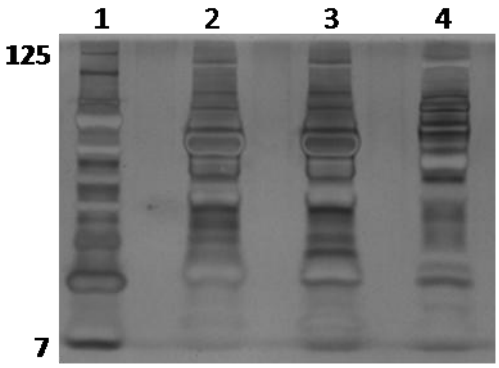


Figure 1.5: Human serum resolved on a 10% acrylamide SDS-PAGE gel stained with silver. 1: 15 ul of 7 to 125 kDa molecular weight marker from New England Biolabs, 2: 100 nl of normal male type AB serum (Sigma Aldrich), 3: 100 nl of serum from an endometrial cancer patient, 4: 1 ul of serum from the same sample as shown in lane 3 after two passages through a ProteoPrep^{TM} immunodepletion spin column (Sigma Aldrich). The immunodepletion column was designed to remove twenty abundant plasma proteins including albumin, Immunoglobulins, transferrin, fibrinogen, compliment components and apolipoproteins.

A sample can continue to influence the LC-MS system long after it has been injected onto the column and analyzed. Elution of analytes from one sample during subsequent chromatographic runs is called carryover. This can be a significant source of error in proteomics experiments, complicating cross-sample comparison. An empirical limit of acceptable carryover for contemporary reversed phase HPLC platforms is an analyte signal intensity ratio of 1:100 between a blank and the sample run preceding it. Large and hydrophobic peptides tend to persist, eluting over numerous runs with signal intensities near their initial values. It follows that extensive washing with acetonitrile or water cannot eliminate the issue. The order in which samples are analyzed is important in label free quantitation in light of these facts. Contamination is minimized with the use of trifluoroethanol or isopropanol, although the latter must be flushed from the system to restore chromatographic performance (114).

Variability in enzymatic digestion efficiency

Most bottom up proteomic studies employ more or less the same trypsin digestion protocol with the implicit assumption that it has substrate and matrix independent high efficiency. Several studies have shown this general approach to sample handling to be suboptimal. Even pure proteins vary in their susceptibility to trypsin at different cleavage sites under different conditions to an extent that sensitivity of quantification differs with the peptide measured (115). In plasma, some proteins resist degradation regardless of the detergent, denaturing agent and treatment time used (116). Such findings suggest treating one sample with several proteolytic conditions would be beneficial to sequence coverage. One surprising development of trypsin research is the emergence of protocols requiring high organic solvent concentration. Since trypsin is a mammalian protein, it seems counterintuitive that it would retain its function at acetonitrile or methanol contents as high as 80% (v/v) however such conditions have been shown to dramatically enhance its rate of catalysis (117-119). Microwave heating is another innovation shown to increase digestion speed, though conventional incubation in water baths or ovens at 37^{o}C is still widely used (120). Regardless of the experimental system it is important to recognize the uncertainty in proteolysis introduced by enzymatic cleavage.

Amino acid modifications

Proteins undergo an assortment of chemical modifications *in vivo* and *in vitro*, mostly at amino acid side chains. Such alterations range in complexity from the addition of single atoms to macromolecules. Biological amino acid modifications within a protein are referred to as post-translational modifications (PTMs). Modification can obstruct protein sequencing and quantification by altering the mass and physical properties of amino acids on which they occur. Modified forms of a peptide reduce or eliminate the MS signal of the unmodified version, which is especially problematic in peptide quantification. The inclusion of PTM-associated hypothetical mass shifts in spectral interpretation models increases computational burden and the number of spurious sequence matches. In practice, only a few common modifications are considered in automated searching in

order to maximize true positive identifications while minimizing false positives, though such a data interpretation paradigm concedes partial sequence coverage of modified proteins. The biological relevance of certain modifications may be obscured by their occurrence both *in vivo* and as artefacts of sample handling, so strict controls are essential to the study of modified peptides. Several protein modifications and their impact on LC-MS and proteomics in general are discussed below.

Glycosylation

The majority of polypeptides are believed to be functionalized with monosaccharides or polysaccharides in what is recognized as the most diverse and common type of post-translational modification (121). The prevalence of protein glycosylation attests to its metabolic importance and a host of biological systems have been shown to depend on this post-translational modification. Glycosylation is both a source of considerable biological information and a hindrance to protein sequencing. Covalently bound sugars interfere with automated spectral interpretation by altering polypeptide mass and fragmentation chemistry. Glycopeptides readily ionize by ESI but current sequencing algorithms cannot discern their highly complex and variable glycan moieties. Appreciation of carbohydrate-protein bonding therefore aids the design of proteomics experiments and interpretation of their results.

Polysaccharides are ethers resulting from condensation reactions of monosaccharide hydroxyl groups forming what are commonly referred to as glycosidic bonds. Since they are polyols, monosaccharides can polymerize at multiple positions to form extensively branched glycans. While leucine and isoleucine are the only structurally isomeric amino acids, many monosaccharides are structural isomers. As a result, glycans of markedly different sizes and shapes are isobaric. A glucose residue has a mass of 180 Da, while glycans can reach masses of several thousands of Daltons, so the mass range added to a peptide by glycosylation is wide. Characterization of specific glycoproteins has revealed that polysaccharide moiety structure reflects glycan synthesis and transport machinery, host-pathogen interactions, genetic variation and disease state so a single amino acid

residue within a protein has the potential to be linked to a multitude of glycans (121, 122). For these reasons, glycosylation complicates protein analysis.

Glycosylation occurs at the side chain nitrogen of asparagine residues as well as the side chain oxygen of serine or threonine residues giving N-linked and O-linked glycosylation respectively. Where the amino acid sequence of a glycoprotein is sought, glycans can be removed by several means. Beta-elimination with bases such as sodium hydroxide or ammonium hydroxide is a simple, MS-compatible method of removing O-linked glycans (123). For the removal of N-linked glycans, a class of enzymes known as endoglycosidases can be used (124). The bacterial protein peptide N-glycosidase F (PNGase F) specifically removes N-linked glycans under conditions commonly used for proteolytic digestion, converting asparagine residues to aspartate in the process. Deglycosylation has a direct influence on peptide identification. For example, an 18% improvement in sequence coverage of influenza virus antigens was reported when PNGase F treatment was incorporated in the sample preparation (125). It is worth noting that glycopeptides can be characterized using the same instrumentation normally applied to peptide analysis (126, 127). In principle, glycopeptide and peptide identification could be integrated to enrich the structural detail of proteomic datasets while simplifying sample preparation.

Phosphorylation

Protein phosphorylation is a reversible signalling mechanism in eukaryotes and prokaryotes. Two classes of enzymes control phosphorylation, kinases which transfer phosphates from adenosine triphosphate (ATP) to protein substrates and phosphatases which remove protein-bound phosphates. In addition, protein phosphorylation by small molecules such as acetyl phosphate occurs in bacteria (128).

Around one third of the vertebrate proteome is thought to undergo this post translational modification (129). In vertebrate proteins, serine is the most commonly phosphorylated amino acid, followed by threonine then tyrosine, with the respective ratio of phosphorylation being approximately 1800 to 200 to 1 (26). The side chain hydroxyl

groups of serine, threonine and tyrosine are bound to phosphate as relatively stable phosphoesters. In bacteria, aspartate and glutamate residues constitute the main phosphate acceptors, forming labile phosphoanhydrides(130). Higher organisms are also known to produce phosphoanhydrides. Prothymosin alpha, a highly conserved mediator of immune function in eukaryotes, is also phosphorylated at numerous glutamic acid residues (131). Side chain nitrogen atoms of arginine, lysine and histidine form a third group of phosphoamino acids, the phosphoramidates, which occur in both prokaryotes and eukaryotes (132).

Phosphosites tend to be modified at substoichiometric levels so the recognition of phosphopeptides has little impact on sequence coverage in most proteomic contexts. The roles of phosphate transfer in metabolism are explored in the related field of phosphoproteomics. Phosphoproteomics workflows typically involve purification of phosphopeptides from proteolytic digests before LC-MS. Both the isolation and identification of phosphopeptides present barriers distinct from those encountered in proteomics. Phosphopeptides behave differently from peptides having identical sequences in both LC and MS. Phosphates are acidic, reducing phosphopeptide ionization efficiency and charge in positive ion mode (133-135). The phosphate group is labile in the gas phase, especially under CAD conditions, and reveals little structural detail after detachment from the phosphopeptide (133). In addition, gas phase phosphate mobility leads to intramolecular relocation in those species having multiple serine, threonine or tyrosine residues (136). These phenomena impede phosphosite assignment. Routines for phosphopeptide MS are diverse and include neutral loss scanning to monitor phosphate dissociation as well as MS/MS in both negative and positive ion mode. Phosphoserine and phosphothreonine derivatization producing less labile functional groups can be used to improve detection (137).

Though their physical properties make for difficult MS analysis, phosphopeptides are well resolved by reversed phase HPLC, having equivalent peak widths but longer retention times on reversed phase over acetonitrile gradients compared to non-phosphorylated peptides with the same sequence(138). In light of the susceptibility of

phosphopeptides to ion suppression, a problem exacerbated by the presence of other analytes, LC is warranted in phosphoproteomics.

Deamidation

Deamidation is the spontaneous rearrangement of the asparagine and aspartic acid side chains through reactive succinimide intermediates following the loss of NH_2^- and OH^- respectively (139). The succinimide ring opens upon addition of a hydroxyl group, resulting in mass shift of 1 Da for asparagine or no mass shift for aspartate. Glutamine and glutamate react analogously, with only the former gaining 1 Da through deamidation (140). The process is much more rapid for aspartate than asparagine (139). The reaction occurs under physiological conditions and is reversed *in vivo* by the enzyme L-isoaspartyl O-methyltransferase, an important reaction in the preservation of protein function since deamidation has been shown to promote protein degradation (141). Deamidation also takes place during sample preparation and sequence dependence has been described, with glycine residues apparently promoting the modification during trypsin digestion when C-terminal to asparagine (142). The proposed mechanism of deamidation involves nucleophilic attack by the amide nitrogen in the peptide bond C-terminal to the affected residue on the Asn or Asp side chain carbonyl carbon (143).

Peptides containing deamidated asparagine and glutamine residues have longer retention times than their unmodified equivalents on C18 columns with water/acetonitrile mobile phase (144). Partitioning of the MS signal for a given peptide into unmodified and deamidated forms is potentially detrimental to quantification and sensitivity. Mass shifts associated with this PTM are commonly considered in MS/MS sequence identification. This effect can be mitigated by performing digestion under mildly acidic conditions (145).

Oxidation

Both biological and abiotic influences affect the oxidation of proteins. The sulfur atoms of cysteine and methionine are readily oxidized. The tryptophan side chain undergoes a

number of oxidation reactions. Addition of one or two oxygen atoms produce hydroxytryptophan and dihydroxytryptophan respectively. Extensive side chain modification yields kynurenine and several derivatives thereof, all of which have masses exceeding that of tryptophan. While tryptophan derivatives are biologically regulated and active, all of the above tryptophan modifications have been shown to result from SDS-PAGE separation by MS/MS of tryptic digests (146).

Disulfide bonding

Disulfide bonding between the side chains of two oxidized cysteine residues presents a challenge to bottom up sequencing, covalently linking distant regions of a protein or two separate proteins. Digestion of sulphur bridged polypeptides yields non-linear peptides with masses and product ions representative of both contributing regions, which present challenges to identification. Treatment with reducing agents such as dithiothreitol (DTT) followed by alkylation controls cysteine reactivity, eliminating disulfide bonds and preventing subsequent oxidation of the sulfur atoms. Reduction and alkylation are practically stoichiometric and widely used in the field of proteomics. Derivatization is essential because sulfhydryl groups spontaneously bond. The technique is not effective for methionine derivatization.

Alkylation

Excess of cysteine derivatization agents can cause alkylation of the N-terminal amine as observed with iodoacetamide (147). This modification is minimized by limiting alkylation agent concentration. *In vivo*, methyltransferases mediate the alkylation of arginine and lysine residues, particularly in proteins controlling gene expression (148). As many as two or three methyl groups may be added to arginine and lysine respectively.

Ubiquitylation

The covalent attachment of two proteins is not restricted to disulfide bonding *in vivo*. In eukaryotes, the primary function of members of the ubiquitin protein family is to regulate other proteins in an enzyme catalyzed process known as ubiquitylation. Such reversible

PTMs maintain homeostasis by marking the modified protein for degradation, altering its activity or its subcellular location (149). Ubiquitylation is known to occur mainly at lysine side chains and to a lesser extent cysteine and amino-terminal residues. Ubiquitin itself harbours seven lysine residues by which it forms homopolymers, the size of which affect the fate of its substrates (150). A diverse group of ubiquitin homologues such as the small ubiquitin-like modifier (SUMO) perform similar roles through analogous mechanisms.

The consequences of ubiquitylation in proteomics experiments include lost sequence coverage due to missed cleavages and unanticipated mass shifts. Trypsin cleavage at lysine is blocked by ubiquitylation in some cases. When fully cleaved, ubiquitin and ubiquitin-like proteins leave di-glycine with a mass of 114 Da on the modified residue. Assigning ubiquitylation sites based on the di-glycine modification is challenging since it does not change the fragmentation pattern of the peptide and is isobaric with 2-acetamidoacetamide, an artefact of iodoacetamide alkylation (151, 152).

1.5 Conclusion

The emergence of proteomics over the past several decades has resulted from the creation of complementary analytical technologies centered on mass spectrometry. Demand for knowledge of polypeptide elemental composition has been met to such a degree that instrument accessibility, rather than quality, has become the major limiting factor in scientific advancement in proteomic applications. The current generation of LC-MS platforms could inform research into protein expression, localization, interaction and post-translational modification for decades to come but improvements in all characteristics of LC-MS performance are ongoing. Bottom-up proteomics stands to benefit especially from advancements in sensitivity, separation and data processing. MS sensitivity and LC separation are complementary in that the negative impact of sample complexity on the detection of any analyte is reduced as both performance characteristics are increased. Current high quality LC separation incompletely resolves all peptides from a typical proteomic sample and MS scan speeds which could make up for the shortfall are

orders of magnitude higher than those currently available (153). The separate acquisition of MS/MS data for all peptides in a single LC run will require either a leap in scan speed and sensitivity or chromatographic resolution. SWATH MS has demonstrated the potential to collect data from all peptides in a sample without the need for improvements in instrumentation and represents a step toward a unified analysis mode with combined high-throughput identification and quantification. Analytics are crucial here, since the relationship between precursor ions and product ions is obscured by the presence of other analyte ions. Even with only a fraction of the observable proteome accessible in a single run, LC-MS/MS experiments generate vast amounts of raw data which must be organized into a meaningful form. From sequence assignment to the recognition of post-translational modifications to quantification, fast and transparent data processing is vital to mass spectral interpretation and its simplification may encourage the wider use of LC-MS based proteomics in biology.

1.6 Statement of Purpose

There are countless uses for LC-MS/MS in the study of protein biochemistry and proteomics. A diverse collection of inquiries centered on LC-MS/MS technology is described in the following chapters, with subjects ranging from specific metabolic events in prokaryotes, plants and animals to theoretical limits of sequence specificity in MS-based proteomics. The main purpose of this dissertation was to solve problems unique to each biological investigation using various LC-MS/MS platforms. A separate theoretical study comparing mass accuracy in MS and MS/MS to MRM transition redundancy was undertaken with the objective of better defining the relationship between proteomics sample heterogeneity and the accuracy of MS/MS-based experiments in terms of peptide sequence.

The scientific results central to this dissertation consist of sequence information derived from data-dependent acquisition and structural information derived from targeted MS/MS. Attributes of the MS/MS datasets on which their interpretation, and the credibility thereof, are based are discussed in order to relate mass spectrometer

performance to data quality. Similarly, LC performance is related to reproducibility and reliability of peptide characterization through illustrations of extracted ion chromatograms and discussions of their properties. Though LC-MS/MS systems may sometimes be regarded as "black boxes" by those who employ them, the data they generate should be assessed on a technical basis in the interest of accurate scientific reporting. Examples of such assessments are presented in the following chapters to show how the qualities of mass spectra influence the confidence of each biological finding.

In the projects described in chapters 2, 3 and 4, LC-MS/MS was used to evaluate the potential of specific amino acids within specific proteins to be phosphorylated. Direct observation of specific amino acid residues with and without phosphates bound to them was achieved with MS/MS, whereas the presence of the phosphate could only be inferred using complementary biochemical techniques. Novel phosphosites identified were shown by other analytical techniques to be relevant to specific protein functions. Chapter 4 details the finding that an assumed phosphosite was not phosphorylated by a particular kinase in a demonstration that LC-MS/MS can be used to disprove as well as prove a hypothetical modification. Aside from phosphosite determination, MS/MS-based polypeptide sequencing was employed to address questions relating to each of the proteins under study, as summarized below.

Chapter 2 describes a study of two transcription factors of the human pathogen *Staphylococcus aureus* whose activities are known to impart resistance to antimicrobial agents. The novel demonstration of phosphorylation at specific residues within both proteins was central to the characterization of their functions in response to antibiotic stress. Chapter 3 is focused on the identification of three phosphosites within an enzyme integral to starch biosynthesis in *Zea mays* and of several kinases responsible for modifying these residues. Direct observation of the phosphorylated residues and the assignment of putative kinases provided a basis for further investigation into the biochemistry of the substrate protein and the protein complex to which it belongs. Chapter 4 describes the discovery of novel interacting proteins for an intracellular signalling protein within a mammalian cell line through a combination of immunoaffinity

chromatography and LC-MS/MS. Additionally, while a novel kinase-mediated modulation of the signalling protein's activity was illuminated by *in vitro* techniques; LC-MS/MS analysis indicated the effect of the kinase was indirect, prompting exploration of an alternative regulatory mechanism. Chapter 5 differs from the others in that it is broadly focused on limitations of tandem mass spectrometry for polypeptide analysis and details a study which employed computation rather than LC-MS/MS in the generation of exploratory data.

LC-MS/MS was an integral component of each research program discussed here, exposing protein-protein interactions and mechanisms of protein regulation. Corroborative data obtained by other techniques is presented to support conclusions drawn from mass spectra. The selected topics covered here reflect the current understanding of the mass spectrometry of polypeptides.

Chapter 2

Phosphorylation-mediated activation of *Staphylococcus aureus* antibiotic resistance response regulators

2.1 Introduction

Antibiotic tolerant pathogens such as methicillin-resistant *Staphylococcus aureus* present a threat to public health, causing hundreds of thousands of potentially life-threatening infections annually (1). To survive antibiotic challenge, a pathogen must sense the stress then coordinate a suitable response. Two-component systems (TCSs), which consist of paired sensor and response regulator (RR) proteins, confer resistance to several classes of antimicrobial agents by mobilizing activities such as drug efflux and cell wall synthesis (2, 3). The sensor protein of a TCS functions as a transmembrane receptor, a kinase and a phosphatase while the RR is typically a transcription factor with an entirely cytosolic distribution. Following recognition of the external stimulus, signal transduction begins with sensor autophosphorylation at a histidine residue. The receptor then activates the RR by phosphorylating it at an aspartate residue. The signal is negatively regulated by phosphatase activity of the receptor on the RR or by RR autophosphatase activity or by other phosphatases (4). Seventeen TCSs have been identified in the ubiquitous Gram-positive bacterium *S. aureus*, of which ApsSR/GraSR, BceSR, BraSR, and VraSR are associated with antibiotic resistance (5-7).

Several distinct properties of TCSs influence the means by which their behaviour can be studied. Both components are homodimers in their active states and the activating phosphate is transferred from ATP by one sensor monomer to its pair, then to the RR. Stimulus is thought to elicit a low degree of sensor autophosphorylation so the degree of sensor protein phosphorylation is unrepresentative of pathway activation (4). RRs

possess potent autophosphatase activity toward their activating aspartate residues, causing rapid turnover of the PTM (8). Furthermore, the aspartate phosphoanhydride group is unstable and readily dissociates under mild conditions in the absence of a phosphatase. Phosphoaspartate degradation occurs rapidly under acidic, basic and neutral conditions (9). Phosphoanhydride fragility and low concentration of TCS phosphoproteins present challenges to detection, however phosphoaspartate may be studied indirectly using a more stable phosphate surrogate or by derivatization. The highly electronegative tetrahedral bonding phosphate analogue beryllofluoride (BeF_3^-) forms more stable active RRs *in vitro* (10, 11). Phosphoaspartate can be reduced with sodium borohydride ($NaBH_4$) to the amino acid homoserine (Hse), though the *E.coli* protein CheY is the only RR that has been studied using the technique (12, 13). Aspartate-57 within the conserved RR activation sequence was shown to be the CheY phosphosite by a combination of reduction, proteolytic digestion, RP-HPLC fractionation and MS/MS.

Phosphoesterification is an emerging mechanism of TCS regulation which is amenable to study using established phosphoproteomics techniques. Like eukaryotes, bacteria possess serine/threonine kinases. CheY has been shown to be activated by phosphorylation at Serine-56, suggesting that more RRs may be similarly controlled (14).

Herein is reported a study of two *S. aureus* antibiotic resistance-mediating RRs, VraR and GraR, by LC-MS/MS in which their activating phosphorylations were observed. The locations and states of VraR and GraR phosphosites can only be inferred using biochemical methods and this is an impediment to the elucidation of their behaviour. LC-MS/MS can supply proof of the phosphorylation state of any residue in a protein and was therefore used here to determine serine-threonine kinase 1 (Stk1) directed phosphosites on GraR and the acetyl phosphate directed phosphosite of VraR. Acetyl phosphate modified VraR at aspartarte-55 (D55) while the endogenous Stk1 modified GraR at threonine residues 128, 130 and 149. The PTM increased the RRs affinities toward their respective promoters while site directed mutagenesis of the phosphosites inhibited functions normally under GraR control. In addition to confirming that VraR is activated by D55 phosphorylation, GraR was shown to be under the control of an entirely different

class of kinase/phosphatase pair than is typical of TCSs. This work contributed to the established model of TCS signalling and uncovered a novel means of response regulation in *S. aureus*. In addition, a sample preparation technique was evaluated to assess its applicability to the measurement of *in vivo* VraR phosphorylation.

VraSR

The VraSR TCS confers resistance to cell wall synthesis-inhibiting compounds such as vancomycin, bacitracin and beta-lactams by inducing the expression of at least 46 genes, among them those of VraSR itself as well as several responsible for cell wall construction (3, 15). Like all TCS names, that of VraSR alludes to both components, the transmembrane sensor VraS and the RR VraR.

Vancomycin resistance associated regulator (VraR), is a 209 amino acid protein with an N-terminal regulatory domain and a C-terminal DNA-binding domain, both of which are conserved among the RRs of a number of prokaryotes. Though never previously observed, D55 was assumed to be the VraR phosphate acceptor based on sequence alignment with other RRs. While the wild-type protein was phosphorylated by phospho-VraS within minutes, mutation of residue 55 to asparagine negated this effect. VraR was also shown to undergo phosphorylation at a single site between residues 1 and 133, albeit more slowly, by acetyl phosphate. Small organophosphates are known to modify regulatory aspartate residues of other RRs though the significance of such reactions *in vivo* remains to be established (16). VraS rapidly dephosphorylates phospho-VraR, reflecting negative regulation by the intrinsic phosphatase activity of the sensor. Both acetyl phosphate-phosphorylated VraR and phospho-VraS-phosphorylated VraR were affected by VraS. In the case of phospho-VraS-phosphorylated VraR, complete dephosphorylation was rapid, occurring in less than five minutes (17).

The phosphorylation of VraR results in structural changes implicated in its activation. Binding of BeF_3^- resulted in conformational remodelling which enabled VraR homodimerization as assessed by crystallography (17, 18). Phosphate-dependent conformational changes of VraR have also been observed by circular dichroism and

hydrogen-deuterium exchange MS (17, 19). DNA binding by VraR was shown to increase substantially upon phosphorylation using regions upstream of the VraSR promoter as substrates (17, 20). VraR dimerization and binding to the VraSR promoter appear to be concentration dependent, so its upregulation following VraS stimulation can be presumed to result in increased VraR activity, possibly with decreased dependence on direct PTM-dependent activation (17). One study found the cytosolic concentration of VraR increased more than three fold in vancomycin-challenged *S. aureus* compared to untreated controls, a change higher than all but one of the ten VraSR-controlled genes found to be upregulated by this stimulus (21). The above studies describe a rapid antibiotic-triggered signalling cascade in which VraS stimulates genetic expression within minutes and VraR concentration quickly reaches levels which may be sufficient for phosphate-independent promoter binding. The capacity of VraS to reverse VraR activation suggests the response to cell wall damage is under both post-translational and transcriptional control.

GraSR

The TCS GraSR, also called apsSR, acts in two ways to confer resistance to cationic antimicrobial compounds causing cell wall damage. The negative charge of the bacterial cell wall is decreased by covalent attachment of amino acids to the teichoic acids and to phosphatidylglycerol which protrude from the cell surface, in turn protecting *S. aureus* from cationic antibiotics. In addition, the membrane transporter VraFG pumps cationic drugs out of the cell. The genes responsible for both of these functions are under the control of GraSR (22, 23). GraSR differs from VraSR in that it does not regulate its own promoter and its expression is unaffected by the class of antibiotics to which it responds (24).

Glycopeptide resistance associated regulator (GraR) is 224 amino acids long. Like VraR, GraR contains a C-terminal DNA binding domain as well as an N-terminal regulatory domain in which Aspartate-51 is understood to be a phosphate acceptor. In addition,

GraR contains 14 serine, 14 Threonine and 7 tyrosine residues, none of which are known phosphosites. GraR is apparently not phosphorylated by acetyl phosphate (25).

Stk1

The *S. aureus* genome contains one known serine/threonine kinase (Stk1) and one serine/threonine phosphatase (Stp1) which dephosphorylates it (26). An extracellular sensory role for Stk1 is suggested by its transmembrane domain and three extracellular penicillin-binding protein serine/threonine kinase-associated (PASTA) domains, features common to STKs of other bacteria (26, 27). Stk1 is a known beta-lactam antibiotic resistance factor and controls the expression VraFG and VraSR among many other genes, though its substrates have not been fully described (27, 28). Notably, VraFG transcription is controlled by GraSR, suggesting that Stk1 influences GraSR activity (24). The possibility of such an interaction is reasonable given that serine/threonine kinases of other prokaryotes are known to regulate RRs (29).

2.2 Methods and Materials

Recombinant protein expression

Full length Stk1, Stp1, GraR and VraR genes from *S. aureus* strain MU50 were ligated into the plasmids then transformed into *E.coli* BL21 DE. Truncated GraR and GraR point-mutants were created using a Stratagene^{TM} site-directed mutagenesis kit. The four mutants created were the effector domain construct GraR^{C} (GraR residues 123-224), the N-termial construct GraR^{N} (GraR residues 1-134), the point mutant GraR-Thr130Ala and the triple mutant GraR-Thr128Ala/Thr130Ala/Thr149Ala. Plasmid pET26b was the expression vector for all genes with the exception of pET141/D-TOPO used for Stk1 expression. Inclusion of the expected genes was confirmed by DNA sequencing of the expression vectors. Recombinant protein expression was induced by addition of isopropyl-beta-D-thiogalactopyranoside at concentrations between 0.5 and 1 mM in the growth medium. After 16 to 18 hours Terrific Broth cultures were sedimented by centrifugation at 7459 g and lysed by sonication then centrifuged at 18138 g to pellet cell

debris. The desired proteins were isolated from the lysate supernatant. Recombinant Stk1 included an N-terminal six-histidine tag and was purified using a nickel-NTA column while the other proteins were purified using DEAE-Sepharose. The molecular weights of the recovered gene products were verified using SDS-PAGE with Coomassie blue staining.

Response regulator phosphorylation

To evaluate its kinase activity on a number of proteins, 5 µM Stk1 was incubated with potential substrates in 25 mM Tris-HCl (pH 7.5), 1 mM dithiothreitol, and 10 mM $MgCl_2$ with ATP or [γ-^{32}P]-ATP at room temperature for 30 minutes in the presence of ATP. Substrate concentrations used were as follows; 10 µM of GraR, 10 µM of the *Bacillus subtilis* GraR homolog BceR, 20 µM of VraR. The interaction was arrested by heating in SDS-PAGE sample loading buffer. The kinase and GraR were separated by electrophoresis on 12.5% acrylamide gels then visualized with Coomassie blue stain.

VraR was phosphorylated in 50 mM Tris-HCl, pH 7.0, 50 mM KCl and 20 mM $MgCl_2$ with 50 mM lithium potassium acetyl phosphate (Sigma-Aldrich, St. Louis, MO, USA) at 37°C for 60 minutes. Acetyl phosphate treatment was followed by incubation on ice during which 1 ml volumes of 10% (w/v) ice-cold trichloroacetic acid (TCA) were added to precipitate the protein. The samples were then brought to 30°C for 10 minutes then chilled and centrifuged at 10,000 g and 4°C for 20 minutes. The pellet was washed with 200 µl of ice-cold 10 mM HCl and dried in a centrifugal concentrator at low heat.

$NaBH_4$ Reduction

The TCA precipitates (approximately 100 µg total protein) were suspended in 100 µL of dimethyl sulfoxide (DMSO) containing 2.7 µmoles of $NaBH_4$ at 30°C for 10 minutes, then the reaction was terminated by the addition of 1 ml of 0.44 M ice-cold perchloric acid ($HClO_4$) on ice. Protein was collected by centrifugation at 10,000 g and 4°C for 20 minutes then washed with 200 µl of ice-cold 10 mM HCl prior to SDS-PAGE or proteolytic digestion.

Reduction of whole cell lysates

Matrix effects on phosphoaspartate reduction and LC-MS performance were determined in two ways. *E.coli* overexpressing VraR was lysed then treated with acetyl phosphate and subsequently with $NaBH_4$ without prior HPLC purification of the recombinant protein. Alternately, acetyl phosphate treated recombinant VraR was spiked into *S. aureus* RN4220 lysates in varying amounts and the lysate was then reduced for the evaluation of matrix effects on LC-MS/MS sensitivity. Each sample contained 52 mg of lysate protein with 0, 0.1, 1, 10 or 100 µg of phospho-VraR in a total sample volume of 500 µl, from which 10 µl portions were loaded onto SDS-PAGE. Borohydride-treated samples were separated by SDS-PAGE on 15% acrylamide gels and bands matching the mass of VraR excised and digested with Glu-C for LC-MS/MS analysis. To test for endogenous phospho-VraR, 40 ml Terrific Soy Broth cultures of *S.aureus* MU50 were grown at 37^{o}C to an OD_{600} of 0.6-0.7 then supplemented with 5 mg/ml oxacillin for one hour pelleted and lysed as described above. The lysates were reduced then separated by SDS-PAGE or immunoprecipitation with a rabbit anti-VraR antibody immobilized on ImmunoCruz IP/WB Optima F matrix (SantaCruz Biotechnology Inc. Dallas, TX, USA). Immunoprecipitates were collected from the support with 0.5 M NH_4OH.

Electrophoretic mobility shift assay

A segment of the S. aureus MU50 genome including the VraRG promoter region was amplified then collected and phosphorylated with [γ-^{32}P]-ATP using recombinant T4 polynucleotide kinase. Recombinant GraR was incubated with 500 µM ATP in the presence or absence of Stk1 for 40 minutes at room temperature. A five-fold excess of GraR to Stk1 was used. Following treatment with ATP, the protein was diluted to concentrations between 3 and 18 µM and mixed with 4 ng of the radiolabelled DNA segment for 30 minutes at room temperature. The mixtures were separated by native PAGE on 6% acrylamide gels and radioactivity visualized on a Typhoon Trio imager (GE Healthcare, Little Chalfont, UK).

LC-MS/MS

Protein bands were excised from polyacrylamide gels then destained with acetonitrile. Whole cell lysates were digested without reduction or alkylation, since VraR contains no cysteine residues while GraR has one. The GraR samples were digested with trypsin (Promega, Madison, WI) while VraR samples were either digested with trypsin or Glu-C (Sigma-Aldrich) overnight at 37^{o}C in 100 mM ammonium bicarbonate. Each sample was confirmed to be pH 8 prior to addition of the enzyme. In general, 0.2 to 0.5 µg of protease were used per gel band. Tryptic peptide solutions were dried in a centrifugal concentrator, dissolved in 0.1% formic acid in water then desalted using C18 ZipTips (Millipore, Billerica, MA). Eluates of the desalting columns were analyzed by HPLC-MS/MS on a NanoLC-Ultra 2D HPLC system coupled to an AB Sciex 5600 quadrupole-quadrupole-time-of-flight (Qq-TOF) mass spectrometer via a NanoSpray^{TM} III ion source (Concord, ON, Canada). Samples were loaded via an autosampler onto a C18 trap column then separated by nanoflow HPLC on a C18 reversed phase analytical column (75 µM I.D., 15 cm long, 3 µM particle size and 120 Å pore size) mounted in a cHiPLC-nanoflex system. The LC separation of GraR peptides was performed using a 45 minute binary mobile phase gradient from 5 to 35 % mobile phase B followed by a two minute wash at 80% mobile phase B then a 20 minute equilibration at 5% mobile phase B. The compositions of mobile phases A and B were 0.1% formic acid in water and 0.1% formic acid in acetonitrile respectively. Mass spectrometry was performed in positive ion mode. The data acquisition cycle consisted of a single MS scan followed by MS/MS scans of the ten most abundant analytes present in the MS. VraR peptides were analyzed over a 30 minute acetonitrile gradient by MRM on either an AS1-TempoLC-4000QTRAP with a NanoSpray^{TM} II ion source (AB Sciex) or on the Qq-TOF platform described above in MRM high resolution (MRMhr) mode. The 4000QTRAP platform was equipped with a 0.5 µm long 200 µm ID C18 trap (Michrom Bioresources, Auburn CA, USA) and an in-house packed 12 cm long, 75 µm ID column containing 3 µm Reprosil-Pur C18-AQ particles with 120 Å pore size (Dr. Maisch, Ammerbuch-Entringen, Germany).

GraR peptides and phosphopeptides were identified from the MS/MS data using ProteinPilot software (AB Sciex). The search conditions used for the sequence assignment were as follows: no cysteine modification, trypsin digestion, phosphorylation emphasis, biological modifications, amino acid substitutions, thorough ID and confidence threshold of 95% for protein detection. Spectra were searched against the entire UniProt protein sequence library (downloaded on September 8, 2012) (30).

2.3 Results

The mass spectrometry of proteolytic peptides containing the VraR D55 phosphosite and GraR phosphosites T128, T130 and T149 is described below. The dependencies of GraR threonine phosphorylation on Stk1 and VraR aspartate phosphorylation on acetyl phosphate were demonstrated with the use of negative controls.

GraR phosphorylation by Stk1

Coverage of the GraR sequence obtained by untargeted MS/MS exceeded 98% in all the samples analyzed. Two singly phosphorylated tryptic peptides from GraR were identified with 99% confidence in the sample treated with Stk1 in the presence of ATP. The sequences of the two phosphopeptides corresponded to amino acids 128 to 140 (TLTWQDAVVDLSK) and 146 to 154 (GDDTIFLSK) of the full length protein. No GraR phosphopeptides were identified from wild-type protein treated with ATP alone or from the triple mutant treated with Stk1 and ATP or from the untreated GraR control whereas two singly phosphorylated peptides were observed by LC-MS/MS with 99% confidence in replicate preparations of GraR incubated with Stk1 in the presence of ATP. The three threonine-to-alanine amino acid substitutions present in the triple mutant were identified with 99% or higher confidence in multiple MS/MS spectra obtained from the mutant sample.

Peptide sequences and locations of the phosphosites shown in Figures 2.1 and 2.2 were assigned with 99% confidence by the ProteinPilot software in a search against the entire UniProt sequence library. Multiple MS/MS spectra were identified with 99% confidence

as those of the phosphopeptides with unambiguous phosphosite assignment. The sequence library to which the MS/MS data were compared by the algorithm was the largest available, including the most complete proteomes of all species on record, yet the only phosphopeptides identified were those from the GraR protein of *S. aureus*. In this context, the likelihood that the algorithm randomly assigned these MS data to the molecular species predicted by other lines of experimental evidence is exceedingly low.

Representative spectra of the two tryptic GraR phosphopeptides, the non-phosphorylated peptides with the same amino acid sequences and the corresponding mutant peptides are shown in Figures 2.1 and 2.2 and include product ion series representative of their complete amino acid sequences. The tentative locations of the phosphoamino acids were assigned based on the presence of characteristic ions in the product ion series. In order to confirm that the phosphates resided on the residues shown to be phosphorylated in wild type GraR, the triple mutant containing alanine at these sites was exposed to the kinase and ATP and analyzed in parallel with the wild-type protein. Mutation of threonine 128, 130 and 149 abrogated GraR phosphorylation which indicated that phosphorylation was dependent on the presence of threonine at these sites independently of the serine residues within both peptides. No spectra obtained from any sample denoted phosphorylation at serine 139 or serine 153. The absence of phosphorylation within the mutant peptide GDDAIFLSK suggests that threonine 149, not serine 153, is the phosphorylation site. There was no evidence for the phosphorylation of serine 139 in wild-type or triple mutant GraR, indicating the phosphate resided at a threonine residue.

Phosphopeptides give characteristically low signal response in positive ion mode due to their acidity (31). Thus the low intensity of the spectra in Figure 2.1B, 2.1C, 2.1D and 2.2B are typical of such peptides. Accurate mass measurement of phosphopeptide product ions compensated for their low s/n to enable their identification and accounts for the high confidence sequence assignments (Table 2.1). Precursor and product ions were observed within hundredths or thousandths of a Dalton of their calculated m/z values. The high resolution Qq-TOF spectra also permitted low abundance product ion peaks to be

distinguished from electronic noise, since the former had width in the m/z domain while the latter had no measureable width.

Mass spectra assigned to non-phosphorylated, singly phosphorylated, doubly phosphorylated and mutant versions of the sequence TLTWQDAVVDLSK were differentiated by their unique precursor ion m/z, chromatographic retention times and their distinctive fragmentation. The MS and MS/MS spectra corresponding to each phosphopeptide were exclusive to the kinase treated samples, indicating they formed only in the presence of Stk1 and ATP.

In the case of the peptide TLTWQDAVVDLSK, phosphorylation occurred both at a single threonine residue (Figure 2.1B and D) and at both threonine residues (Figure 2.1C). Phosphorylation of T128 and T130 could not be distinguished by MS/MS since they were contained in a single peptide. Where a phosphopeptide contains multiple serine, threonine or tyrosine residues the phosphate may transfer from one of these amino acids to another within the mass spectrometer, due to unintentional ion activation, obscuring the initial location of the phosphate (32). Spectra were collected showing the wild-type peptide singly phosphorylated on either threonine so that the order of dual phosphorylation could not be determined, however phospho-T128 in the T130A mutant revealed this event to be independent of T130 (Figure 2.1D). Glutamine deamidation is a common and uncontrollable reaction in LC-MS experiments which causes an increase of peptide ion mass of 0.98 Da (33, 33). Since the peptide TLTWQDAVVDLSK has two phosphorylation states and a glutamine residue, four modified variants with unique masses are present, reducing the MS signal of any one variant. High mass accuracy, mass resolution and chromatographic resolution were therefore necessary to isolate and confidently identify GraR phosphopeptides.

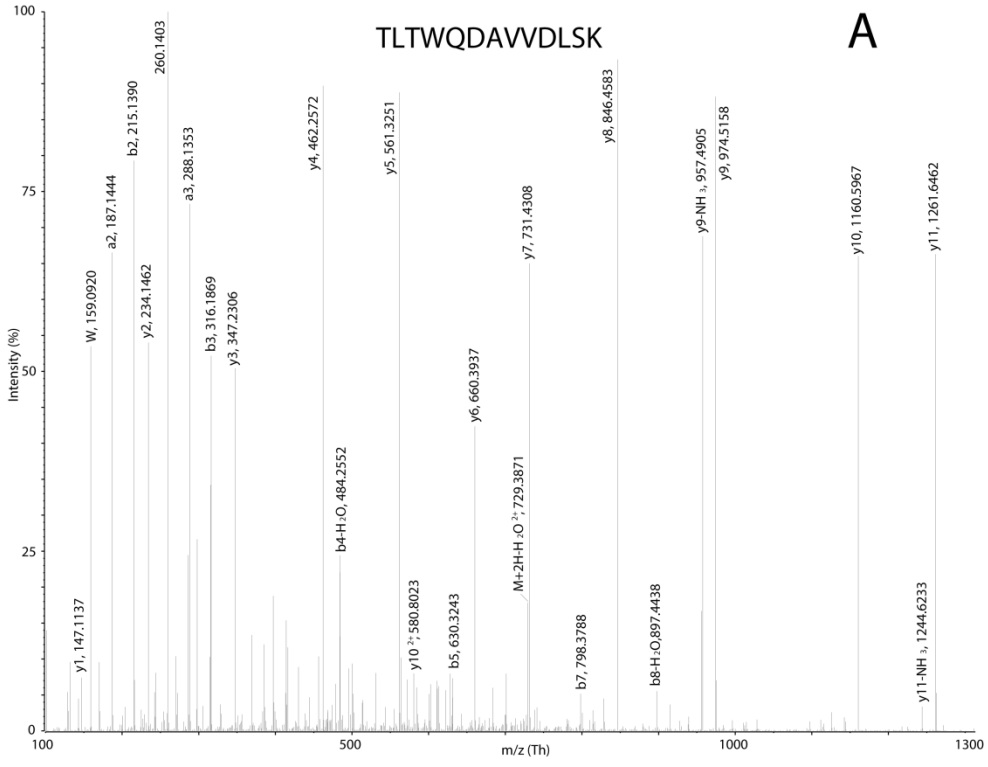
Loss of phosphate as a neutral species is a prominent and well documented feature of phosphopeptide CAD MS/MS. Phosphate loss results in a mass shift of 18.015 Da relative to the unmodified ion because H_2O is lost from the serine or threonine side chain along with the phosphate. The s/n of the y_{11}-H_3PO_4 peak in Figure 2.1B is 6.67:1 and

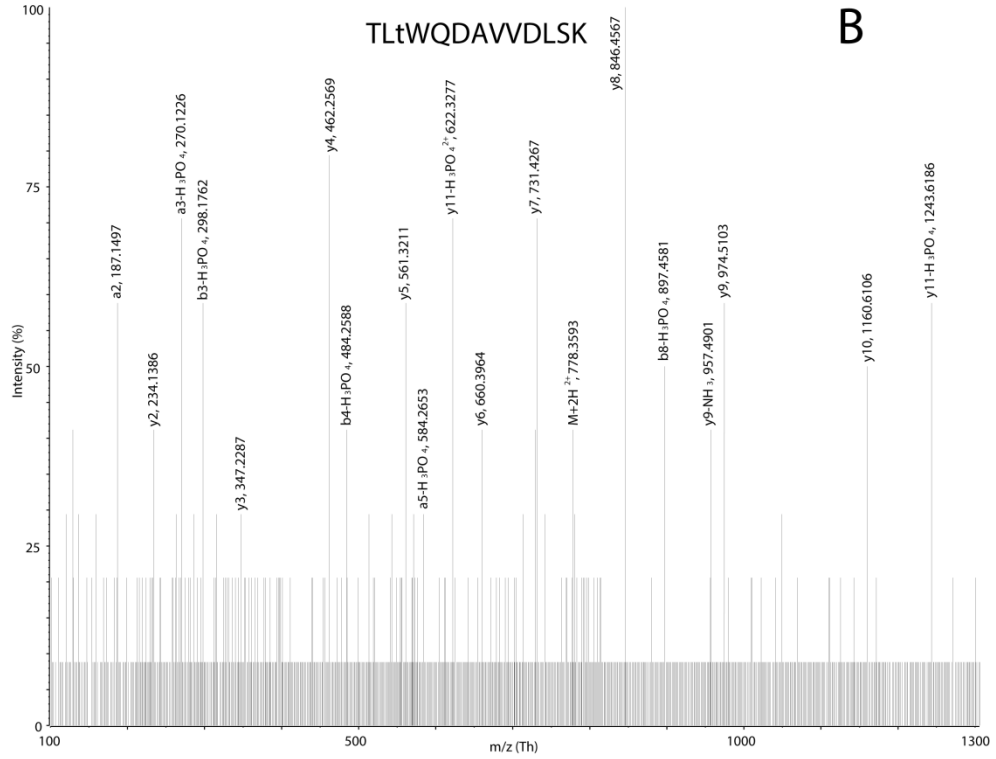
unlike the baseline noise, this peak had width characteristic of genuine ions. These peak characteristics are sufficient for MS/MS sequencing algorithms to differentiate them from noise so that the phosphopeptide identity was automatically assigned based on the y_{11}-H_3PO_4 ion among others. The doubly charged version of the y_{11}-H_3PO_4 was also identified in this spectrum with s/n of 8:1. In the same spectrum the a_3-H_3PO_4 and b_3-H_3PO_4 ions were identified with s/n of 8:1 and 6.7:1 respectively while b_4-H_3PO_4, b_8-H_3PO_4 and doubly charged y_{13} were also present above s/n of 5:1. Taken together, these product ions all support the assignment of the phosphosite to the second threonine from the N-termius of the peptide. The measured masses of the y_6 and y_7 ions in Figure 2.2A and the y_6-H_3PO_4 and y_7-H_3PO_4 ions in Figure 2.2B differ by 18.0124 Daltons and 18.020 respectively, indicating that water loss has occurred in both cases. Though non-phosphorylated amino acids undergo some water loss during CID of peptides, the major fragments of non-phosphorlyated peptides are typically y- and b-ions. Similarly, non-phosphorylated residues within phosphopeptides tend not to undergo water loss so that phosphopeptide fragments in which phosphosites are absent are mainly b- and y-types. In data from Stk1/ATP-treated GraR, both the absence of non-phosphorylated y_6 and y_7 ions and the presence of y_6-H_3PO_4 and y_7-H_3PO_4 implicate T149 as the phosphosite.

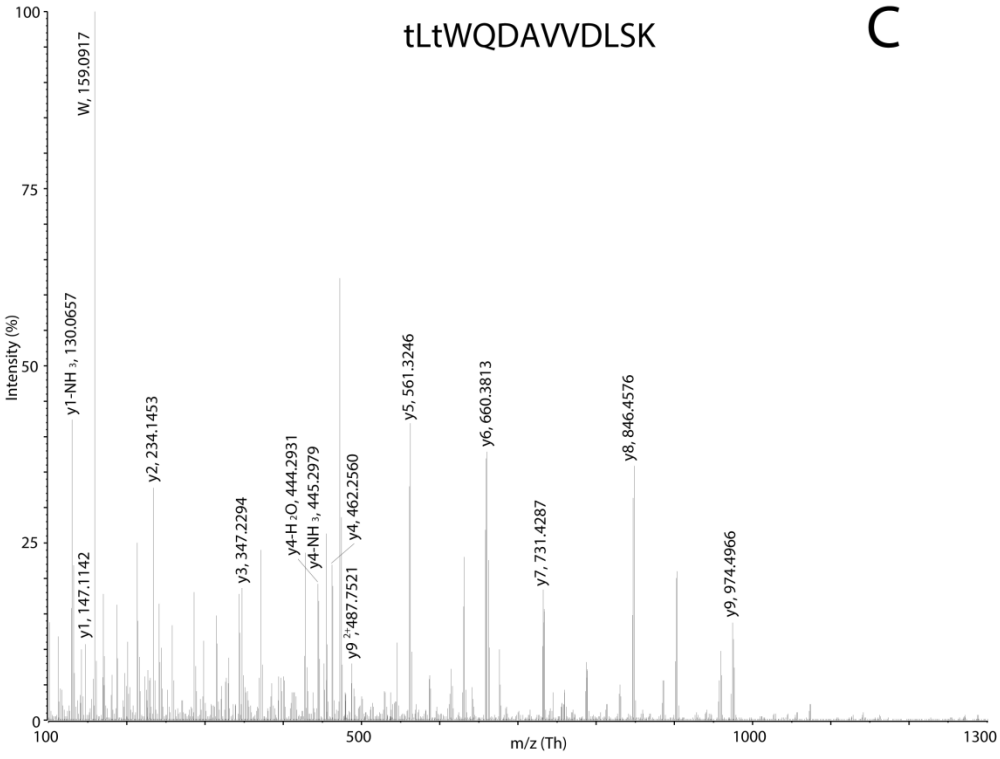
To increase instrument sensitivity for GraR phosphopeptides, MRMhr experiments were performed with the extracted mass width of each product ion maintained at 0.05 Th. Chromatograms of MRMhr transitions obtained from Stk1/ATP-treated GraR specific to doubly phosphorylated TLTWQDAVVDLSK, which have s/n between 1500:1 and 3745:1 are shown in Figure 2.3. The peptide was shown to be doubly phosphorylated based on its precursor m/z of 454.9 Th, while y_4, y_5, y_6 and y_9 ions were consistent with a non-phosphorylated serine residue. The iminium ion at m/z 159 Th was the most prominent fragmentation product of the doubly phosphorylated species, confirming the presence of tryptophan.

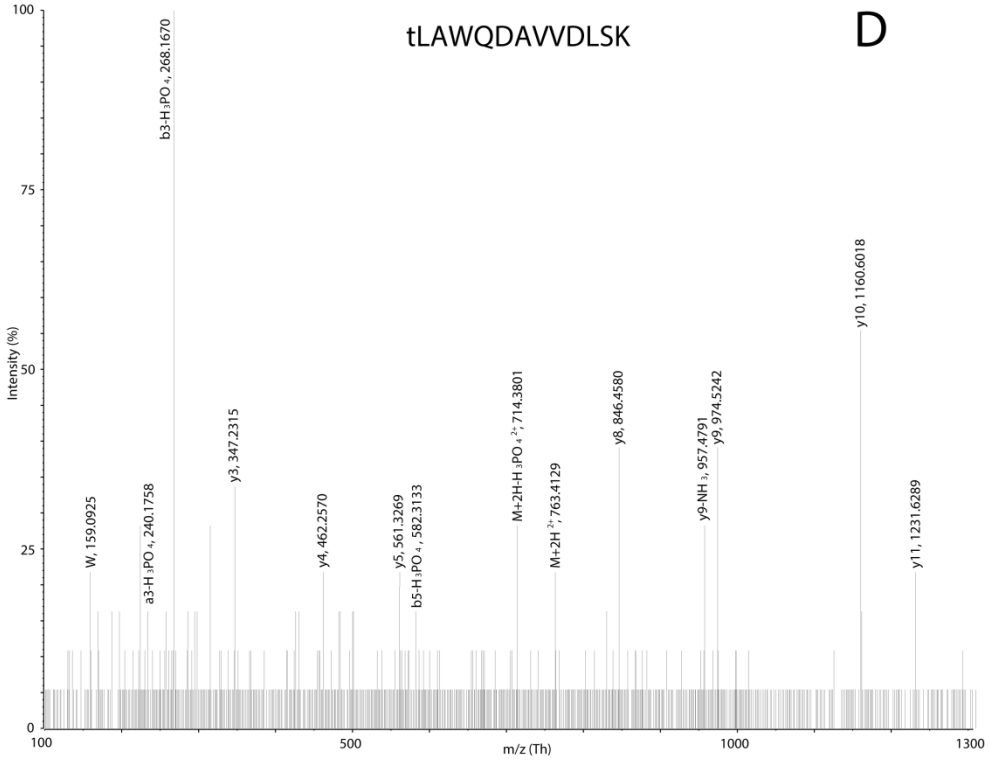
Figure 2.1: Phosphorylation of GraR at T128 and T130. Qq-TOF MS/MS spectra of tryptic GraR peptides containing the Stk1 phosphorylation sites and corresponding peptides from GraR point mutants in which threonine residues were replaced by alanine.

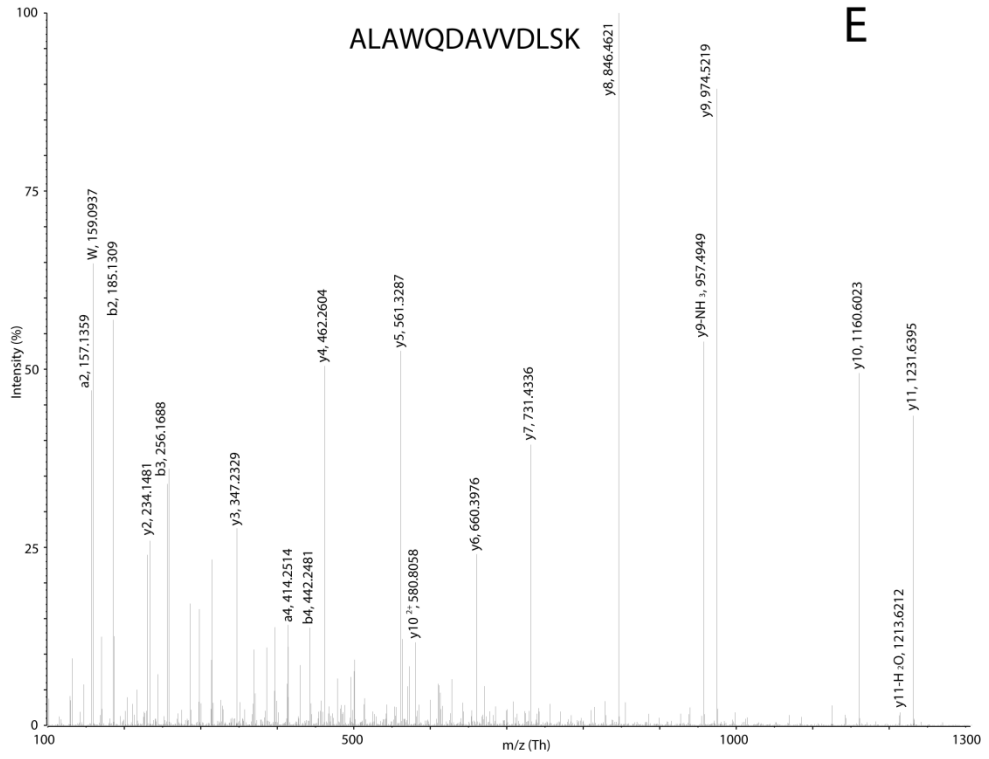
Sequence coverage is illustrated by labelling of fragment ions of each peptide according to Roepstorff–Fohlmann–Biemann nomenclature. Phosphoamino acids are indicated by lower case letters. Phosphate dissociation is indicated by the loss of phosphoric acid (H_3PO_4) from fragment ions which initially contained phosphorylated amino acid residues. The MS/MS spectra of (A) the peptide 128-TLTWQDAVVDLSK-140 from untreated GraR, (B) the phosphopeptide 128-TLtWQDAVVDLSK-140 from GraR incubated with ATP in the presence of Stk1, (C) the doubly phosphorylated peptide tLtWQDAVVDLSK from wild-type GraR incubated with ATP in the presence of Stk1, (D) the phosphopeptide 128-tLAWQDAVVDLSK-140 from the GraR-Thr130Ala mutant incubated with ATP in the presence of Stk1 and (E) the peptide 128-ALAWQDAVVDLSK-140 from the GraR triple mutant incubated with ATP in the presence of Stk1.




TLtWQDAVVDLSK
B
Intensity (%)
m/z (Th)
a2, 187.1497
y2, 234.1386
a3-H3PO4, 270.1226
b3-H3PO4, 298.1762
y3, 347.2287
y4, 462.2569
b4-H3PO4, 484.2588
y5, 561.3211
a5-H3PO4, 584.2653
y11-H3PO4 2+, 622.3277
y6, 660.3964
y7, 731.4267
M+2H 2+, 778.3593
y8, 846.4567
b8-H3PO4, 897.4581
y9-NH3, 957.4901
y9, 974.5103
y10, 1160.6106
y11-H3PO4, 1243.6186


tLtWQDAVVDLSK
C
Intensity (%)
m/z (Th)
W, 159.0917
y1-NH3, 130.0657
y1, 147.1142
y2, 234.1453
y3, 347.2294
y4-H2O, 444.2931
y4-NH3, 445.2979
y4, 462.2560
y9 2+, 487.7521
y5, 561.3246
y6, 660.3813
y7, 731.4287
y8, 846.4576
y9, 974.4966


tLAWQDAVVDLSK
D
Intensity (%)
m/z (Th)
b3-H3PO4, 268.1670
y3, 347.2315
W, 159.0925
a3-H3PO4, 240.1758
y4, 462.2570
y5, 561.3269
b5-H3PO4, 582.3133
M+2H-H3PO4 2+, 714.3801
M+2H 2+, 763.4129
y8, 846.4580
y9-NH3, 957.4791
y9, 974.5242
y10, 1160.6018
y11, 1231.6289


ALAWQDAVVDLSK
E
Intensity (%)
m/z (Th)
y8, 846.4621
y9, 974.5219
W, 159.0937
b2, 185.1309
a2, 157.1359
y4, 462.2604
y5, 561.3287
y9-NH3, 957.4949
y10, 1160.6023
y11, 1231.6395
y7, 731.4336
b3, 256.1688
y2, 234.1481
y3, 347.2329
y6, 660.3976
a4, 414.2514
b4, 442.2481
y10 2+, 580.8058
y11-H2O, 1213.6212

Attempts to observe endogenous phosphorylation were made using *S.aureus* RN6390 transformed with the pMK4 vector wild type GraR. All three phosphosites were evident, but only in their unmodified states, within two peptides by LC-MS/MS of a tryptic digest of the lysate. A gel band excised at the expected position of GraR gave the same result. No GraR peptides or phosphopeptides appeared in LC-MS/MS data from a MOAC eluate of the same digest in which many other *S. aureus* phosphopeptides were identified. Lack of MS evidence for *in vivo* PTM points to an inability in maintaining RR activation at a high level. Similar challenges were encountered in the case of VraR D55, a phosphosite of known biological significance.

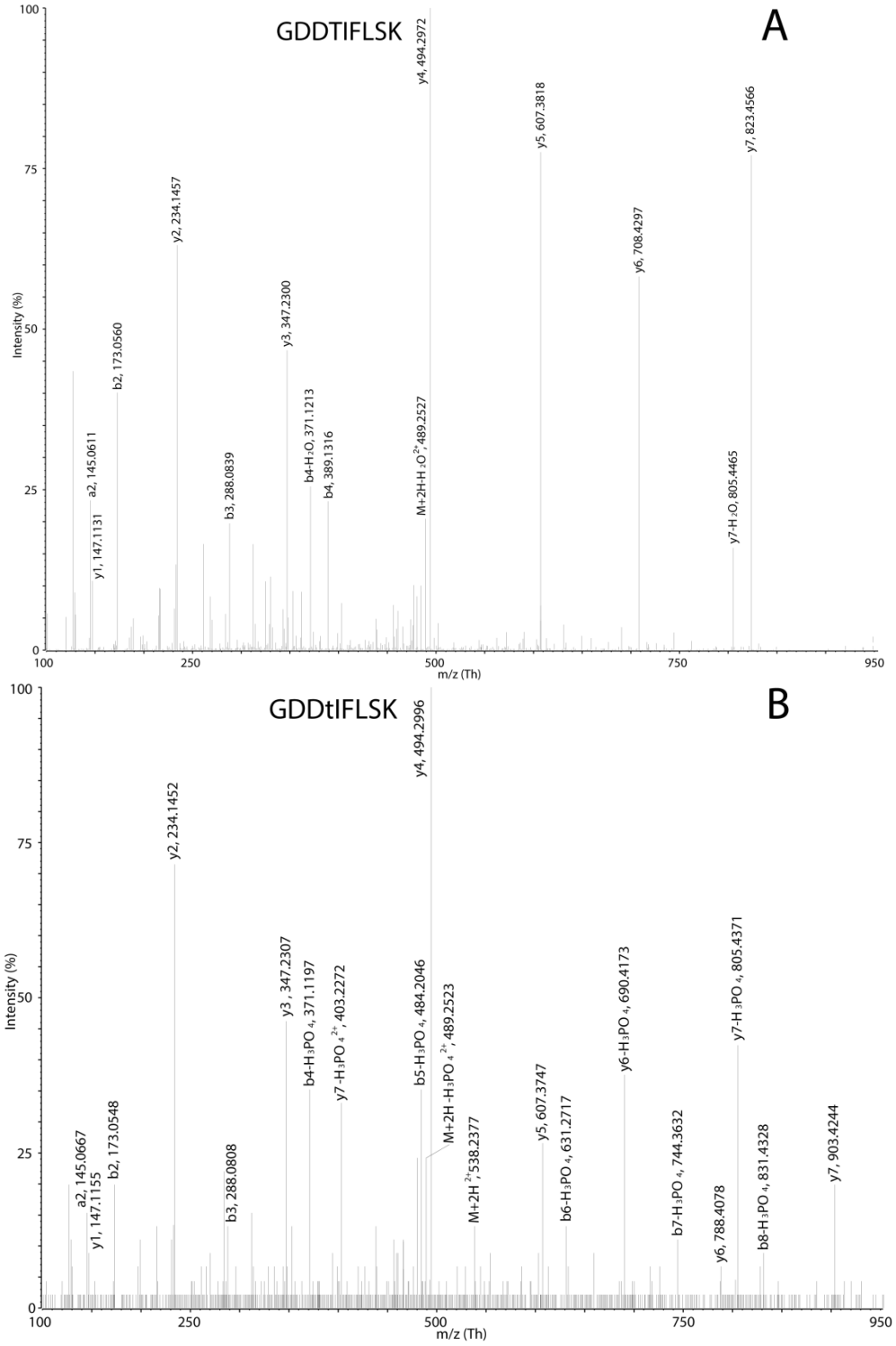
Biochemical tests and Nuclear Magnetic Resonance Spectroscopy suggest Stk1 phosphorylates GraR in vivo

In agreement with LC-MS/MS, a kinase assay, a phosphatase assay, a DNA binding assay and ^{1}H-NMR analysis of cell wall teichoic acid supported the hypothesis that Stk1 phosphorylates GraR *in vivo* in response to antibiotic stress (25).

Unlike BceR, heat-denatured GraR or VraR, GraR became radioactive after exposure to Stk1 and [γ-^{32}P]-ATP, demonstrating the interaction was substrate specific. Since *in vitro* phosphorylation assays combining substrate and kinase in relatively high concentrations are prone to false-positives, BceR and denatured GraR were important negative controls. Whereas VraR and GraR have low homology, BceR and GraR share 56% sequence identity including T128 and T130 (Figure 2.4). That GraR only bound phosphate in its native state showed the PTM to depend on conformation rather than sequence alone. Stk1-induced GraR phosphorylation was restricted to the C-terminal DNA binding domain since GraR and GraR^{C} were radiolabelled by Stk1 while the GraR^{N} construct was not. Despite high reagent concentrations used in the kinase assay, only an estimated 9% of total GraR was modified, an observation supported by the low abundance phosphopeptide signals relative to the corresponding peptide signals in MS. The addition of Stp1 following Stk1 treatment abrogated the phospho-GraR signal, suggesting a mechanism for direct negative modulation of the RR *in vivo*.

An electrophoretic mobility shift assay was used to determine the effect of GraR phosphorylation on its function as a transcription factor. Treatment with ATP and Stk1 increased the affinity of GraR for the vraRG promoter from *S. aureus* strain MU50 by 1.5-fold, despite the phosphorylation equilibrium favouring the unmodified protein. This result implies a relationship between the transcription factor action of GraR and its Stk1-depenendent phosphorylation state.

The biological role of the phosphosites was further investigated using *S. aureus* GraR knockout mutant strain RN6390ΔgraR, which grows normally but is twice as susceptible to vancomycin as RN6390. Alternate strains were produced by inserting the genes for wild-type GraR or the double mutant GraR-Thr130Ala/Thr149Ala or GraR-Thr128Ala/Thr130Ala/Thr149Ala into RN6390ΔgraR. Wall teichoic acids were isolated from cultures of each modified RN6390 strain then analyzed by 1D proton NMR on a Bruker AV III 700 MHz NMR spectrometer. The effect of each GraR variant on wall teichoic acid composition was assessed by the intensity ratios of peaks corresponding to the N-acetyl group of N-acetyl-glucosamine (2.10 ppm) and D-alanine (1.63 ppm). RN6390ΔgraR had substantially lower wall teichoic acid D-alanine content than RN6390. wall teichoic acid from RN6390ΔgraR complimented with the wild-type GraR gene was indistinguishable from that of RN6390 though complimenting the same knockout with the GraR double mutant or triple mutant reduced D-alanine content by approximately half. That NMR showed the alanine content of wall teichoic acid to depend on a GraR isoform having intact Stk1 phosphosites is indicative of a phosphorylation-induced GraR activating effect. The results presented above depict a serine/threonine kinase as a sensor complimentary to GraS. Since two-component systems confer tolerance to antibiotics, this work suggests Stk1 or the threonine phosphorylation sites of GraR are potential therapeutic targets in the management of multidrug resistant *S.aureus* infection.


GDDTIFLSK
A
y4, 494.2972
y5, 607.3818
y7, 823.4566
y2, 234.1457
y6, 708.4297
y3, 347.2300
b2, 173.0560
b4-H2O, 371.1213
b4, 389.1316
M+2H-H2O2+, 489.2527
b3, 288.0839
a2, 145.0611
y1, 147.1131
y7-H2O, 805.4465
Intensity (%)
m/z (Th)
GDDtIFLSK
B
y4, 494.2996
y2, 234.1452
y3 , 347.2307
b4-H3PO4, 371.1197
y7 -H3PO4 2+, 403.2272
b5-H3PO4, 484.2046
M+2H -H3PO4 2+, 489.2523
y6-H3PO4, 690.4173
y7-H3PO4, 805.4371
M+2H 2+, 538.2377
y5, 607.3747
b6-H3PO4, 631.2717
b7-H3PO4, 744.3632
b2, 173.0548
b8-H3PO4, 831.4328
a2, 145.0667
y7, 903.4244
y1, 147.1155
b3, 288.0808
y6, 788.4078
Intensity (%)
m/z (Th)

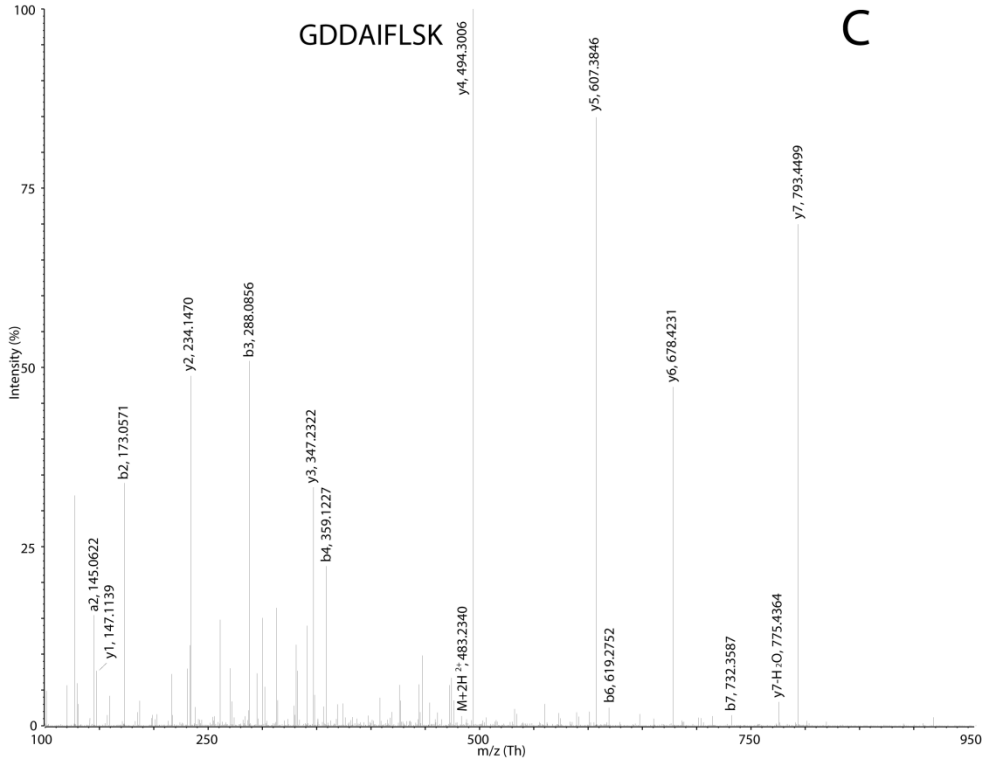


Figure 2.2: Phosphorylation of GraR at T149. Qq-TOF MS/MS spectra of (A) the tryptic peptide 146-GDDTIFLSK-154 from untreated GraR (B) the phosphopeptide 146-GDDtIFLSK-154 from GraR incubated with ATP in the presence of Stk1, (C) 146-GDDAIFLSK-154 from the GraR triple mutant incubated with ATP in the presence of Stk1.

Table 2.1: Mass accuracies of ions observed in MS/MS spectra of tryptic GraR peptides harbouring the three Stk1 phosphorylation sites and the corresponding alanine point mutants. Scans from which the listed masses were obtained are shown in the figure panels indicated.

Ion	Observed monoisotopic m/z (Th)	Error (Th)
TLTWQDAVVDLSK, 2+	Figure 2.1A	
$M+2H^{2+}$	738.3908	0.0003
y1	147.1137	0.0009
W	159.0920	0.0003
b2	215.1390	<0.0001
y2	234.1462	0.0014
a3	228.1353	-0.0565
b3	316.1869	0.0002
y3	347.2306	0.0017
y4	462.2572	0.0014
$b4-H_2O$	484.2552	-0.0002
y5	561.3251	0.0008
$y10^{2+}$	580.8023	0.0014
b5	630.3246	<0.0001
y6	660.3937	0.0010
$M+2H-H_2O^{2+}$	729.3871	0.0017
y7	731.4308	0.0010
b7	798.3788	0.0007
y8	846.4583	0.0016
$b8-H_2O$	897.4438	-0.0027
b8	915.4359	0.0212
$y9-NH_3$	957.4907	0.0017
y9	974.5158	0.0005
y10	1160.5967	0.0021
y11	1261.6462	0.0039
TLtWQDAVVDLSK, 2+	Figure 2.1B	
$M+2H^{2+}$	778.3737	-0.0002
a2	187.1497	0.0056
y2	234.1386	-0.0062
$a3-H_3PO_4$	270.1226	-0.0586
$b3-H_3PO_4$	298.1762	0.0001
y3	347.2287	-0.0002
$a4-H_3PO_4$	456.2486	0.0119
y4	462.2569	0.0011
$b4-H_3PO_4$	484.2588	0.0034
y5	561.3211	-0.0032
$a5-H_3PO_4$	584.2653	0.0538
$y11^{2+}-H_3PO_4$	622.3277	0.0082
y6	660.3964	0.0037
y7	731.4267	0.0031
y8	846.4567	<0.0001
$b8-H_3PO_4$	897.4581	0.0116
$y9-NH_3$	957.4901	-0.0013
y9	974.5103	-0.0050
y10	1160.6106	0.0160
$y11-H_3PO_4$	1243.6186	0.0131
tLtWQDAVVDLSK, 3+	Figure 2.1C	
$y1-NH_3$	130.0638	-0.0225
y1	147.1150	0.0022
W	159.0918	0.0001
y2	234.1453	0.0005
y3	347.2340	0.0051

y4-H_3PO_4	444.2952	0.0499
y4-NH_3	445.2989	0.0696
y4	462.2564	0.0006
$y9^{2+}$	487.2493	-0.5120
y5	561.3338	-0.0589
y6	660.3725	-0.0202
y7	731.4297	-0.0001
y8	846.4626	-0.0059
b7-H_3PO_4	958.4760	0.0193
y9	974.5057	-0.0096
ALAWQDAVVDLSK, 2+	Figure 2.1D	
a2	157.1359	0.0024
b2	185.1309	0.0024
y2	234.1481	0.0033
b3	256.1688	0.0032
y3	347.2329	0.0040
a4	414.2514	0.0014
b4	442.2481	0.0032
y4	462.2604	0.0046
y5	561.3287	0.0044
$y10^{2+}$	580.8058	0.0049
y6	660.3976	0.0049
$M+2H^{2+}$	708.3832	0.0031
y7	731.4336	0.0038
y8	846.4621	0.0054
y9-NH_3	957.4949	0.0061
y9	974.5219	0.0066
y10	1160.6023	0.0077
y11-H_2O	1213.6212	<0.0001
y11	1231.6395	0.0078
tLAWQDAVVDLSK, 2+	Figure 2.1E	
a3-H_3PO_4	240.1758	0.0051
b3-H_3PO_4	268.1670	0.0014
y3	347.2315	0.0017
y4	462.2570	0.0012
y5	561.3269	0.0026
b5-H_3PO_4	582.3133	0.0098
$M+2H-H_3PO_4^{2+}$	714.3801	<0.0001
$M+2H^{2+}$	763.3769	0.0084
y8	846.4580	0.0013
y9-NH_3	957.4791	-0.0097
y9	974.5242	0.0089
y10	1160.6018	0.0072
y11	1231.6289	-0.0028
GDDTIFLSK, 2+	Figure 2.2A	
a2	145.0611	0.0003
y1	147.1131	0.0003
b2	173.0560	0.0003
y2	234.1457	0.0009
b3	288.0839	0.0013
y3	347.2300	0.0012
b4-H_2O	371.1213	0.0016
b4	389.1316	0.0013
$M+2H-H_2O^{2+}$	489.2527	0.0021
y4	494.2972	-0.0001
$M+2H^{2+}$	498.2563	0.0008

y5	607.3818	0.0004
y6	708.4297	0.0006
y7-H_2O	805.4465	0.0011
y7	823.4566	0.0006
GDDtIFLSK, 2+	Figure 2.2B	
a2	145.0667	0.0059
y1	147.1155	0.0027
b2	173.0548	-0.0009
y2	234.1452	0.0004
b3	288.0808	0.0018
y3	347.2307	0.0018
b4-H_3PO_4	371.1197	<0.0001
y7-$H_3PO_4^{2+}$	403.2272	0.0008
b5-H_3PO_4	484.2046	0.0008
M+2H-$H_3PO_4^{2+}$	489.2523	0.0017
y4	494.2996	0.0023
M+$2H^{2+}$	538.2377	-0.0026
y5	607.3747	0.0067
b6-H_3PO_4	631.2717	-0.0005
y6-H_3PO_4	690.4173	-0.0012
b7-H_3PO_4	744.3632	0.0069
y6	788.4078	0.0124
y7-H_3PO_4	805.4371	-0.0083
b8-H_3PO_4	831.4328	0.0445
y7	903.4244	0.0021
GDDAIFLSK, 2+	Figure 2.2C	
a2	145.0622	0.0014
y1	*147.1139*	0.0011
b2	173.0571	0.0014
y2	234.1470	0.0022
b3	288.0856	0.0030
y3	347.2322	0.0033
b4	359.1227	0.0030
M+$2H^{2+}$	483.2340	-0.0166
y4	494.3006	0.0033
y5	607.3846	0.0032
b6	619.2752	0.0030
y6	678.4231	0.0046
b7	732.3587	0.0024
y7-H_2O	775.4364	0.0015
y7	793.4499	0.0045

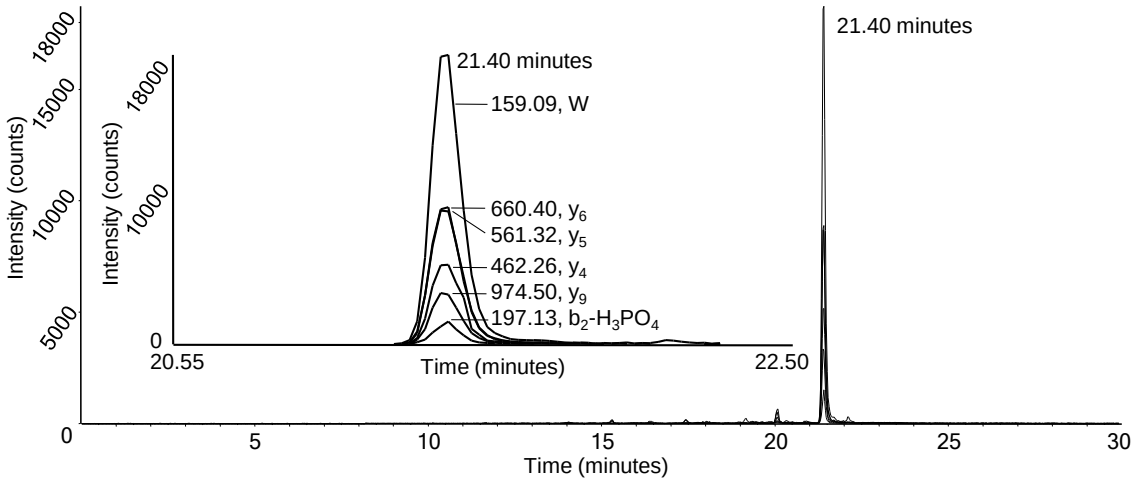


Figure 2.3: Overlaid chromatograms of six transitions specific to the doubly phosphorylated GraR peptide tLtWQDAVVDLSK from a tryptic digest of GraR after treatment with Stk1 and ATP. Data from the entire analytical gradient is shown to illustrate the specificity of the transitions, which appeared simultaneously in a single peak. All transitions shared the same precursor ion with m/z 545.9 Th, corresponding to the triply charged phosphopeptide ion. The identities of the product ions are shown in the inset which includes data from 20.5 to 22.5 minutes. The product ion mass window was ±0.025 Th for each transition.

```
                                  128 130                    149
                                   ↓   ↓                     ↓
GraR (gi: 81854576) 121-FTAEEKRTLTWQDAVVDLSKDSIQKGDDTIFLSKTEMIILEILITKKNQIVSRDTIITAL
                        +  E     TW  A VD  ++ +     ++ L+K EM IL+ LI +KN+IVSR+ +I +L
BceR (gi: 41688440) 122-YNTEPSTIKTWCGAAVDAEQNLVSNDKGSVELTKNEMFILKQLIEQKNKIVSREELIRSL
                                   ↑   ↑
```

Figure 2.4: Sequence alignment of GraR and BceR, its homologue in B. Subtilis, in the region containing the Stk1 interaction site. The Stk1 phosphorylation sites on GraR and their equivalents on BceR are indicated by arrows.

VraR phosphorylation by acetyl phosphate

The region flanking VraR D55 has few Lys and Arg residues, resulting in a fully tryptic peptide 28 amino acids long (AHELKPDLILMDLLMEDMDGVEATTQIK), the C-terminus of which is next to a Lys and therefore potentially forms ragged ends. The fully

tryptic peptide eluted late in the LC gradient as a narrow, well resolved peak; however, it was highly retained on the chromatographic system and carried over in considerable amounts into successive runs following the sample. Glu-C digestion produced a group of shorter peptides containing D55 having equally well resolved but earlier chromatographic peaks than their tryptic counterpart. Glu-C gave the fully cleaved species LILMD (Figure 2.5) as well as the partially cleaved LILMDLLME, LKPDLILMD and LKPDLILMDLLME (Figure 2.6).

Exposure to $NaBH_4$ did not alter the mass of D55 from purified VraR, however a 14 Da mass shift was observed after the protein was treated with acetyl phosphate then reduced (Figures 2.5 and 2.6). The mass shift corresponded to the replacement of Asp with Hse, a structural isomer of threonine. The masses of small neutral species commonly lost from peptide ions differ significantly from the change induced by Asp to Hse conversion. Water and ammonia loss alter the precursor ion mass by 18 and 17 Da, respectively, and therefore phospho-Asp reduction resulted in a distinctive mass shift, one discernable even on a relatively low resolution mass spectrometer such as a triple-quadrupole instrument. Product ion spectra of each Glu-C cleavage variant confirmed D55 to be the reduced residue, indicating that it was phosphorylated by acetyl phosphate (Figure 2.7). As evidenced by MS and MS/MS data from peptides having missed cleavage sites, neither D50 nor E59 was affected by $NaBH_4$ (Figures 2.6, 2.7 and 2.8 and Table 2.2). A representative product ion spectrum from LILM-Hse-LLME depicted in Figure 2.7 includes the entire b-ion series and all but one y-ion with b_5 to b_8 and y_5 to y_7 having m/z values reflective of the loss of one oxygen along with the gain of two hydrogen atoms. The b_2 to b_4 and y_1 to y_4 ions in contrast had m/z values identical to those from the unmodified peptide ion. The peptide LKPDLILM-Hse-LLME also dissociated extensively so that all but one of its y-ions and every b-ion save one were detected (Figure 2.8, Table 2.2). From these findings it can be concluded that both the phosphorylation and the reduction were highly site-specific. Since acetyl phosphate is a small molecule which can presumably access much of the protein sequence, it follows that VraR actively participates in its own phosphorylation at D55. Reduced D55

remained a Glu-C cleavage site, though the prominence of peptides with missed cleavages suggested Glu-C efficiency may be compromised by removal of the aspartate carboxyl-group.

LC elution profiles of the peptides and analogous reduced peptides were consistent and their similar chromatographic behaviour is reasonable since threonine and aspartic acid differ in their indices of hydropathy by 2.8 on a scale 9 units in magnitude (34). Coelution could be advantageous in the quantification of activated VraR since this assures uniformity of LC-MS matrix effects.

Phosphorylated-reduced VraR peptide ions were consistently lower in abundance than the unmodified versions. This finding was consistent with the observation that acetyl phosphate yields phospho-VraR hundreds of times more slowly than does phospho-VraS at the concentration used here (17). The fully cleaved peptides LILMD and LILM-Hse differed in abundance by three orders of magnitude but both fragmented extensively to give b- and y-ions which had consistent relative abundances between the reduced and unmodified species (Figure 2.5). Despite their limited size and low basicity relative to tryptic peptides, these species formed mainly doubly charged ions.

Both LILMDLLME and LKPDLILMDLLME formed doubly protonated ions yielding extensive sequence information upon CAD in their unmodified and reduced forms. The peptide LKPDLILMDLLME at m/z 772.43 Th eluted less than a minute earlier than LILMDLLME at m/z 545.80 Th as did their respective reduced versions (Figure 2.6). MS signals of LILM-Hse-LLME and LKPDLILM-Hse-LLME prepared from the purified recombinant were approximately 10% those of unmodified forms remaining after treatment but reached 50 to 70% in *E.coli* lysate. The chromatographic properties of the LILMDLLME and LKPDLILMDLLME ions were comparable in the presence and absence of soluble *E.coli* protein however their signals decreased by an order of magnitude in the lysate which may account for the apparent increase in phosphorylation efficiency. While purified VraR samples contained single analytes with m/z values corresponding to those of the peptides of interest, many isobaric species were present in

cell lysates though with markedly dissimilar retention times and product ion spectra (Figure 2.6 C and D).

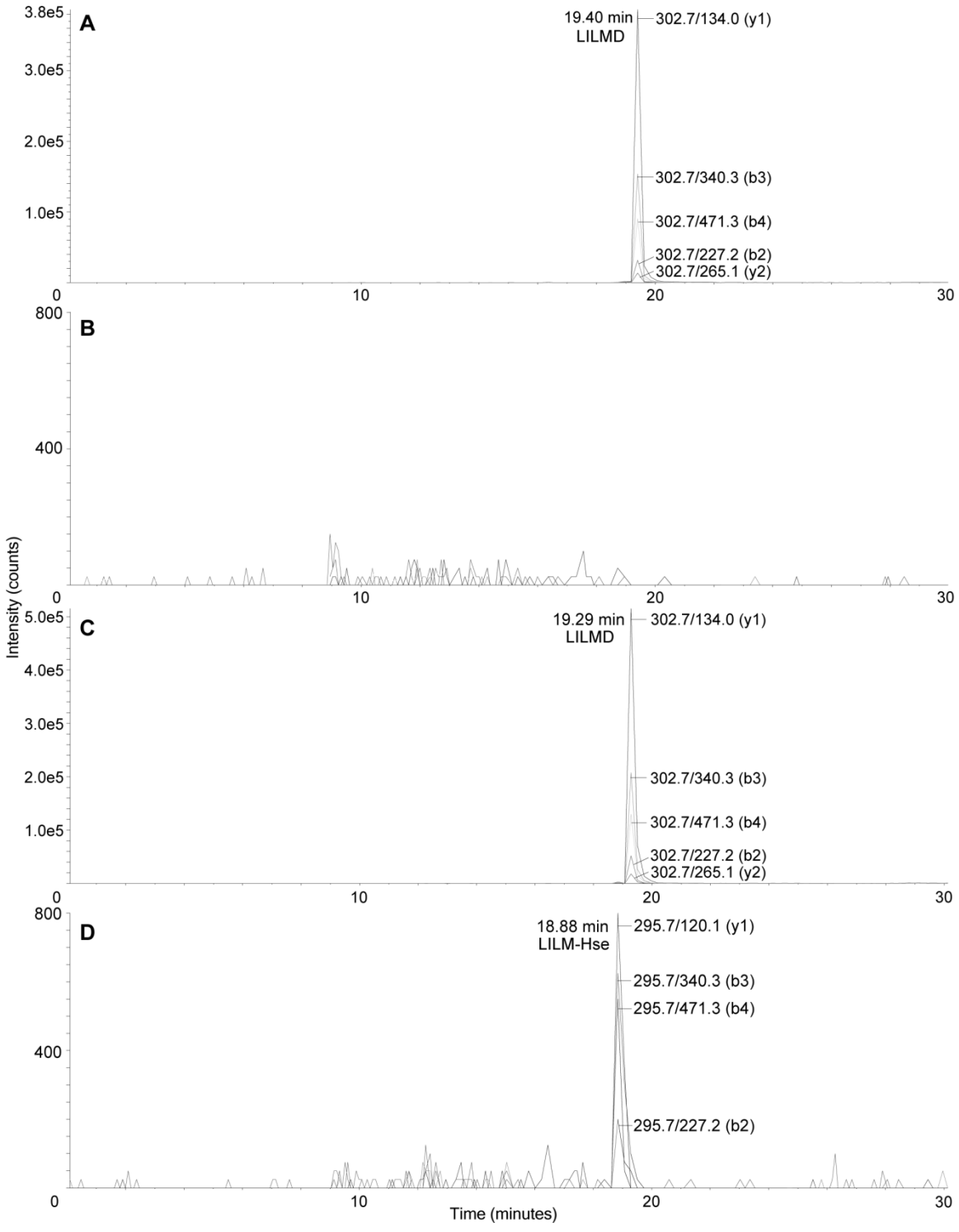


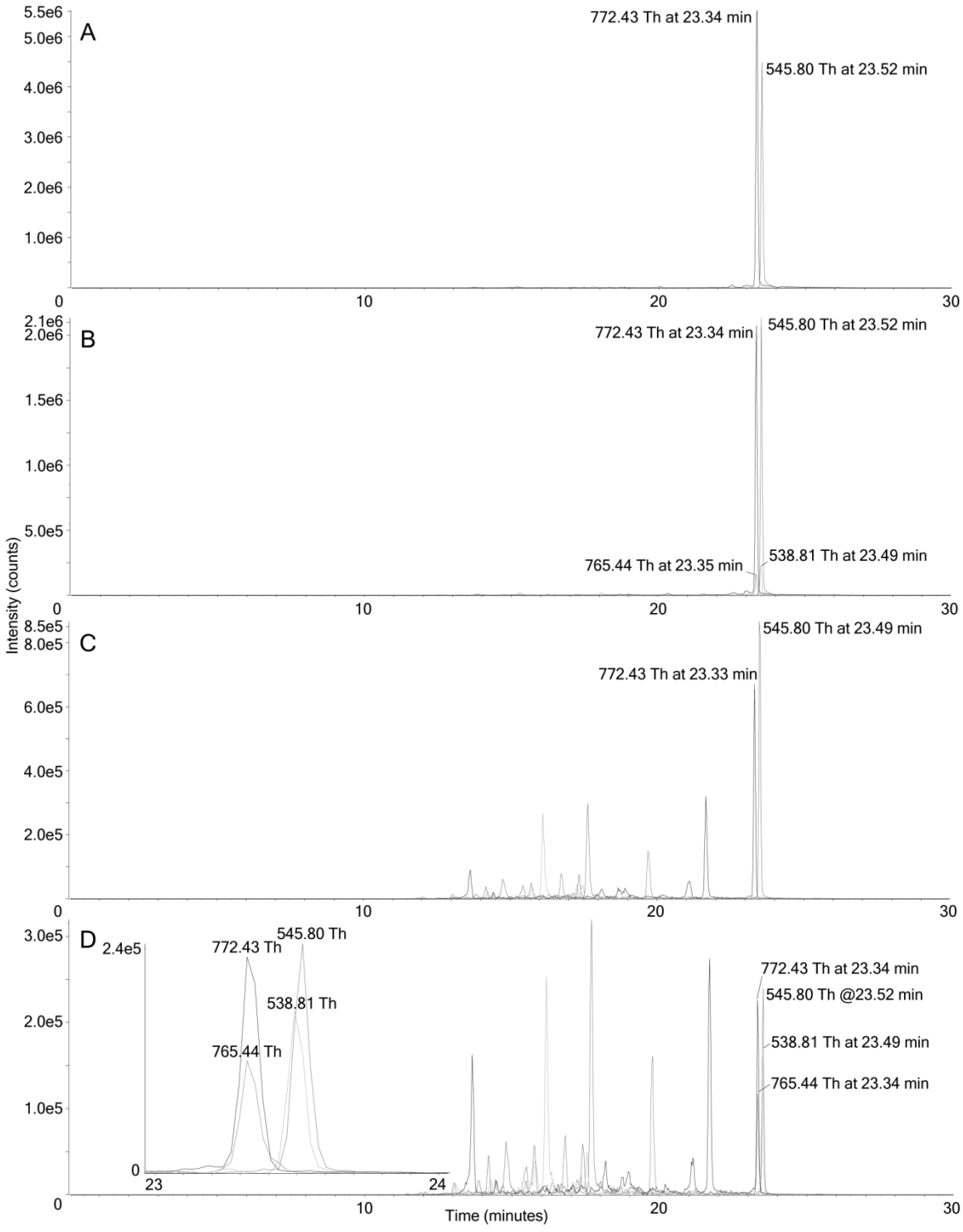
Figure 2.5: Chromatograms of MRM transitions corresponding to the fully cleaved Glu-C peptide LILMD in unmodified and reduced forms from purified recombinant VraR. Peaks from each transition are labelled according to their precursor/product ion m/z. Signals for (A) LILMD from $NaBH_4$-treated protein, (B) LILM-Hse from $NaBH_4$-treated protein, (C) LILMD from acetyl phosphate phosphorylated $NaBH_4$-treated protein (D) LILM-Hse from acetyl phosphate phosphorylated $NaBH_4$-treated protein. In order to eliminate potential carryover, three blank runs were performed between each sample run, which reduced the transition signals to baseline intensity. A single acquisition method was used to run the blanks and samples.

Sample complexity obscures VraR D55 phosphorylation state

Hse55-containing peptides were detected in crude lysates of VraR-overexpressing *E.coli* subjected to acetyl phosphate/$NaBH_4$ treatment (Figure 2.6D and 2.8), demonstrating that conditions under which VraR was modified were largely unaffected by the presence of other soluble proteins, however recombinant VraR represented a large fraction of total protein in this sample. Lower limits of detection for both D55 and Hse55 were established by spiking *S. aureus* lysate with phospho-VraR. Signals for both LKPDLILMD (529.3 Th, doubly charged) and LKPDLILM-Hse (522.3 Th, doubly charged) were observed only at the highest phospho-VraR concentration used, an amount of lysate equivalent to 2 µg of recombinant VraR in a total protein background of 1 mg. D55 alone was observed in the sample spiked with ten times less phospho-VraR, or 0.2 µg in 1 mg of lysate protein. Based on these finding, the limit of detection for Hse55 was estimated to be 1 µg per 1 mg lysate protein based on the assumption that half the total amount of VraR added was phosphorylated (17) (data not shown). Only SDS-PAGE was used to purify the protein prior to digestion and the VraR band was not discernable by Coomassie blue staining against a background of other proteins. Inclusion of matrix proteins, rather than VraR concentration, therefore increased the limit of detection beyond biologically relevant concentrations.

Hse55 was not observed in *S.aureus* MU50 cultured in the presence of oxacillin. Low intracellular VraR concentration was confirmed by SDS-PAGE and resulted in low signals for the unmodified peptides of interest. The apparent VraR concentration of the immunoprecipitate was low and the expected modification was not detected. It is feasible that the antibiotic exposure period used was excessive since phospho-VraR is dephosphorylated in less than five minutes following phosphorylation by VraS (17). It may be possible to follow the kinetics with a time-course study of MU50 oxacillin or vancomycin exposure.

Figure 2.6: Qq-TOF-MRMhr spectra of Glu-C cleaved D55-containing peptides from purified recombinant VraR and *E.coli* lysates. All samples were $NaBH_4$ reduced, digested with Glu-C then collected by zip tip prior to LC-MS/MS. The data was processed with the XIC Manager function of PeakView™ using a mass width of 0.02 Da over the entire data acquisition period. (A) Untreated recombinant VraR, (B) acetyl phosphate treated VraR, (C) untreated lysate from *E.coli* overexpressing VraR, (D) acetyl phosphate treated lysate from *E.coli* overexpressing VraR. Inset of panel D is the region of the chromatogram between 23 and 24 minutes by which the reduced peptides can be distinguished from their co-eluting unmodified variants.


A
772.43 Th at 23.34 min
545.80 Th at 23.52 min
B
772.43 Th at 23.34 min
545.80 Th at 23.52 min
765.44 Th at 23.35 min
538.81 Th at 23.49 min
C
545.80 Th at 23.49 min
772.43 Th at 23.33 min
D
2.4e5
772.43 Th
545.80 Th
538.81 Th
765.44 Th
772.43 Th at 23.34 min
545.80 Th @23.52 min
538.81 Th at 23.49 min
765.44 Th at 23.34 min
Intensity (counts)
Time (minutes)

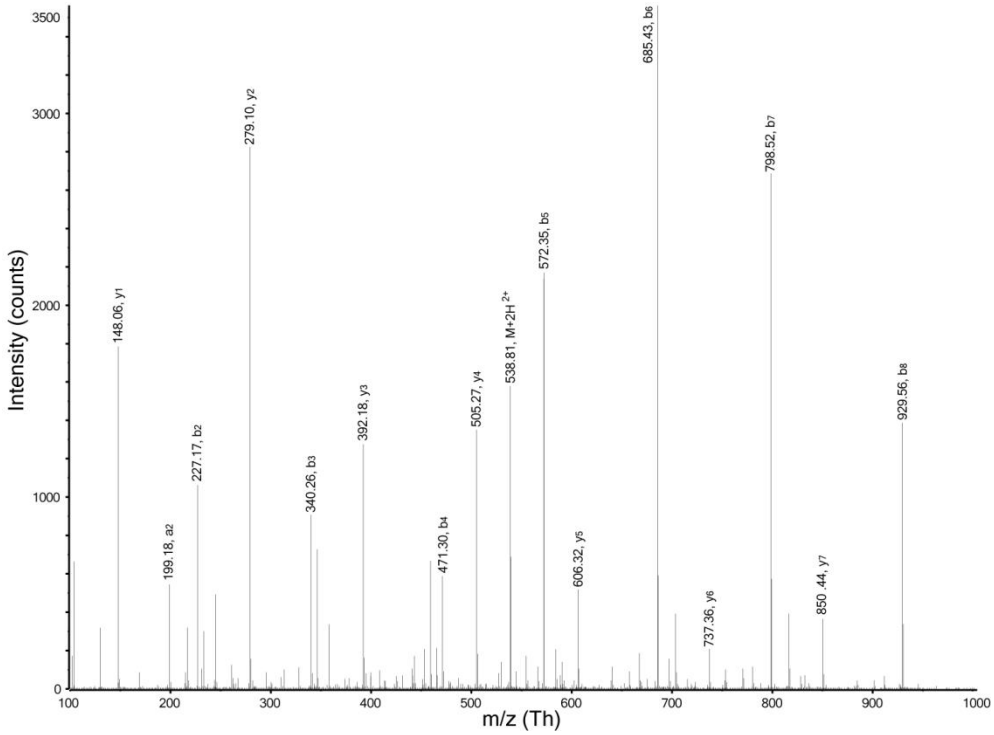


Figure 2.7: MS/MS spectrum of the doubly protonated peptide ion LILM-Hse-LLME (m/z 538.8 Th) from a Glu-C digest of acetyl phosphate-treated NaBH4-reduced purified recombinant VraR. This spectrum was collected 23.512 minutes into the LC-MS run shown in Figure 2.6B.

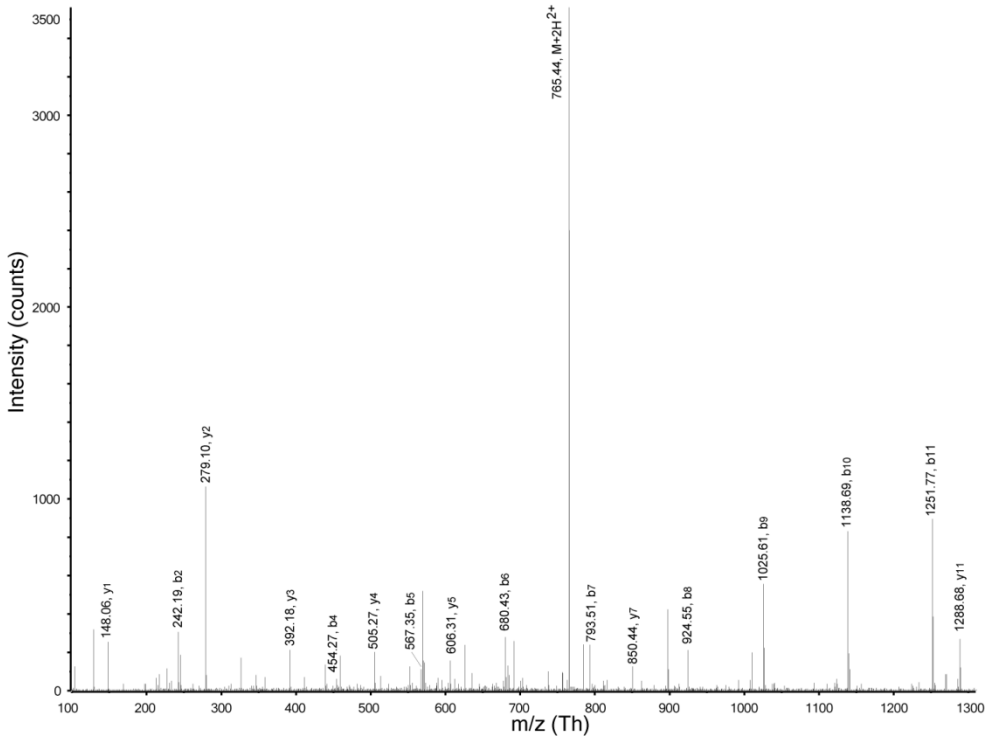


Figure 2.8: MS/MS spectrum of the doubly protonated peptide ion LKPDLILM-Hse-LLME (m/z 765.4 Th) from a Glu-C digest of lysate from E. coli expressing VraR after incubation with acetyl phosphate then NaBH4. This spectrum was collected 23.338 minutes into the LC-MS run shown in Figure 2.6D.

Table 2.2: Mass accuracies of ions observed in MS/MS spectra of Glu-C VraR peptides enclosing reduced D55.

Ion	Observed monoisotopic m/z (Th)	Error (Th)
LILM-Hse-LLME, 2+		
y1	148.0620	0.0016
b2	227.1778	0.0024
y2	279.1033	0.0024
b3	340.2620	0.0025
y3	392.1877	0.0027
b4	471.3025	0.0025
y4	505.2725	0.0035
$M+2H^{2+}$	538.8119	0.0036
b5	572.3512	0.0036
y5	606.3213	0.0046
b6	685.4363	0.0046
y6	737.3607	0.0035
b7	798.5158	0.0000
y7	850.4441	0.0028
b8	929.5632	0.0070
LKPDLILM-Hse-LLME, 2+		
y1	148.0592	-0.0012
b2	242.1865	0.0002
y2	279.1015	0.0006
y3	392.1844	-0.0006
b4	454.2678	0.0018
y4	505.2675	-0.0015
b5	567.3532	0.0031
y5	606.3143	-0.0024
b6	680.4336	-0.0005
y6	737.3584	0.0012
$M+2H^{2+}$	765.4372	-0.0005
b7	793.5116	-0.0066
y7	850.4422	0.0009
b8	924.5501	-0.0086
y8	963.5248	-0.0005
b9	1025.6072	0.0008
y9	1076.6061	-0.0033
b10	1138.6905	0.0001
b11	1251.7744	-0.0001
y11	1288.6826	-0.0065
b12	1382.8101	-0.0049
y12	1416.7818	-0.0023

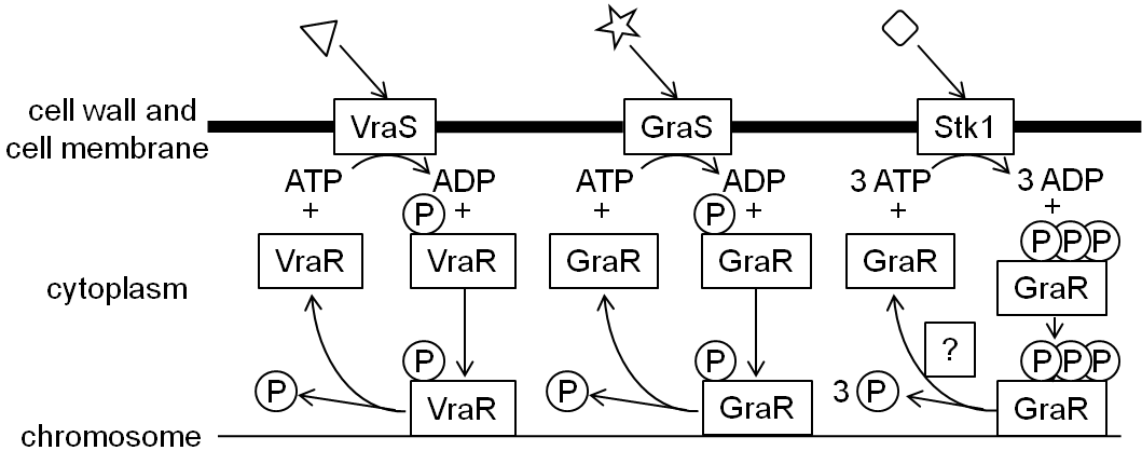


Figure 2.9: Proposed antibiotic response mechanisms involving VraSR, GraSR and Stk1/GraR. The unique extracellular ligands of each cell surface receptor are represented by different shapes. Each receptor is autophosphorylated upon binding its extracellular ligand then phosphorylates its cognate response regulator. VraS and GraS phosphorylate residues D55 and D51 of VraR and GraR respectively. Stk1 phosphorylates GraR T128, T130 and T149. Modification of the response regulators enhances their affinity for their respective promoters, leading to DNA binding and subsequent transcription initiation. Dephosphorylation causes detachment of the response regulators from the chromosome. GraR and VraR aspartate dephosphorylation can occur in the absence of a separate phosphatase whereas an unknown phosphatase, possibly Stp1, may act on GraR phosphothreonine residues.

2.4 Conclusion

In the above study, LC-MS/MS allowed the determination of phosphorylation sites within two *S. aureus* RRs known to mediate antibiotic resistance, namely VraR and GraR. A cell lysate-compatible method for the study of VraR activation showed that VraR D55 is a phosphate acceptor, in agreement with indirect evidence previously reported. Isolation of the region containing D55 required compromise between a tryptic peptide with unfavourable RPLC retention and ineffective Glu-C digestion. Endogenous VraR concentration remains the most significant barrier to its characterization.

The Stk1 and GraS phosphorylation sites are notably at the C-terminal DNA-binding domain and N-terminal TCS regulatory domain of GraR respectively. The two GraR phosphodonors therefore act on different regions of the protein, unlike VraR which has two possible *in vivo* phosphodonors for the same site. This suggests differential regulation, wherein the two sensors induce alternate active forms of the RR. Threonine residues 128, 130 and 149 are distributed throughout a sequence rich in potential phosphate acceptors so that site-directed mutagenesis was required to discern them as Stk1 targets.

Novel therapeutics are needed for the treatment of infection by multidrug resistant pathogens. Elucidation of the systems by which these organisms survive current treatments may inform the design of compounds exploiting or circumventing them. The complete description of VraSR and GraSR signalling, which initiates much of the *S. aureus* adaptive response, represents a step toward the management of this clinical challenge.

2.5 Contributors

All cloning, recombinant protein expression, recombinant protein purification, SDS-PAGE, *in vitro* protein phosphorylation and wall teichoic acid isolation was performed by Michael Fridman (M.Sc.). The reduction protocol was optimized by this author in collaboration with Michael Fridman. Howard Hunter (Ph.D.) performed ^{1}H-NMR analysis and interpreted the NMR results. This author performed all sample preparation beyond SDS-PAGE including $NaBH_4$ reduction and proteolytic digestion in the work presented above, as well as the acquisition, processing and interpretation of all LC-MS data. The VraR immunoprecipitation experiment was carried out entirely by this author.

Chapter 3

Phosphorylation of *Zea mays* Starch branching enzyme IIb by multiple calcium-dependent kinases of the amyloplast

3.1 Introduction

Starch is a mixture of highly-branched, high molecular weight polysaccharides synthesized by green algae and plants as an energy store. The notable hydrophobicity and density of this macromolecular assembly have stabilizing effects advantageous to their long-term intracellular sequestration. Seed endosperms contain the starch used by developing plant embryos in early growth and are a key source of dietary carbohydrates for livestock and humans. In addition to its use as a foodstuff, starch has myriad applications as an industrial feedstock. For these reasons it is a globally traded commodity. Corn, *Zea mays*, accounts for the majority of global cereal production and is therefore the most widely available source of starch.

Amylopectin and amylose, which are both composed exclusively of glucose monomers, are the two polysaccharide components of starches. Amylopectin, the more prevalent of the two, contains both α-(1→4)-O bonds and α-(1→6)-O bonds whereas amylose has only α-(1→4)-O-linkages. Chains of α-(1→4)-O bound glucose make up amylopectin strands, with α-(1→6)-O-linkages forming branch points between them (1). The length of the core chain as well as the length and number of branches vary, reflecting the behaviour of the enzyme complex responsible for their biosynthesis. Physicochemical properties of amylopectin depend on its structure, with the pattern of branching contributing more than chain length (2).

At the macroscopic level, starch is assembled into granules whose conformation and chemistry have ramifications for the survival of the plant as well as its utility to humans. These factors are affected by amylopectin structure, for example a wheat mutant lacking

the starch granule protein (SGP-1) involved in amylopectin synthesis makes deformed granules and rice mutants with similar defects produce starch of low digestibility (3, 4).

Developing endosperms generate starch for long term storage in specialized organelles called amyloplasts. Proteins make up around 1% of the mass of the starch granule, many of them mediating its formation (1). The starch synthesis machinery includes multiple gene products which differ among tissues of the same plant, presumably due to their unique energy requirements. In maize, four soluble starch synthases, (SSI to IV) are responsible for the production of amylopectin strands, each SS isoform producing α-glucan (polyglucose) chains of different length. The related granule-bound starch synthase I (GBSSI) is thought to produce amylose of varying chain length and is strongly associated with the granule (5). The Starch branching enzyme (SBE) family of proteins add glucose residues to the α-glucan via hydrolytic cleavage of internal α-(1→4) bonds and reaction of the freed reducing ends with the C6 hydroxyl group of glucose (6). Debranching enzymes, which are required for normal granule formation, are also present (7). Two SBE classes are known in maize, termed SbeI and SbeII, and these differ in the lengths of the glucan chains they transfer, the size of substrate they modify and their sites of action. SbeI has been shown to possess high amylose affinity, while SbeII displays a preference for amylopectin to which it transfers shorter α-glucan chains (8). The two maize SbeII subtypes are known as SbeIIa and SbeIIb and they differ in their tissue distribution and catalytic behaviour (9, 10). As the most abundant protein in the maize amylopast stroma, SbeIIb is expressed in at concentrations many times higher than SbeIIa throughout the endosperm (11, 12).

Amylopectin synthesising proteins work in multi-subunit complexes. In corn SSI, SSIIa and SbeIIb have been shown to congregate (13, 14) . The components of this trimeric network eventually become entangled within the starch granule through the glucan-binding capacity of SSIIa (15). SbeIIb also appears to form a protein complex with SbeI and starch phosphorylase (SP) (16). These complexes are dynamic in that the loss of one member through mutation can be compensated by other members so that while their final product is altered, the plant remains viable. For example, SbeIIa associates with many

SbeIIb binding partners and some additional proteins in the absence of functional SbeIIb and similarly SbeI replaces its interaction with SbeIIb in the knockout mutant, instead binding SSI along with SSIIa, SbeIIa and SP (Table 3.1). That such protein-protein associations can vary dramatically, arise from mutations of a single gene and are found in nature suggests a powerful mechanism of evolutionary adaptation.

Protein-protein interaction within the starch-synthesis machinery is dependent on its phosphorylation but no phosphorylation sites within starch synthesis proteins have been characterized prior to the study reported here (17). SbeIIb was chosen for phosphosite analysis since it is known to be phosphorylated and phenotypes arising from its mutation have been extensively characterized.

Genotype/Bait protein	no treatment	phosphatase	ATP	phosphatase+ATP
wild type				
SSI	SSI, SSIIa, SbeIIb	NA	NA	NA
SSIIa	SSI, SSIIa, SbeIIb	SSIIa, SbeIIb	SSI, SSIIa, SbeIIb	SSI, SSIIa, SbeIIb
SbeI	SbeI, SbeIIb	NA	NA	NA
SbeIIb	SSI, SSIIa SbeI, SbeIIb	SbeIIb	SSI, SSIIa SbeI, SbeIIb	NA
ae^{-}1.1				
SSI	SSI, SSIIa, SbeI, SbeIIa, SP	NA	NA	NA
SSIIa	SSI, SSIIa, SbeI, SbeIIa, SP	SSI, SSIIa, SbeI	SSI, SSIIa, SbeI	NA
SbeI	SSI, SSIIa SbeI, SbeIIa, SP	NA	NA	NA

Table 3.1: Co-immunoprecipitation of starch synthesizing proteins from functional endosperm amyloplast lysates of wild type and SbeIIb deficient maize. Bait protein refers to the antigen of the antibody used. Amyloplasts were untreated or incubated with 1 mM ATP or *E. coli* alkaline phosphatase for 30 minutes at 25^{o}C prior to immobilization of the bait proteins indicated. The presence of each interacting protein was probed by western blot. Adapted from Liu *et al*, 2009 (16). Data for SP were reported only where SSI, SSIIa and SbeI were used as bait.

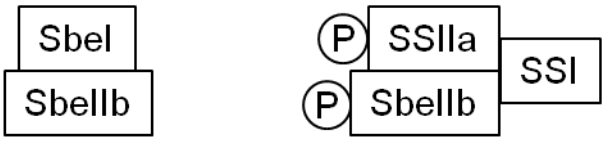


Figure 3.1: Two known starch biosynthesis protein complexes containing SbeIIb in maize. SP may also bind the SbeI/SbeIIb complex.

Maize SbeIIb structure and function

The full length maize SbeIIb transcript is translated into a 91 kDa, 799 amino acid preproprotein including a 61 residue N-terminal domain which mediates its migration to the amyloplasts and is then removed to form the mature protein (18). The mature form contains 48 serine, 31 threonine and 41 tyrosine residues, each of them a potential phosphosite. The N-terminal carbohydrate binding domain, C-terminal substrate specificity domain and central catalytic domain contains conserved among SBEs are all present in SbeIIb (9, 10).

Maize SbeIIb is phosphorylated in the granule and its catalytic activity and affinity for various proteins of the starch biosynthesis complex have been shown to be modulated by its phosphorylation state (13, 16, 19). A viable maize mutant lacking SbeIIb called amylose extender ($ae^-1.1$) exists. A related catalytically inactive SbeIIb mutant with a deletion spanning Val272 to Pro299, called $ae^-1.2$, is associated with starch that differs from both $ae^-1.1$ and wild type maize. Granules from wild type, $ae^-1.1$ and $ae^-1.2$ are successively smaller, lower in starch content and more irregular in texture. The $ae^-1.1$ and $ae^-1.2$ starches have fewer but longer glucan branches than wild type. Table 3.1 compares the protein networks of SbeIIb under conditions favourable to and inhibitory to protein phosphorylation and indicates that interactions between SbeIIb and SSI as well as SbeIIb and SbeI are dependent on phosphorylation but that attachment of SSIIa to SbeIIb is not. Addition of alkaline phosphatase to wild type amyloplast lysates appeared to abrogate the association of SbeIIb with all other proteins as assessed by SbeIIb immunoprecipitation however SbeIIb was detected in SSIIa immunoprecipitates under the same conditions. It is not clear which interactions of SbeIIb are direct. In addition, free

SbeIIb was shown to be phosphorylated within the plastid stroma, suggesting a role for the modification beyond coordinating protein binding (17). The sequence from Val272 to Pro299 (VFRHAQPKRPKSLRIYETHVGMSS) harbours three serine, one tyrosine and one threonine residue and could therefore accommodate multiple phosphates. At least one phosphosite outside of this region was assumed to exist since ae$^-$1.2 was phosphorylated *in vitro* (13).

LC-MS/MS was used to investigate SbeIIb modification in two ways. First, targeted analysis of proteolytic SbeIIb peptides shown to be phosphorylated by amyloplast extracts revealed their differential phosphorylation at three sites by two or more endogenous kinases. The affected residues were S286, S297 and S649. Secondly, shotgun sequencing of partially purified amyloplast stroma identified putative SbeIIb kinases. The kinase activity acting upon S649 and S286 was termed K1 and that activity acting on S649 and S297 was termed K2. Fractions K1 and K2 contained different protein kinases which was consistent with the observation that they had different substrate specificities. The contributions of this author to the work described in this chapter include all sample preparation for LC-MS/MS after the excision of gel bands, the collection and interpretation of all mass spectra.

3.2 Methods and Materials

Corn culture and endosperm amyloplast recovery

The CG102 line of maize was grown in a field in Guelph, Ontario, Canada. Whole cobs were harvested 20-25 days after pollination and from them were collected endosperm amyloplasts. Amyloplasts isolation and subsequent protein collection were performed according to previously described methods (20). Plastid disruption was conducted with 100 mM Tricine/KOH, pH 7.8, 1 mM Na_2-EDTA, 1 mM dithiothreitol (DTT), 5 mM $MgCl_2$ and 10 µl/ml of protease inhibitor cocktail (G-Biosciences, St. Louis, MO, USA). The isolated plastid preparations were stored at -80˚C. A single maize cob gave approximately 5 ml of amyloplasts with a total protein concentration of 1-2 mg/ml.

Recominant and synthetic SbeIIb polypeptides

Synthetic peptides spanning Serine-containing sequences of SbeIIb with tri-Arginine C-termini were produced by CanPeptide Inc. (Pointe-Claire, QC, Canada). These served as substrates for the amyloplast lysate kinase assay. The sequence corresponding to processed, or mature, SbeIIb (residues 62-799, GenBank Accession L08065) with a fused N-terminal S-tag contained in the plasmid vector pET 29 (Novagen, EMD Millipore) was a gift from Dr. Alan Myers of Iowa State University, USA. Truncated SbeIIb variants were created by PCR amplification of various segments of the coding sequence from the plasmid described above with subsequent ligation of each amplicon into pET29 and expression in *E. coli* ArcticExpress (Stratagene, La Jolla, CA, USA) upon induction with 1 mM isopropyl-beta-D-thiogalactopyranoside. Turbo polymerase was from Stratagene and deoxyribonucleoside 5'triphosphates were from Invitrogen (Burlington, ON, Canada). Recombinant proteins were purified using a Novagen^{TM} Protein Refolding Kit then with size-exclusion chromatography on a Superdex 200 10/300GL gel permeation column from which monomeric and inactive aggregates eluted separately. Sequences of all constructs were verified by DNA sequencing. Segments making up the truncated proteins were residues 62-732 (ΔC1), 62-646 (ΔC2), 62-562 (ΔC3), 128-799 (ΔN1), 214-799 (ΔN2) and 291-799 (ΔN3). Monomeric truncated proteins were used to isolate the location of phosphosites within the full length enzyme by amyloplast lysate kinase assay.

The QuikChange site-directed mutagenesis kit (Stratagene) was employed to separately replace S286, S297 and S649 of the recombinant SbeIIb with alanine residues. The S649 recombinant gene served as the template for the production of double and triple mutants. Reaction mixtures included 2.5U/reaction of PfuTurbo™ DNA polymerase II (Agilent, Santa Clara, CA, USA), 50 ng of vector and 125 ng of each forward and reverse primer. Point mutants were expressed and purified according to the same protocol used for the truncated mutants.

Measurement of amyloplast kinase kinetics and activity

The properties of kinases present in amyloplasts were determined based on their ability to radiolabel synthetic peptides or recombinant SbeIIb. The assay was performed with 200 µM of substrate and 10 µl of kinase preparation at pH 7.0 in 50 mM Tris-HCl, 1 mM DTT, 5 mM $MgCl_2$ and 200µM of [γ-32P]-ATP for 10 min at 25˚C. After this reaction period, the mixture was blotted onto phosphor-cellulose paper, washed in 1% (v/v) phosphoric acid then acetone and ^{32}P measured with a Beckman Coulter LS6500 Multi-Purpose liquid scintillation counter. Radioactivity was normalized to protein content assessed with Coomassie blue staining.

In vitro phosphorylation of recombinant SbeIIb and amyloplast stromal proteins

In vitro SbeIIb phosphorylation was performed by two methods, both of which included amyloplasts proteins as the source of kinase. In one method, 20 µg of SbeIIb was added to 100 µl of endosperm plastid lysate (1-1.5 mg protein/ml) or partially purified protein kinase in the presence of ATP or [γ-^{32}P]-ATP (Perkin-Elmer, Waltham, MA, USA) and 10 µl/ml Plant ProteaseArrest protease inhibitor mixture (G-Biosciences) for 30 minutes. Following exposure to ATP and kinases, S-tagged recombinant SbeIIb was collected on S-agarose beads over one hour. In the second method, 20 µg of monomeric recombinant SbeIIb was immobilized on S-protein agarose in the presence of protease inhibitors, unbound proteins were washed away in 20 mM Tris-HCl (pH 7.8), 0.15 M NaCl, 0.1% (v/v) Triton-X100 with 1 mM DTT then protein kinase and ATP were added to the washed beads. Phosphorylation was arrested in both methods by addition of Na_2-EDTA to a final concentration of 20 mM. Each phosphorylated SbeIIb sample was heated at 95˚C in loading buffer for 5 minutes prior to SDS-PAGE or digested with trypsin directly. Crude amyloplasts lysates were phosphorylated *in vitro* by the addition of [γ-^{32}P]-ATP (0.5 µCi).

Detection of radiolabelled-phosphopeptides in RP-HPLC eluates

In order to isolate phosphosites, kinase/[γ-^{32}P]-ATP–treated SbeIIb was digested with trypsin and the products separated by RP-HPLC using a Gemini-NX 3µ C18 110A column (Phenomenex, Torrance, CA, USA) on an Agilent 1100 Series HPLC pump with a 38 minute mobile phase gradient from 0.1% (v/v) formic acid and acetonitrile with formic acid, which was converted from 95% to 0% water over. Fractions containing phosphorylated peptides were identified with a Geiger counter, stored until no longer radioactive and subsequently analyzed by LC-MS/MS.

Partial Purification of SbeIIb kinases

SbeIIb kinase was partially purified from amyloplasts with an AKTA Explorer FPLC pump (Amersham Biosciences) at 4˚C. Initially, 10 to 12 ml of plastid lysate (5-10 mg of protein) was separated on a 1 ml HiTrap DEAE Sepharose Fast Flow column (GE Healthcare, Little Chalfont, UK) at a flow rate of 0.5 ml/min over a KCl gradient from 0 to 1 M in mobile phase containing 100 mM Tricine-NaOH (pH 7.5), 7.5 mM $MgCl_2$, 1 mM DTT. Kinase-containing fractions of the anion-exchange eluate were determined based on their ability to phosphorylate recombinant SbeIIb or synthetic SbeIIb peptides then pooled, desalted on a PD-10 column (GE Healthcare) and further separated on a 1 ml HiTrap Blue HP column (GE Healthcare) over a KCl gradient from 0 to 1.5 M. Proteins were separated by glycine SDS-PAGE on 10% acrylamide gels. Western blotting conducted with primary polyclonal rabbit anti-SbeIIb and anti-S-tag antibodies (AbCam, Cambridge, UK) confirmed the presence of SbeIIb.

LC-MS/MS of SbeIIb and its kinases

Prior to LC-MS/MS analysis of SbeIIb phosphorylation sites, the protein was separated by SDS-PAGE and bands equal in molecular mass to SbeIIb excised then subjected to in-gel digestion. Coomassie Blue stain was removed from the isolated bands with water containing 50% acetonitrile (v/v). The samples were washed twice with 100 mM ammonium bicarbonate (NH_4HCO_3) and reduced with 10 mM dithiothreitol in 100 mM

NH_4HCO_3 for 30 minutes at 60˚C then cooled to 21˚C. Cysteine alkylation was performed with 20 mM iodoacetamide in 100 mM NH_4HCO_3 in the dark for one hour. Excess iodoacetamide was removed with two washes of 100 mM NH_4HCO_3. The proteins were digested at 37˚C using 0.4 µg of endoproteinase Glu-C (Glu-C, Sigma-Aldrich) per gel band for 16 h, then with 0.4 µg of endoproteinase Lys-C (Lys-C, Sigma-Aldrich) per sample for 16 h. Alternately 0.1 to 0.4 µg of trypsin was used in place of Glu-C and Lys-C. Gel bands were washed with 100 µl volumes of water then 50% acetonitrile (v/v) and the washes were pooled, dried in a centrifugal concentrator, dissolved in aqueous 0.1 % (v/v) formic acid then purified using C18 zip tips (EMD Millipore, Darmstadt, Germany).

Purified Glu-C/Lys-C and tryptic SbeIIb digests as well as tryptic digests of kinase isolates were analyzed by online nanoflow HPLC-MS/MS using a NanoLC-Ultra 2D HPLC pump coupled to a Nanoflex cHiPLC system and a TripleTOF 5600 Qq-TOF mass spectrometer (AB Sciex, Concord ON, Canada). Peptides were loaded and washed on a 0.5 mm long, 200 µm diameter trap (AB Sciex) and resolved on a 150 mm long, 75 µm diameter chromatographic column (AB Sciex). The trap and the analytical column packing material consisted of 3 µm diameter C18 particles with 120Å pores. Peptide separation was performed using a binary mobile phase gradient. Mobile phases A and B contained 0.1% (v/v) formic acid in water and acetonitrile respectively. The following mobile phase B compositions were used: 2% at 0 min, 35% at 45 min, 80% from 46 to 50 min, 2% from 51 to 60 min. Nanospray emitters used were uncoated fused silica with 10 micrometer ID tips (New Objective, Woburn, MA, USA). The mass spectrometer was operated in positive ion mode with an electrospray voltage of 2300 V. The MS data acquisition cycle consisted of a single MS scan with a range of 400 to 1250 Th followed by MS/MS scans from 100 to 1800 Th of the most abundant ions detected in the preceding MS scan. The number of MS/MS scans per MS scan was limited to twenty. Peptides and phosphopeptides were identified from the tandem mass spectra using ProteinPilot™ software with phosphorylation emphasis and amino acid substitutions specified in the search parameters. Separate searches of the dataset were conducted using

the RefSeq *Zea mays* library which contained 22444 sequences representing 7089707 residues for peptide identifications (downloaded November 14, 2012). Glu-C/Lys-C digestion and cysteine carbamidomethylation were specified in the search terms and the confidence threshold for protein detection was set to 95%.

For the identification of amyloplast kinases, plastid lysates were separated into the K1 and K2 fractions as described above before reduction, alkylation with iodoacetamide and trypsin digestion. Digests were dried then dissolved in aqueous 0.1% (v/v) formic acid and the peptides resolved over an hour long analytical gradient on the nano-LC system described above coupled to an Orbitrap Elite mass spectrometer via a Flex^{TM} ion source (Thermo Scientific, Waltham, MA) connected to the same type of emitter used for the Qq-TOF. The data acquisition cycle began with a 30,000 resolution Orbitrap MS scan followed by MS/MS scans of the 10 most abundant precursor ions on the LTQ linear ion trap section of the instrument at normalized collision energy of 35, activation Q of 0.250 and a 10 millisecond activation time. Dynamic precursor exclusion was used with a repeat count of 1, repeat duration and exclusion duration both set to 15 seconds and an exclusion mass width of 10 ppm. Mass calibration was continually adjusted over the course of the analysis using the lock mass function set to 371.101240 Th, the m/z of a ubiquitous indoor air contaminant.

MS/MS data were searched with SEQUEST navigated through ProteomeDiscoverer software using the same FASTA format *Zea mays* sequence library as for the Qq-TOF data. Search settings were as follows: precursor mass range: 290-5000 Da; minimum peak count: 1; enzyme: trypsin; maximum missed cleavages: 2; product ion series considered: b and y; dynamic modifications: methionine oxidation (+15.995 Da), phosphorylation (S, T and Y at +79.966 Da) and glutamine deamidation (+0.984 Da); static modification: carbamidomethyl cysteine (+57.021 Da).

The predicted tertiary structure of SbeIIb

Three-dimensional modelling of SbeIIb predicted the position and role of each phosphosite in its tertiary structure. The sequence of the processed 738 residue protein

with the N-terminal sub-cellular localizing transit peptide region absent was used for *in silico* simulations. The initial structure was predicted using I-Tasser (21). Interrogation of the Protein Data Bank structure library showed SbeI from Asian rice (*Oryza sativa L.*) to be the most homologous to maize SbeIIb so this was used in the initial structure prediction. Due to low sequence homology between SbeIIb and rice SBEI, the 55 N-terminal residues were not simulated. SbeIIb molecular dynamics were modelled on the GROMACS version 4.5.5 software suite with the Gromos96 ffG53a6 force-field adapted for phosphorylated serine residues using an existing model obtained from the GROMACS user contribution forum, with partial charges and Van der Waals values for the atoms in the phosphate group (22). Phosphates were added to serine residues with SYBYL-X 1.3 molecular modeling software (Tripos Associates, St. Louis, MO, USA) with charge of -2 applied to each group. Terminal residues were assigned net charges of zero as were histidine side chains. Eight different SbeIIb structures were modelled, including the unmodified protein, the protein phosphorylated at S286, phosphorylated at S297, phosphorylated at S649, simultaneously phosphorylated at S286 and S297 or at S286 and S649 or at S297 and S649 and triply phosphorylated at S286, S297 and S649. Simulations were run on Compute Calcul Canada's SHARCNET computer cluster, accessed via www.sharcnet.ca. The hypothetical folding space enclosing each model was a cube allowing a minimum gap of 1 nm distance from the protein to the perimeter. The cube was populated with water molecules, sodium and chlorine ions to a salt concentration of 0.15 M. Folding was simulated for 100 ns.

3.3 Results

Three modified amino acids from amyloplast treated SbeIIb were observed by targeted MS/MS while data-dependent acquisition of kinase-active amyloplast anion exchange chromatography eluates showed them to contain multiple fraction-specific protein kinases.

SbeIIb phosphosite localization using truncated variants and synthetic peptides

S649 was indicated as the main site of SbeIIb phosphorylation using a kinase assay performed on a series of recombinant and synthetic polypeptide substrates. The same assay also showed S286 and S297 to be potential kinase substrates. Recombinant SbeIIb was readily phosphorylated in the presence of amyloplast lysates and ATP, as the endogenous equivalent and the ae^{-}1.2 mutant were shown to be in previous work (13). Recombinant maize SbeIIb mutants with portions of their N- or C-terminal sequences were produced in *E. coli* and tested for reactivity with amyloplasts lysates to establish through a process of elimination the regions of the protein affected by kinase activity. The truncations were designed to maximize sequence coverage of the mature sequence of SbeIIb while isolating potential phosphosites. The sequence spanning residues 62-732 underwent ^{32}P-labelling to a degree equivalent to the full-length protein while radiolabelling of the 62-646 and 62-562 truncations was very low. From this data it could be concluded that multiple residues between M62 and S732 were modified and that the region from 646-732 received the most labelling. Putative phosphorylation sites within this region determined by NetPhos 2.0 at probability greater than 90% were incorporated into synthetic peptides which separately covered S649, S286 and S297 (23). A synthetic peptide containing the sequence surrounding S649 (PRGPQRLPSGKFIPGNRRR) was synthesized and shown to be phosphorylated by the same amyloplast preparation. Addition of the synthetic S649 peptide to either native or recombinant SbeIIb in micromolar concentrations inhibited their phosphorylation whereas other endogenous amyloplast proteins were readily ^{32}P-labelled in its presence.

Two endogenous kinase modalities with distinct specificities for three SbeIIb regions were resolved by anion exchange chromatography of amyloplast lysates and these were termed K1 and K2, reflecting the order of their elution (Figure 3.2). That the SbeIIb kinases were present in the same sub-cellular compartment suggests a biological role for its modification. Synthetic peptide S649 was readily modified by both fractions. Peptide S286 was phosphorylated exclusively by K1 while peptide S297 was phosphorylated exclusively by K2. Both S286 and S297 underwent much less modification than did

S649 with either fraction. Amyloplast isolates typically contained less than 1% protein contamination from the remainder of the cell, suggesting the SbeIIb kinases were specific to the organelle (24).

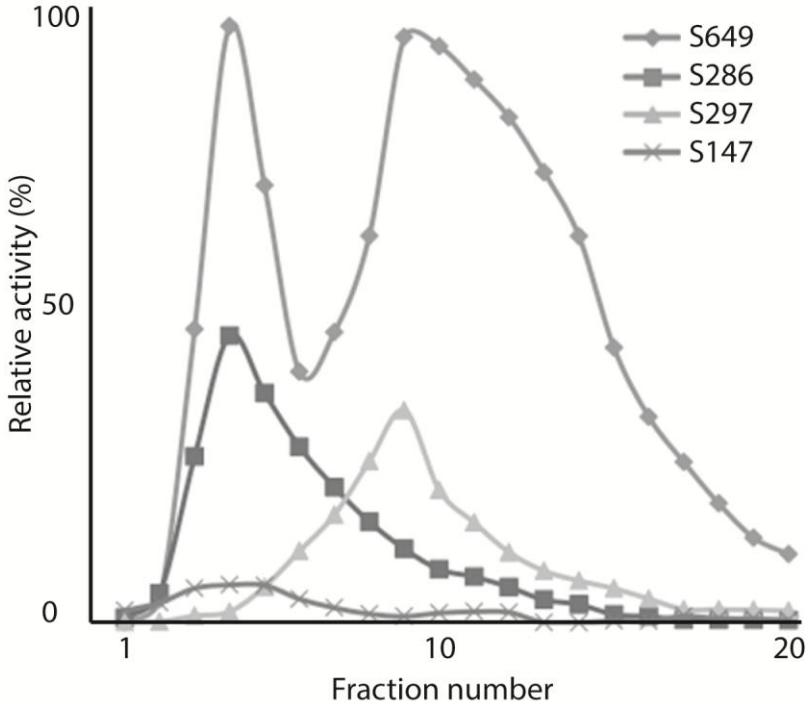


Figure 3.2: Rates of synthetic peptide phosphorylation upon treatment with anion exchange chromatography fractions of amyloplast lysate. Each curve shows the reaction of a different synthetic peptide. K1 appeared most concentrated in fraction 4 whereas K2 was apparently most concentrated in fraction 9. A synthetic peptide equivalent to SbeIIb S147 and the region flanking it was used as a negative control. Adapted from Makhmoudova *et al.* 2014 (24).

Wild type and S649A point mutants were both radiolabelled with [γ-^{32}P]-ATP by K1 and K2, though the signals of S649A were much weaker. S649A had equivalent enzymatic activity to the wild type protein based on *in vitro* phosphorylase stimulation, suggesting the site to be more important for protein-protein interaction than catalysis. In contrast, mutation of S286 or S297 significantly lowered catalytic activity. That S286 and S297 are conserved in all known plant SBEs supports the conclusion that they are involved in

the starch branching reaction. S286A/S649A and S297A/S649A double point mutants were used to verify the specificities of K1 and K2 since both fractions affected S649 disproportionately. The S286A/S649A mutation abrogated ^{32}P-labelling by K1 and S297A/S649A was not labelled by K2.

K1 and K2 had high specificities for their substrates as indicated by micromolar Michaelis constants (K_m) using the synthetic peptides (24). Catalytic activities of both kinase fractions for each of the three peptides varied with Ca^{2+} concentration and even 1 mM of Na_2-EGTA reduced activities of both fractions on the S649 synthetic peptide by two orders of magnitude while Ca^{2+} at concentrations as low as 5 μM eliminated the inhibitory effect of EGTA. Kinase activity was lowest in the case of K2 acting on the S297 peptide (Figure 3.2).

Molecular dynamics simulation analysis of SbeIIb

Though structural rearrangement was simulated for 100 ns, most took place in the first 10 ns at which point a steady state was achieved, consistent with a highly favoured conformation. S286, S297 and S649 all migrated to the protein surface at disordered loops where they were presumably kinase accessible. S286 and S297 located to opposite sides of a central β-barrel. Simulations of the phospho-SbeIIb variants showed minimal change in the overall structure took place. Salt-bridging between phosphate groups and several Arg and Lys residues stabilized conformation, for instance phospho-S286 interacted with K285 as well as R288 as did phospho-S297 and Arg665 (24). These data indicate that each serine residue is a reasonable phosphosite and that the phosphorylation states of S286 and S297 control the structure of the catalytic domain.

LC-MS/MS of SbeIIb phosphorylation sites

The three putative SbeIIb residues were observed in their unmodified, alanine point mutated and singly phosphorylated states by LC-MS/MS. It is apparent from the sequence of maize SbeIIb (Figure 3.3) that S286 and S649 are each separated from tryptic cleavage sites by no more than two amino acids at their N- and C-termini. To produce

larger peptides encompassing these sites, double digestion with Glu-C and Lys-C was employed. The larger peptides gave more MS/MS sequence information, had masses better suited to the instrument mass range and potentially had more favorable C18 retention properties than shorter tryptic equivalents. Signal intensities of the phosphopeptide precursor ions were low compared to the corresponding peptide ions however this was compensated by high resolution Qq-TOF spectra having low noise. Table 3.2 shows that product ions of SbeIIb phosphopeptides had m/z values differing from theory by less than 0.02 Th on average.

(MAFRVSGAVLGGAVRAPRLTGGGEGSLVFRHTGLFLTRGARVGCSGTHGAMRAAAAARKAV**)**MVP
EGENDGLASGADSAQFQSDELEVPDISEETTCGAGVADAQALNRVRVVPPPSDGQKIFQIDPMLQGY
KYHLEYRYSLYRRIRSDIDEHEGGLEAFSRSYEKFGFNRSAEGITYREWAPGAFSAALVGDFNNWDP
NADRMSKNEFGVWEIFLPNNADGTSPIPHGSRVKVRMDTPSGIKDSIPAWIKYSVQAPGEIPYNGIYY
DPPEEVKYVFRHAQPKRPK**S**LRIYETHVGM**S**SPEPKINTYVNFRDEVLPRIKKLGYNAVQIMAIQEHSY
YGSFGYHVTNFFAPSSRFGTPEDLKSLIDRAHELGLLVLMDVVHSHASSNTLDGLNGFDGTDTHYFHS
GPRGHHWMWDSRLFNYGNWEVLRFLLSNARWWLEEYKFDGFRFDGVTSMMYTHHGLQVTFTGNF
NEYFGFATDVDAVVYLMLVNDLIHGLYPEAVTIGEDVSGMSTFALPVHDGGVGFDYRMHMAVADKWI
DLLKQSDETWKMGDIVHTLTNRRWLEKCVTYAESHDQALVGDKTIAFWLMDKDMYDFMALDRPSTP
TIDRGIALHKMIRLITMGLGGEGYLNFMGNEFGHPEWIDFPRGPQRLP**S**GKFIPGNNNSYDKCRRRFD
LGDADYLRYHGMQEFDQAMQHLEQKYEFMTSDHQYISRKHEEDKVIVFEKGDLVFVYNFHCNNSYF
DYRIGCRKPGVYKVVLDSDAGLFGGFSRIHHAAEHFTADCSHDNRPYSFSVYTPSRTCVVYAPVE

Figure 3.3: The sequence of SbeIIb with S286, S297 and S649 indicated in bold and the nearest tryptic cleavage sites underlined. The 61-amino acid transit peptide is shown in round brackets.

Phosphorylation at S286 was detected in trace amounts following K2-treatment and in greater abundance in K1-treated samples by LC-MS/MS. Serine and tyrosine in the Glu-C/Lys-C peptide SLRIYE are both potential phosphosites; however product ion spectra showed serine to be the modified residue (Figure 3.4A). A complete y-ion series matched that expected of the unmodified peptide, while all a-ions and b-ions present were phosphorylated. A peak at m/z 136.1 was attributed to an iminium ion of unmodified tyrosine.

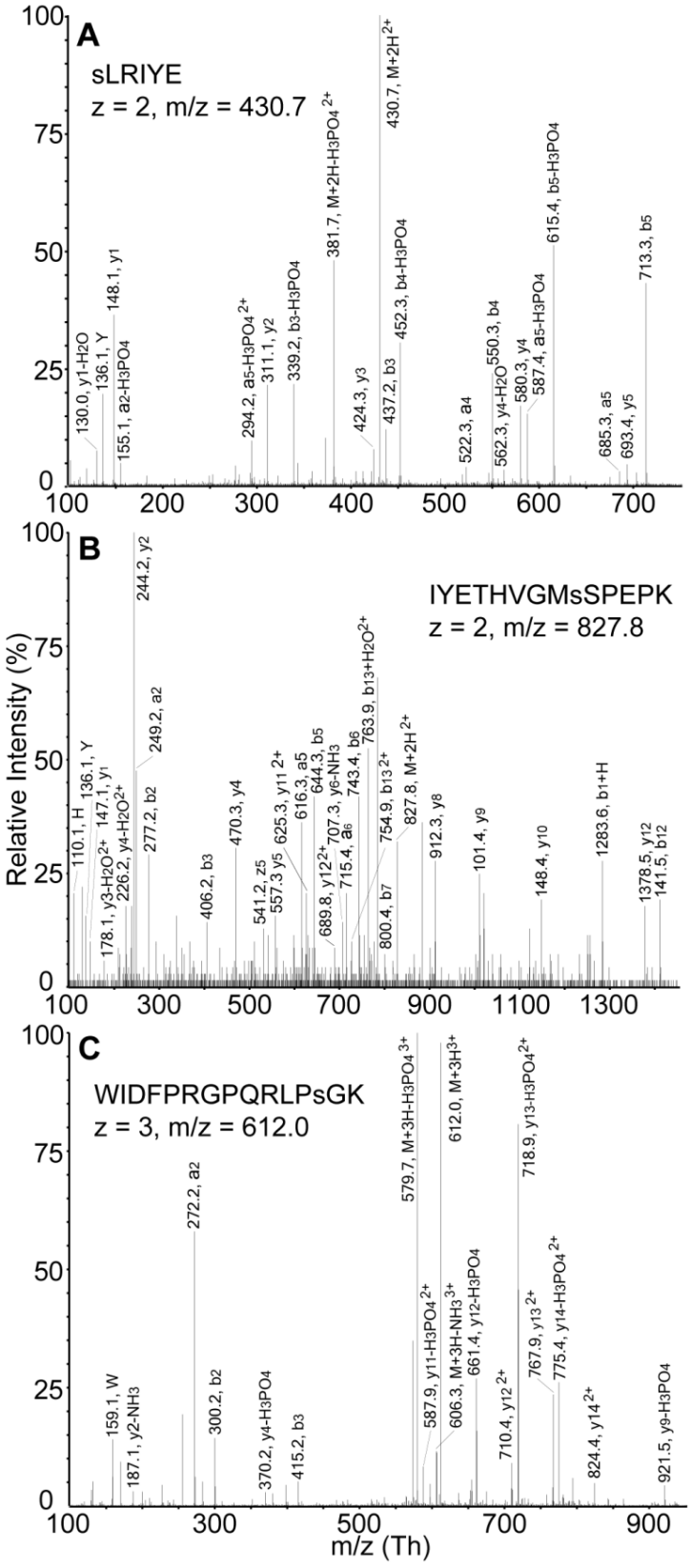

A
sLRIYE
z = 2, m/z = 430.7
B
IYETHVGMsSPEPK
z = 2, m/z = 827.8
C
WIDFPRGPQRLPsGK
z = 3, m/z = 612.0
Relative Intensity (%)
m/z (Th)

Figure 3.4: Tandem mass spectra of SbeIIb phosphopeptides with phosphates located at amino acid residues corresponding to S649, S286 and S297. Top panel: MS/MS of the doubly protonated Glu-C/Lys-C-generated phosphopeptide SLRIYE with the phosphate located at the S286 site from the K1-treated sample. Middle panel: MS/MS of the of the doubly protonated tryptic phosphopeptide IYETHVGMSSPEPK with the phosphate located at the S297 site, obtained from the fourth C18 RP-HPLC fraction of trypsin-digested K2-treated triple mutant of SbeIIb S286A/S297A/S649A. This spectrum was acquired from endogenous SbeIIb in the triple mutant sample, co-purifying with the protein kinase preparation. The presence of SbeIIb in kinase preparations was confirmed by LC-MS/MS (data not shown). Bottom panel: MS/MS of the triply protonated Glu-C/Lys-C-generated phosphopeptide WIDFPRGPQRLPSGK from the K1-treated sample with the phosphate located at the S649 site.

The S297 phosphopeptide was observed exclusively in K2-treated SbeIIb, from which it was identified with 99% confidence by a ProteinPilot search against the whole *Zea mays* library. The peptide THVGMSSPEPK, which contains the S297 site, was observed from Glu-C/Lys-C digests in both methionine-oxidized and unmodified forms with retention times between 12 and 13 min into the 60-min chromatographic run. The Glu-C/Lys-C peptides containing S297 had consistently low LC-MS signals. For instance, an untargeted MS/MS analysis of K2-treated wild-type SbeIIb resulted in relative signal intensities of 1 and 8 and 78 for the monoisotopic precursor ions of oxidized THVGMSSPEPK, phosphorylated WIDFPRGPQRLPSGK (S649) and non-phosphorylated WIDFPRGPQRLPSGK respectively. The concentrations of these species should be comparable since all originated from a single protein, however signals of the S649 containing peptide and phosphopeptide were orders of magnitude higher than those of the S297 containing peptide, suggesting the analytical platform had differential sensitivity for these analytes. Chromatograms of the oxidized and unmodified peptide THVGMSSPEPK were compared to those of other peptides observed in the same runs in order to determine if sensitivity was limited by low HPLC performance. Oxidized THVGMSSPEPK and unmodified THVGMSSPEPK had chromatographic qualities

comparable to other analytes. Despite considerable difference in their areas, the peaks of the oxidized S297-containing peptide and the unmodified S649-containing peptide were 0.2 min and 0.22 minutes full width at half maximum (FWHM) respectively, suggesting that HPLC performance did not selectively reduce either signal. As evidenced by its early elution time, THVGMSSPEPK is a hydrophilic peptide and may not be effectively retained by the C18 trap. The grand averages of hydropathy for the Glu-C/Lys-C peptide THVGMSSPEPK and the corresponding tryptic peptide IYETHVGMSSPEPK were -0.945 and -0.764 respectively, indicating both were hydrophilic (25). The low signals of THVGMSSPEPK suggest observation of the corresponding phosphopeptide was unfavoured, since MS sensitivity for phosphopeptides is lower than for peptides.

Phosphorylated S297 was not observed in Glu-C/Lys-C digests of SbeIIb, but was identified by LC-MS/MS of tryptic digests previously fractionated off-line by RP-HPLC. Consistent with its higher calculated hydrophobicity, IYETHVGMSSPEPK eluted later than the Glu-C/Lys-C equivalent digest, with a retention time of 21 min. Phospho-IYETHVGMSSPEPK had a retention time of 23 min. The monoisotopic ion of the triply charged tryptic phosphopeptide had a maximum signal of 65,000 counts in the precursor ion scan. While the signal may have increased due to C18 HPLC pre-fractionation, the sensitivity of the LC-MS system for this species was apparently much higher than for the Glu-C/Lys-C equivalent digest. Modification of S297 was the slowest of all reactions studied here so the population of phospho-S297 and the corresponding MS signal were expected to be low. Consistent with this expectation, MS/MS of the phospho-S297 species had the lowest s/n of all three phosphopeptides studied. The apparent absence of the Glu-C/lys-C phosphopeptide could therefore be due to the combination of low LC performance and low concentration.

Potential kinase targets within the peptide IYETHVGMSSPEPK include two Ser, one Tyr and one Thr residue. Product ion spectra showed the phosphate to be absent from all but the S297 site (Figure 3.4B). Abundant a_2, b_2, and b_3 ions indicated that tyrosine was not phosphorylated while b_5, a_6, b_6 and b_7 ions suggested both Tyr and Thr were unmodified. The Ser closest to the C-terminal of the peptide was observed in an unmodified state

within a y_5 ion having a mass accuracy of 0.0037 Da and s/n between 5.5:1 and 11:1. All product ions containing the S297 site also contained a phosphate. Two series of phosphorylated ions, from y_8 to y_{12} and b_{11} to b_{13}, supported the assignment of the ninth residue within the peptide as the phosphorylation site.

S649 is flanked by arginine and lysine so that trypsin digestion yielded the pentapeptide LPSGK with an m/z of 251.15 for the doubly protonated ion. Glu-C/Lys-C digestion was used to extend the length of the S649-containing analyte to match the mass range of the mass spectrometer. A S649 phosphopeptide was present in both K1- and K2-treated recombinant SbeIIb samples. A missed Glu-C cleavage was consistently observed around the S649 site, resulting in a peptide with the sequence WIDFPRGPQRLPSGK, which was identified with 99% confidence in searches against the *Zea mays* protein library. Low digestion efficiency at the third Asp of the S649 peptide may be due to the Pro located two residues C-terminal to it, since Pro at this position has been shown to inhibit protease activity (26). A Ser residue is the only potential phosphorylation site within this sequence and MS/MS data acquired from phospho-WIDFPRGPQRLPSGK featured product ions consistent with serine phosphorylation (Figure 3.4C). The fragments y_{12}, y_{13} and y_{14}, containing both the C-terminus and the phosphate, were detected with s/n values of 56:1, 150:1 and 29:1 respectively. Six prominent peaks corresponded to fragments having undergone phosphoric acid loss, an energetically favored fragmentation pathway of phosphopeptides. A series of N-terminal fragments which did not contain the Ser were without modifications. Using the combined MS/MS evidence, the presence and location of the phosphate at S649 in both K1-treated and K2-treated SbeIIb were deduced.

Untreated SbeIIb, as well as kinase-treated SbeIIb S649A and a SbeIIb S286A/S297A/S649A triple mutant were digested then analyzed by LC-MS/MS under conditions identical to those used for kinase-treated wild-type SbeIIb to further assess the kinase-dependence of the phosphorylation. While unmodified peptides containing S649, S286 and S297 were observed in the untreated controls, no phosphorylation of these sites was detected in these samples (data not shown).

Ion	Observed monoisotopic m/z (Th)	Theoretical monoisotopic m/z (Th)	Error (Th)
WIDFPRGPQRLPsGK, 3+, S649			
W	159.0923	159.0917	0.0006
y2-NH3	187.0873	187.1077	-0.0204
a2	272.1771	272.1757	0.0014
b2	300.1719	300.1707	0.0012
y4-H3PO4	370.2095	370.2085	0.0010
b3	415.1980	415.1976	0.0004
M+3H-H3PO43+	579.3198	579.3179	0.0019
y11-H3PO42+	587.8452	587.8438	0.0014
M+3H-NH33+	606.3033	606.3013	0.0020
M+3H3+	611.9785	611.9768	0.0017
y12-H3PO42+	661.3788	661.3780	0.0008
y122+	710.3735	710.3665	0.0070
y13-H3PO42+	718.8930	718.8915	0.0015
y132+	767.8812	767.8799	0.0013
y14-H3PO42+	775.4346	775.4335	0.0011
y142+	824.4234	824.4220	0.0014
y9-H3PO4	921.5237	921.5265	-0.0028
sLRIYE, 2+, S286			
y1-H_2O	130.0499	130.0499	0.0000
Y	136.0760	136.0757	0.0003
y1	148.0610	148.0604	0.0006
a2-H_3PO_4	155.1181	155.1179	0.0002
a5-$H_3PO_4^{2+}$	294.1885	294.1868	0.0017
y2	311.1244	311.1238	0.0006
b3-H_3PO_4	339.2158	339.2139	0.0019
M+2H-$H_3PO_4^{2+}$	381.7122	381.7109	0.0013
y3	424.2090	424.2078	0.0012
M+2H^{2+}	430.7016	430.6993	0.0023
b3	437.1910	437.1908	0.0002
b4-H_3PO_4	452.2996	452.2980	0.0016
a4	522.2810	522.2800	0.0010
b4	550.2769	550.2749	0.0020
y4-H_2O	562.3009	562.2984	0.0025
y4	580.3097	580.3089	0.0008

a5-H_3PO_4	587.3655	587.3664	-0.0009
b5-H_3PO_4	615.3621	615.3613	0.0008
a5	685.3398	685.3433	-0.0035
y5	693.3952	693.3930	0.0022
b5	713.3400	713.3382	0.0018
IYETHVGMsSPEPK, 2+, S297			
H	110.0719	110.0713	0.0006
Y	136.0755	136.0757	-0.0002
y1	147.1114	147.1128	-0.0014
y3-H_2O^{2+}	178.1372	178.1024	0.0348
y4-H_2O^{2+}	226.1562	226.6288	-0.4726
y2	244.1667	244.1656	0.0011
a2	249.1617	249.1598	0.0019
b2	277.1556	277.1547	0.0009
b3	406.1993	406.1973	0.0020
y4	470.2632	470.2609	0.0023
z5	541.2242	541.2742	-0.0500
y5	557.2967	557.2930	0.0037
a5	616.3092	616.3089	0.0003
$y11^{2+}$	625.2445	625.2678	-0.0233
b5	644.3119	644.3039	0.0080
$y12^{2+}$	689.7637	689.7891	-0.0254
y6-NH_3	707.3421	707.2648	0.0773
a6	715.3755	715.3774	-0.0019
b6	743.3716	743.3723	-0.0007
$b13^{2+}$	754.8792	754.8100	0.0692
b13-H_3O^{2+}	763.8825	763.8153	0.0672
b7	800.4016	800.3937	0.0079
M+$2H^{2+}$	827.8365	827.8628	-0.0263
y8	912.2966	912.3533	-0.0567
y9	1011.3678	1011.4217	-0.0539
y10	1148.4197	1148.4806	-0.0609
b11+H	1283.5984	1283.5252	0.0732
y12	1378.5131	1378.5709	-0.0578
b12	1411.5067	1411.5600	-0.0533

Table 3.2: Mass accuracy of each product ion attributed to the phosphopeptide spectra depicted in Figure 3.4.

LC-MS/MS identified protein kinases in amyloplasts extracts

Information dependent acquisition of amyloplast anion exchange chromatography fractions revealed the presence of four kinases, two of which appeared to be unique to K1 and two which appeared to be unique to K2. K1 contained the uncharacterized putative protein kinase LOC100383602 (GI: 293336498) as well as its homolog Calcium-dependent protein kinase 2 (CDPK2, GI: 162463011). The two proteins share 69% sequence identity from V60 of LOC100383602 to their C-termini and were distinguished by six unique tryptic peptides. Calcium-dependent protein kinase ZmCPK11 (GI: 162460670) and abscisic acid (ABA) stimulation MAP kinase (GI: 226495415) were identified with high confidence by several peptides each in K2 samples. Unlike the K1 kinases, the two K2 kinases have less than 30% sequence identity overall, however ZmCPK11 shares around 60% of its sequence with both K1 kinases. Though the functions of the aforementioned proteins within the endosperm remain unknown, ZmCPK11 is known to be transcribed at high concentrations in the maize seed (27). ABA stimulation MAPK is unique among the K1 and K2 kinases in its regulation of gene expression within the seed in response to abscisic acid (28). Starch synthase and SBE gene expression in the endosperm are regulated by ABA, therefore ABA stimulation MAPK likely has functions other than SbeIIb activation (29). That three similar gene products, namely CDPK2, ZmCPK11 and LOC100383602, are expressed within one organelle suggests similar but distinct roles for each. It is possible that two or more of these kinases respond to a single stimulus, namely calcium signalling, by separately functionalizing those SbeIIb residues controlling its productivity. Such an activation mechanism could exert multiple levels of control on the starch biosynthesis complex by modulating the proximity of its members as well as the activity of SbeIIb. It is striking that two serine residues affecting the behaviour of the catalytic domain are activated by separate kinases, so that a wide range in the starch branching rate is possible.

Tandem mass spectra used in the assignment of protein identities to K1 and K2 are illustrated in Figures 3.5 and 3.6. Peptides identified at 95% confidence spanned 11% of ABA stimulation MAPK, 11% of ZmCPK11, 8% of CDPK2, and 9% of LOC100383602

respectively. Though overall sequence coverage was low, each protein was identified by multiple unique peptides sequenced with greater than 99% confidence. Extensive product ion coverage was obtained from MS/MS data, for example each of the spectra in Figures 3.5 and 3.6 contain overlapping b-series and y-series ions which together account for all residues. Observed product ion mass accuracy was within the typical range of a linear ion trap, with most ions having absolute mass deviations within 0.1 Th. Precursor masses measured on the orbitrap agreed with predicted values within 5 ppm. The LTQ ion trap spectra consistently featured low intensity noise over a wide range of m/z values below 1000 Th. In general, smaller product ions had lower s/n than those above 800 Th, however their signals were adequate in differentiating them from noise and establishing their charge states.

In addition to protein kinases, K1 was incidentally found to contain serine/threonine-protein phosphatase 5 (GI: 226505390), a possible negative regulator of SbeIIb. This phosphatase was recognized by four unique peptides identified with 99% confidence. MS/MS spectra from these species contained nearly complete b- and y-ion series with high s/n. Though its role in maize has not been reported, the *Arabidopsis thaliana* homolog is known to mediate heat stress tolerance (30). The amyloplast fractions contained other proteins not directly associated with starch synthesis such as actin, ribosomal proteins and enolase. In total, 67 and 89 proteins were identified by two or more peptides at 95% confidence in the K1 and K2 respectively. These contaminants reduced acquisition time available for sequencing the kinase peptides.

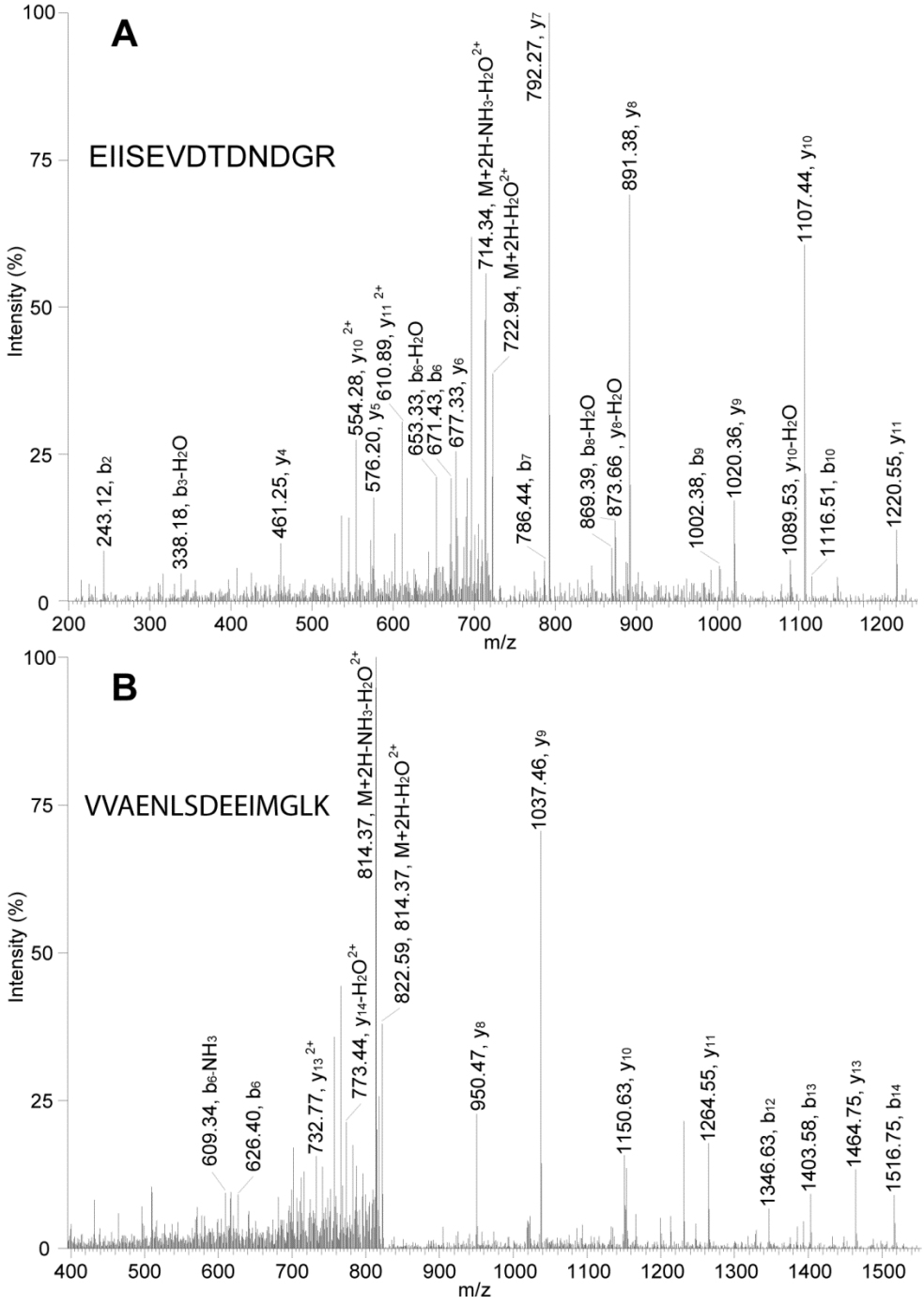


Figure 3.5: Representative tandem mass spectra of tryptic kinase peptides obtained from amyloplast fractions K1. A: doubly protonated EIISEVDTDNDGR from CDPK2 with a retention time of 31.58 minutes, a SEQUEST cross correlation score of 2.97 and a precursor ion mass accuracy of -0.00042 Da. B: methionine oxidized

VVAENLSDEEIMGLK from LOC100383602 with a retention time of 38.38 minutes, a SEQUEST cross correlation score of 4.16 and a precursor ion mass accuracy of -0.00237 Da.

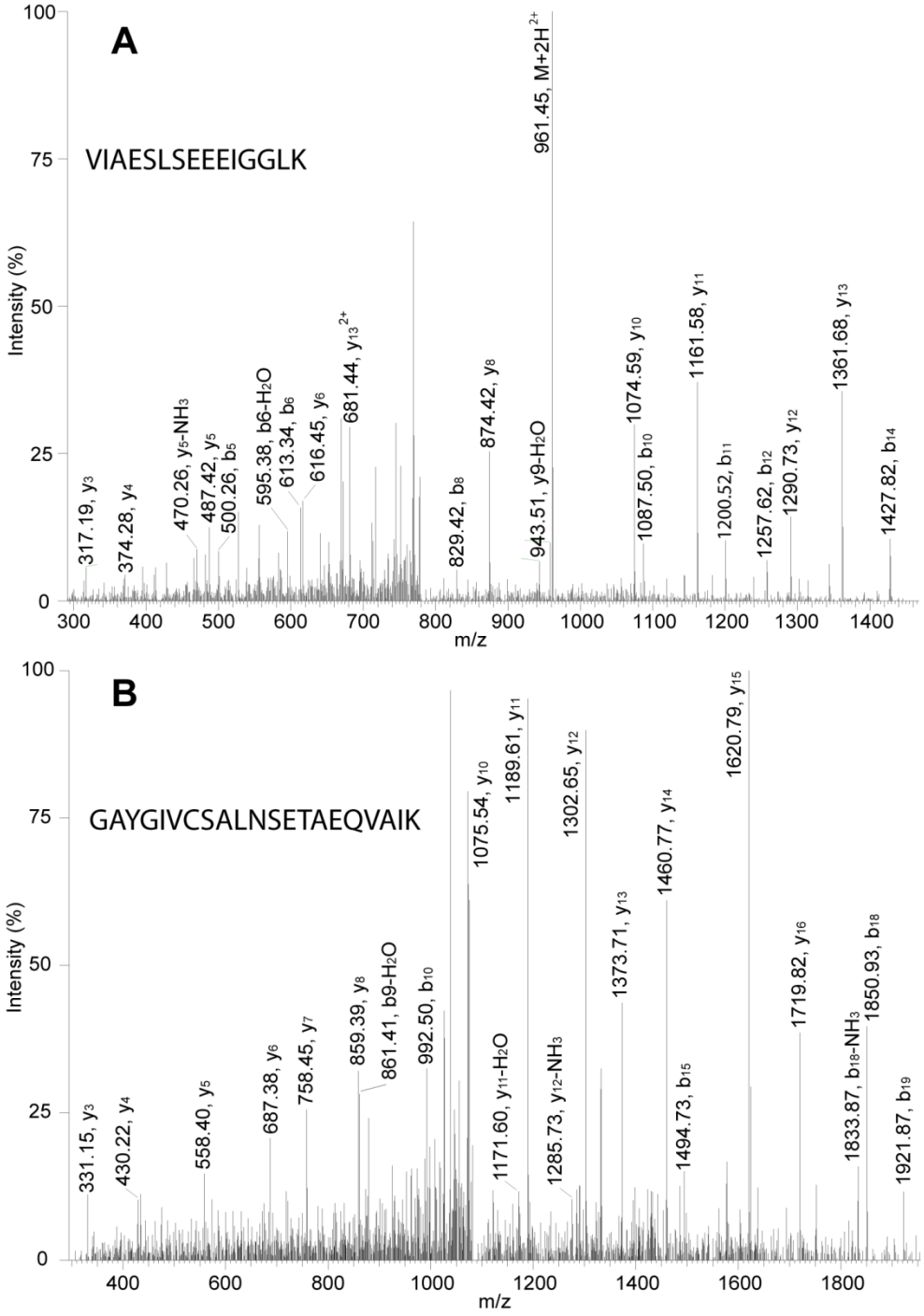


Figure 3.6: Representative tandem mass spectra of tryptic kinase peptides obtained from amyloplast fractions K2. A: VIAESLSEEEIGGLK from ZmCPK11 with a retention time of 39.37 minutes, a SEQUEST cross correlation score of 4.96 and a precursor ion mass accuracy of -0.00083 Da. B: GAYGIVCSALNSETAEQVAIK from ABA stimulation MAP kinase with a retention time of 53.75 minutes, a SEQUEST cross correlation score of 3.23 and a precursor ion mass accuracy of 0.00527 Da.

3.4 Conclusion

SBEIIb plays a central role in amylopectin production within maize endosperm, working in concert with several other enzymes. The complex was suggested to be coordinated by phosphorylation though the phosphosites, kinase(s) and phosphatase(s) exerting such control remained to be discovered. To explore activation of the starch synthesis complex, native SbeIIb, recombinant SbeIIb and several mutants were exposed to crude amyloplast isolates and found to undergo phosphorylation to different extents at three serine residues, namely S649, S286 and S297. Two WAX fractions of amyloplast lysates differed in their ability to phosphorylate residues S286 and S297 and were therefore assumed to contain different kinases. The two plastidial isolates phosphorylated each of three synthetic peptides harbouring S286 or S297 or S649 with the same differential specificity they showed for the full length enzyme. Amyloplast isolates were partially sequenced by bottom-up LC-MS/MS and found to contain three distinct members of a calcium-dependent serine/threonine kinase family. Modification of the SbeIIb phosphosites by each isolate was found to depend on Ca^{2+} concentration, suggesting that the identified kinases were those acting on SbeIIb. The discovery of calcium dependence establishes a novel upstream regulatory mechanism and will improve the efficiency of *in vitro* SbeIIb activation in ongoing research.

Sites S286 and S297 were shown to modulate catalytic function and are conserved in all known plant SBEs, which suggests their role is maintained across plant species. S649 likely serves a more specialized role in protein-protein interaction as it occurs only in a subset of SBEs and its replacement with alanine does not affect SbeIIb catalysis.

Whatever the function of S649, it is clearly modified more readily than the other two sites as evidenced by biochemical characterization as well as LC-MS.

Three dimensional molecular dynamic simulations indicated that S286, S297 and S649 all orient to the protein surface in disordered loops accessible to other proteins. S286 and S297 are located on either end of the central catalytic β-barrel where they could influence its structure or interact directly with the substrate. Since replacement of S286 or S297 with alanine decreased *in vitro* starch branching catalysis significantly, it could be concluded that these sites have important roles *in vivo* (24).

Here are shown for the first time the locations of phosphosites on a protein which catalyzes a key reaction in starch biosynthesis and is known to recruit enzymes with related roles. Though previously shown to undergo phosphorylation along with other amyloplasts proteins, SbeIIb is the first enzyme of the group to have its phosphosites identified. That corn varieties featuring inactive SbeIIb yield distinct starches points to the potential of strains containing SbeIIb phosphosite mutants to produce novel variants of this important commodity. The roles of the three phosphosites in the assembly of the starch synthesis complex and the behaviour of SbeIIb itself with respect to phosphorylation kinetics remain to be determined. Since the phosphosites are modified to different degrees by at least two kinases, their functions in activity regulation are likely more sophisticated than switching the enzyme between completely inactive and completely active forms.

3.5 Contributors

Amina Makhmoudova (Ph.D.) performed all cloning and site-directed mutagenesis for this study, as well as maize seed processing, recombinant protein isolation and biochemical assays. Jenelle Patterson (B.Sc.) performed chromatographic purifications of the kinases as well as the peptide kinase assay. Kenrick A. Vassall (Ph.D.) performed the molecular dynamics simulations.

Chapter 4

Identification of Beta-catenin interacting partners in a vascular smooth muscle cell model

4.1 Introduction

Cellular proliferation and differentiation are essential processes for development, growth and repair of animal tissues and organs coordinated through intercellular and intracellular signalling. The elucidation of pathway components and their interaction mechanisms holds important promise in regenerative medicine and oncology. Novel therapies promoting dedifferentiation and growth may enable the replacement of damaged nerves and cardiac muscle, which normally lack regenerative capacities, while cancer treatments are intended to suppress these processes (1, 2). Additionally, excessive proliferation of vascular smooth muscle cells (VSMC), the major component of blood vessels, has been implicated in cardiovascular diseases such as atherosclerosis and restenosis, leading causes of death in Western countries (3). VSMC growth is therefore a potential target of pharmacological intervention in cardiovascular illness.

The secreted signalling proteins wingless-type mouse mammary tumor virus integration site (Wnt) are key agonists of VSMC proliferation and muscle cell differentiation (3, 4, 5). Upon binding to the receptors Frizzled and low-density lipoprotein-related receptor 5 and 6 (LRP5/6), Wnt causes them to form a heteromeric cell surface receptor complex, which promotes the transcription of numerous genes (6).

In the basal state, a cytosolic protein complex including Axin, casein kinase 1 (CK1), glycogen synthase kinase 3 beta (GSK3B) and adenomatous polyposis coli (APC) inhibits Wnt target gene expression through its effect on the transcription mediator Beta-catenin. This degradation complex is inactivated following phosphorylation of the intracellular region of LRP5/6 and of the protein Dishevelled, which possibly promotes the interaction

of Axin and Dishevelled, upon stimulation by Wnt (7, 8, 9). Interestingly, membrane bound forms of GSK3 and CK1, the enzymes responsible for negative regulation of the Wnt signal cascade, are responsible for the activating phosphorylation of LRP (10, 11). When unconstrained by CK1/GSK3B/APC, Beta-catenin accumulates in the nucleus where it disrupts interaction between the transcription repressor transducin-like enhancer of split (TLE) and the transcription factors T cell factor/Lymphoid enhancer factor, displacing the former to allow DNA binding by the latter (12). This assembly is known as the canonical Wnt signalling network since Wnt is known to bind other receptors with different downstream effectors.

Beta-catenin is an 86 kDa protein with a sequence highly conserved in animals. In the nucleus, Beta-catenin binds to and enhances the activity of a host of transcription factors including TCF/LEF, Nuclear Factor-Kappa Beta, Sex-determining region Y-related high-mobility-group box (SOX) and Forkhead box-O (FOXO) but cannot itself bind DNA (9, 12). In the absence of Wnt, Beta-catenin distribution is restricted to the inner surface of the plasma membrane where it is bound to the intracellular region of the transmembrane protein cadherin. Here Beta-catenin links cadherin to the actin cytoskeleton through alpha-catenin (13). This interaction protects Beta-catenin from sequential polyphosphorylation at its amino terminal end by CK1 and GSK3, a modification which targets it for ubiquitylation and ultimately degradation (14). In the cytoplasm Beta-catenin is rapidly consumed by this process, however where the degradation complex is disrupted by LRP5/6 and Dishevelled, it accumulates in the cytoplasm, migrates to the nucleus and coordinates transcription.

As might be inferred from the variety of transcription factors it influences, Beta-catenin activity is modulated by many proteins beyond the canonical Wnt signalling network. Other than CK1-dependent phosphorylation at S45 and GSK3-dependent phosphorylation at S33, S37 and T41, Beta-catenin is known to undergo 18 additional phosphorylations and two acetylations, and these modifications are associated with functions such as stabilization and relocalization of the protein (15). For instance, epidermal growth factor receptor stimulation of the glioma cell line U87MG resulted in Beta-catenin Y333

phosphorylation by Src, which apparently lead to interaction between Beta-catenin and pyruvate kinase muscle 2 isoform, causing cyclin D1 transcription (16). Protein kinase B (also called Akt) phosphorylates GSK3B at S9 as well as Beta-catenin at S552, inactivating the former and causing the latter to recruit the signalling protein 14-3-3-zeta (17, 18). Furthermore, phosphorylation by CK1 and GSK3 is reversed by protein phosphatase 2A in the absence of APC, resulting in Wnt-independent protection of cytosolic Beta-catenin (19). Acetylation of Beta-catenin at K49 by cyclic AMP response element binding protein binding protein (CBP) suppresses transcription of c-myc with no effect on cyclin D1 transcription, suggesting that these genes are controlled by Beta-catenin in different PTM states (20). Over 20 enzymes catalyze Beta-catenin PTMs. This diversity of modification suggests the protein participates in a complex and incompletely described series of interactions.

One protein of emerging significance in both Wnt-dependent and Wnt-independent Beta-catenin signalling is p38 mitogen-activated protein kinase (p38). Wnt signalling activates p38 and this effect depends upon Dishevelled as well as the Wnt-activated protein $G\alpha_q$ (21). Subsequently, p38 protects Beta-catenin by phosphorylating GSK3B to render it inactive (21, 22). The intracellular domain of LRP6 is activated through phosphorylation by p38, further controlling the CK1/GSK3B/APC degradation complex (23). Platelet-derived growth factor (PDGF) both activates p38 in VSMCs and disrupts the association of Beta-catenin and cadherin, causing Beta-catenin to migrate to the nucleus, a process associated with VSMC proliferation (24, 25). Two kinases, MAPK kinase 3 (MKK3) and MKK6, directly activate p38 (26). Phosphorylation of p38 by MKK6 is required for muscle differentiation (27). Given the variety of modifications influencing Beta-catenin functionality and the involvement of p38 in Beta-catenin directed-gene expression, the interaction between the two was examined here.

Two experiments employing LC-MS to describe the behaviour of p38 and Beta-catenin are described in this chapter. Potential phosphorylation of Beta-catenin by p38 was explored, yielding strong evidence that such a PTM does not occur and implying that Beta-catenin regulation by p38 is indirect. Secondly, novel Beta-catenin interacting

partners were revealed by sequencing of Beta-catenin immunoprecipitates from the A10 cell line, an undifferentiated VSMC model (28). Components of the transcription machinery were enriched along with Beta-catenin. Eight mediators of gene expression with no previously established connection to Beta-catenin were identified. WDR5 was subsequently shown to interact with Beta-catenin by western blot analysis.

4.2 Methods and Materials

Phosphosite prediction and in vitro protein phosphorylation

The 781 residue full length Beta-catenin sequence was evaluated for potential phosphosites using the computer program NetPhosK (29). Three serine residues were identified as p38 consensus sequences in this manner; S191, S246, and S605 and were targeted in subsequent LC-MS analysis.

The plasmid vector 6xHis pet28a containing the full length Beta-catenin gene was inserted into *E.coli* BL21 to produce histidine-tagged Beta-catenin. The bacterial culture was grown in LB broth at 37^{o}C to an OD_{600} of 1 then vector expression was induced overnight in 500 nM IPTG at 16^{o}C. The recombinant protein was isolated from the crude cell lysate on a nickel column. A truncated form of the known p38 substrate Activating transcription factor 2 (ATF2) purchased from Cell Signalling (Danvers MA, USA), spanning residues 19 to 96 of the full length protein, was used as a positive control (30). *In vitro* phosphorylation was carried out with 2-5 µg of purified substrate, 50-500 ng recombinant p38 (Millipore, Billerica, MA) and 200 µM ATP at 37^{o}C in 50 mM Tris-HCl pH 7.5 with 10 mM MgCl2, 0.1 mM EDTA, 2 mM DTT and 0.01% Brij 35 for 30 minutes. Following treatment the proteins were separated by SDS-PAGE on 10 % acrylamide gels, Coomassie blue stained and bands matching the molecular weight of beta catenin and ATF2 were excised. The molecular weight marker was from Pierce (Rockford IL, USA). Each gel band was treated destained in 60% acetonitrile in water (v/v), washed in 100 mM NH_4HCO_3, reduced in 110 µl of 9 mM dithiothreitol at 60^{o}C in 100 mM NH_4HCO_3 for 15 minutes, brought to ambient temperature then alkylated with

iodoacetamide in the dark for 30 minutes. Each band was then treated with 1 µg of trypsin at 37°C for 16 hours.

Immunoprecipitation

A10 cells (ATTC catalogue number CRL-1476) were grown at 37°C with an atmosphere of 5% CO_2 on 10 cm diameter plates to 80% confluency in high glucose DMEM with 100 U/ml penicillin, 100 µg/ml streptomycin, 1.5 g/l sodium bicarbonate and 10% (v/v) fetal bovine serum. One hour after replacing the spent media, the cells were transfected with the vector pCDNA3 (Invitrogen, Burlington ON, Canada) or pCDNA3 containing full length genes of p38 and pCDNA3 the constitutively active MKK6 mutant MKK6-EE as well as DsRed-pCDNA3 using 25 µL of Lipofectamine 2000 (Life Technologies Inc., Burlington, ON) per plate. The point mutations making MKK6-EE constitutively active are Ser207-Glu and Thr211-Glu. A total of 0.5 µg of each plasmid was used for each plate. Following 2 hours of incubation in plasmid-supplemented media, the plates were washed in phosphate buffered saline (PBS) then incubated in fresh media for 24 hours. Transfection efficiency was assessed based on DsRed fluorescence and was found to be approximately 70%. The pCDNA-transfected and pCDNA3-MKK6-EE/p38C-transfected A10 cells were cultured in triplicate. Cells were harvested by scraping after two washes in PBS and immediately lysed in 5 volumes of 50 mM Tris-HCl, pH 7.6, 150 mM NaCl, 100 mM NaF, 10 mM sodium pyrophosphate, 2 mM EDTA, 1 mM of the phosphatase inhibitor Na_3VO_4, 1 mM phenylmethylsulfonyl fluoride, 1 µg/ml leupeptin, 1 µg/ml aprotinin, 1 µg/ml pepstatin A and 0.5% NP-40 (v/v). Lysates were centrifuged at 14,000 rpm at 4°C for 10 minutes and the supernatants collected.

Anti-Beta-catenin IgG (Cell Signalling) was immobilized at a concentration of 5 µg per 50 µL of ImmunoCruz IP/WB Optima F substrate (Santa Cruz Biotechnology, Dallas, Texas, USA) overnight at 4°C in 500 µL of phosphate buffered saline with 0.1% Tween 20 (v/v). For negative controls, the substrate was incubated with rabbit serum (Sigma Aldrich, Oakville ON, Canada). Unbound material was removed by two 500 µL PBS washes then the resin was suspended in 1 ml of cell lysate overnight at 4°C.

The immunoprecipitation washing procedure was optimized using SDS-PAGE to determine the degree of non-specific binding in the eluates. In one case, the columns were treated with 1 ml of PBS following incubation with the sample. Alternately, two 500 µl washes in lysis buffer followed by two 500 µl washes in lysis buffer without NP-40 were used. Bound proteins were eluted for LC-MS analysis in 500 µl of 0.5 M NH_4OH whereas the washed beads were boiled in 34 mM Tris-HCl pH 6.8, 10% glycerol (v/v), 1% SDS (w/v) with 2.5% 2-mercaptoethanol for SDS-PAGE. Samples analyzed by LC-MS were washed with lysis buffer as described above. Separate immunoprecipitation columns were used for each sample so that a total of six anti-beta catenin eluates were generated. Duplicate rabbit serum controls were generated from separate biological replicates of A10 cells transfected with either vector. Eluates were dried by centrifugal concentration then dissolved in 100 mM NH_4HCO_3 and treated with dithiothreitol then iodoacetamide as in the p38 phosphorylation experiment. The alkylated immunoprecipitates were incubated at 37^{o}C overnight with 0.75 ug of trypsin.

LC-MS/MS

Trypsin digests were dried in a centrifugal concentrator then dissolved in aqueous 0.1% (v/v) formic acid and the peptides resolved by RP nanoflow-HPLC over a 60 minutes 300 nl/minute acetonitrile gradient via a NanoLC-Ultra 2D HPLC pump coupled to a Nanoflex cHiPLC system. The pump and columns were from AB Sciex (Concord ON, Canada). Mobile phase A consisted of 0.1% formic acid in water (v/v) and mobile phase B was 0.1% formic acid in acetonitrile (v/v) (Honeywell, Morristown, NJ, USA). Over the course of each LC-MS/MS analysis, the mobile phase B concentration went from 2% (v/v) at 0 minutes to 35% at 35 minutes, then to 80% from 35.5 to 38.5 minutes back to 2% from 39 to 60 minutes. Both the trap and analytical column were packed with 3 µm diameter C18 stationary phase with 120Å pores. The trap was 0.5 mm long and 200 µm in diameter and the column was 150 mm long and 75 µm in diameter. The column eluate was sampled by an Orbitrap Elite mass spectrometer though an uncoated fused silica capillary having a 10 µm internal diameter mounted into a Flex^{TM} ion source (Thermo Scientific, Waltham, MA).

For examination of Beta-catenin phosphorylation, the acquisition cycle included an Orbitrap MS scan at 60,000 resolution followed by MS/MS scans of the 10 most abundant precursor ions. The orbitrap and the linear ion trap were used to acquire MS/MS data in alternate runs of the kinase-treated samples so two complimentary datasets were produced; one at high resolution and one at high sensitivity. For immunoprecipitate analysis, precursor ion scans were collected on the Orbitrap at a resolution of 30,000 and the linear ion trap was used for MS/MS. All tandem mass spectrometry was carried out at normalized collision energy of 35, activation Q of 0.250, isolation width of 2 Da and a 10 millisecond activation time. Dynamic precursor ion exclusion with a repeat count of 1 and an exclusion mass width of 10 ppm was applied. A 5 second dynamic exclusion duration and 5 second repeat duration were used in data acquisition for the recombinant proteins while these values were 10 seconds and 20 seconds respectively for the immunoprecipitate analyses. The lock mass function was used at 371.101240 Th to maintain mass calibration.

For phosphosite identification, an inclusion list consisting of m/z values for fully tryptic peptide and phosphopeptide ions having the tentative p38 phosphosites assigned by NetPhosK was generated. The sequences SPQMVSAIVR, MLGSPVDSVLFYAITTLHNLLLHQEGAK and GLNTIPLFVQLLYSPIENIQR contained S191, S246, and S605 respectively. Mass shifts from methionine oxidation and glutamine deamidation in conjunction with those due to phosphates were accounted for in the inclusion list. During immunoprecipitate analysis, three blank runs were performed between each pair of samples in order to control carryover, which may otherwise have obscured the origins of those proteins identified. Blank and sample data were acquired under identical instrumental parameters.

MS/MS data were searched against the UniProt *Rattus norvegicus* sequence library (September 18, 2013 edition downloaded from the UniProt Knowledge Base on September 26, 2013) with the SEQUEST algorithm navigated through Proteome Discoverer^{TM} software version 1.3. Search settings were as follows: precursor mass range: 350-5000 Da; minimum peak count: 1; enzyme: trypsin; maximum missed

cleavages: 1; precursor ion tolerance: 10 ppm, product ion tolerance: 0.4 Da, product ion series considered: b and y; dynamic modifications: methionine oxidation (+15.995 Da), phosphorylation (S, T and Y at +79.966 Da), glutamine deamidation (+0.984 Da) and tryptophan dioxidation (+31.990); static modification: cysteine carbamidomethylation (+57.021 Da). Monoisotopic ion masses were considered in the search.

Identification of Beta-catenin associated proteins

Proteins specific to the Beta-catenin immunoprecipitate were determined by processing the SEQUEST outputs. Cutoff scores at the peptide level were determined by automated false discovery rate (FDR) estimation for each sample independently. Lists of proteins identified in each sample at ≥95 and ≥99% peptide confidence were compared separately and those meeting the 95% cutoff were used in subsequent data processing stages. Proteins common to anti-Beta-catenin column eluates and rabbit serum controls were rejected as potential Beta-catenin binding partners.

4.3 Results

Using targeted LC-MS/MS as the detection system in an *in vitro* phosphorylation experiment, no direct effect of p38 on Beta-catenin could be discerned. Known and novel interacting partners co-immunoprecipitated along with Beta-catenin from A10 cell lysates, though p38 activity could not be correlated to any particular interaction.

Beta-catenin is indirectly activated by p38

Only three sites within Beta-catenin match the highly specific p38 substrate consensus sequence serine-proline or threonine-proline, and these are isolated within separate peptides upon trypsin digestion (31). The three candidate phosphosites were present in their unmodified states with high s/n by MS and MS/MS in samples treated with p38 as well as the untreated controls, however the corresponding phosphopeptides were not observed (Figure 4.1). Ions of the S191, S246, and S605 peptides had maximum signals on the order of 200,000 to 300,000,000 counts (Figure 4.1) in the same chromatographic

run. Given the magnitude of signals from the unmodified species, the absence of signals for the corresponding phosphopeptide precursor ions suggested the tentative phosphosites were unmodified by p38. SPQMVSAIVR and GLNTIPLFVQLLYSPIENIQR were reproducibly identified by SEQUEST at 99% confidence whereas MLGSPVDSVLFYAITTLHNLLLHQEGAK was assigned low confidence and confirmed by manual inspection. Similarly, deamidated and oxidized forms of SPQMVSAIVR were present while the phosphorylated equivalents were not. In contrast, the known ATF2 phosphosite was observed in its modified state in p38-treated samples but not in the absence of p38 (Figure 4.1G). Tandem mass spectra from the expected ATF2 phosphopeptide NDSVIVADQTPTPTR, spanning residues 60 to 74 of the canonical sequence were identified by SEQUEST. The peak height of this species was 13% that of the unmodified version, consistent with active p38. The observation that each candidate p38 phosphorylation site remained unmodified under conditions which readily phosphorylate ATF2 indicated that the previously reported activating effect of p38 on Beta-catenin in VSMCs is indirect. By inactivating GSK3B and activating LRP6, p38 maintains elevated cytosolic Beta-catenin concentration, which could account for its increased nuclear function.

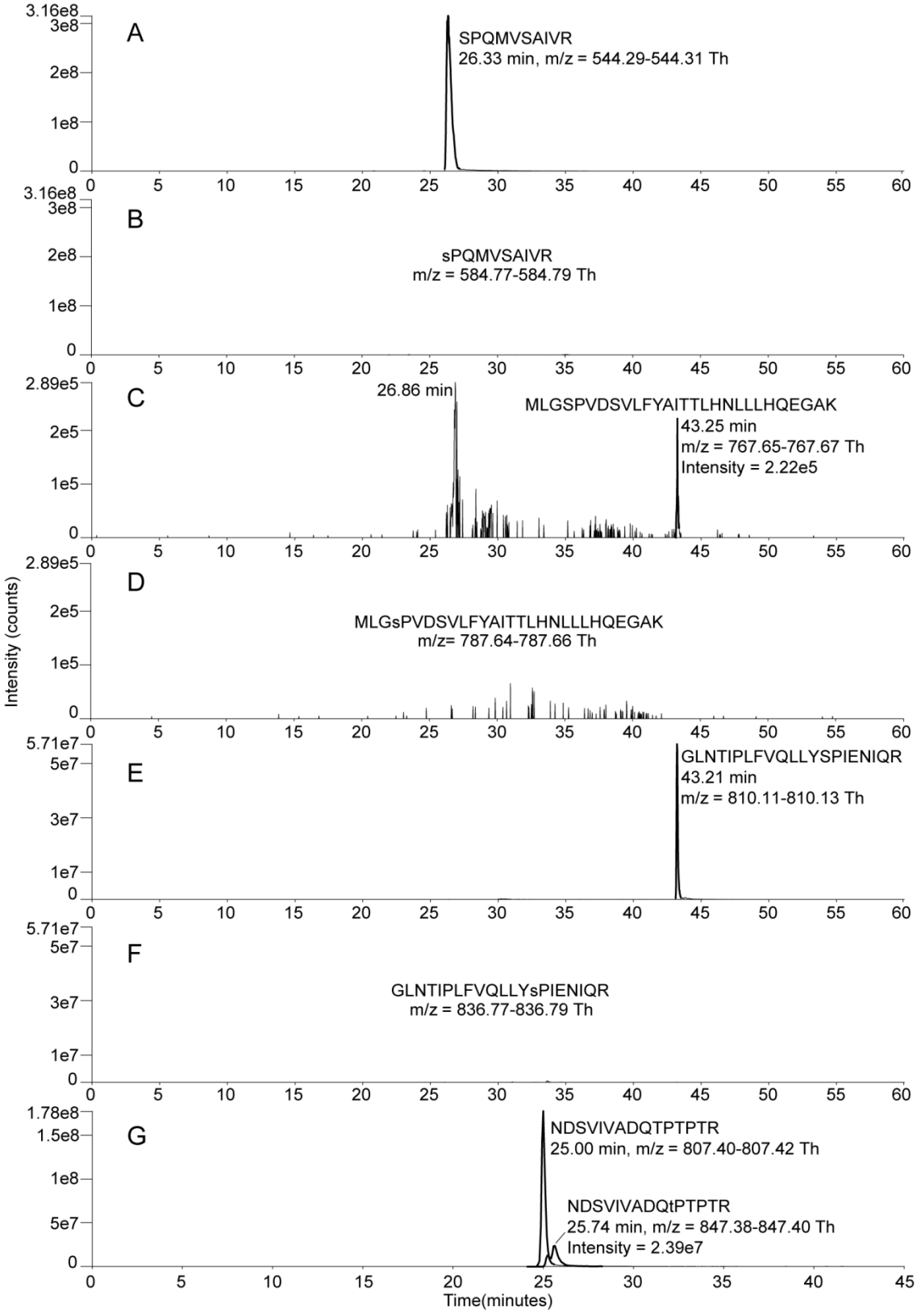

A
SPQMVSAIVR
26.33 min, m/z = 544.29-544.31 Th
B
sPQMVSAIVR
m/z = 584.77-584.79 Th
C
26.86 min
MLGSPVDSVLFYAITTLHNLLLHQEGAK
43.25 min
m/z = 767.65-767.67 Th
Intensity = 2.22e5
D
MLGsPVDSVLFYAITTLHNLLLHQEGAK
m/z= 787.64-787.66 Th
E
GLNTIPLFVQLLYSPIENIQR
43.21 min
m/z = 810.11-810.13 Th
F
GLNTIPLFVQLLYsPIENIQR
m/z = 836.77-836.79 Th
G
NDSVIVADQTPTPTR
25.00 min, m/z = 807.40-807.42 Th
NDSVIVADQtPTPTR
25.74 min, m/z = 847.38-847.40 Th
Intensity = 2.39e7
Intensity (counts)
Time(minutes)

Figure 4.1: Ion chromatograms of tryptic Beta-catenin peptides containing candidate p38 phosphosites and a tryptic ATF2 peptide and phosphopeptide containing known p38 phosphosites. Phosphorylated residues are shown as lower case letters. Doubly protonated SPQMVSAIVR (A) from Beta-catenin gave a prominent chromatographic peak while the corresponding phosphopeptide (B) did not. The quadruply protonated peptide MLGSPVDSVLFYAITTLHNLLLHQEGAK (C) was confirmed by MS/MS at 43.25 minutes while the ion at 26.86 minutes was an isobaric species. No chromatographic peak from phosphorylated MLGSPVDSVLFYAITTLHNLLLHQEGAK was present (D). Doubly protonated GLNTIPLFVQLLYSPIENIQR (E) was observed at 43.21 minutes while the extracted ion chromatogram of the phosphorylated counterpart (F) had no signal. The ATF2 peptide NDSVIVADQTPTPTR is shown in both its unmodified and phosphorylated state in panel G.

The determination of Beta-catenin binding partners

Eluates of immunoaffinity columns functionalized with Beta-catenin antibody and rabbit serum were compared by SDS-PAGE in order to assess their specificity as well as to optimize the stationary phase washing protocol. Eluates from anti-Beta-catenin columns and those of rabbit serum columns were indistinguishable following washes with PBS alone (data not shown), whereas washes with lysis buffer reduced the complexity of the former (Figure 4.2). Lysis buffer washing was therefore applied in preparation for LC-MS. Several prominent bands were common to eluates of both columns, indicating that some degree of contamination from the cell lysate remained in the experimental preparations as well as the rabbit serum controls. For example, two resolved bands exceeding 130 kDa and another between 70 and 100 kDa were found in all samples.

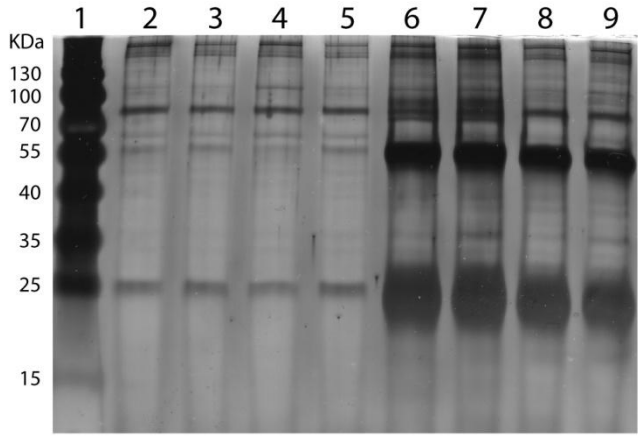


Figure 4.2: SDS-PAGE of immunoprecipitates from A10 cells. Lane 1: molecular weight marker. Lanes 2 to 5 are eluates from anti-Beta-catenin IgG columns. Lanes 6 to 9 are eluates from rabbit serum columns. Lanes 2, 4, 6 and 8 are eluates from A10 cell lysate transfected with the empty vector. Lanes 3, 5, 7 and 9 are eluates from A10 cells transfected with pCDNA3-MKK6-EE/p38.

Proteins identified by MS/MS of rabbit serum column eluates were presumed to be non-specifically bound. Consistent with the SDS-PAGE results, the majority of proteins identified in the rabbit serum control were also found in Beta-catenin column eluates. Given the stochastic nature of ion sampling in shotgun proteomics and differences between the two immunoaffinity columns used, the numbers of peptides assigned to any one protein could be expected to differ among analyses. To address sequence coverage bias introduced by such variation, the presence of a single unique peptide within a single rabbit serum immunoprecipitate was used as the criterion by which the corresponding protein was excluded as a potential Beta-catenin interacting partner. This approach was justified by the absence of Beta-catenin in all rabbit serum column eluates and its appearance in anti-Beta-catenin column eluates from all biological replicates. Protein identifications from anti-Beta-catenin column eluates were confirmed to be sample-specific based on their absence in SEQUEST results from the preceding blank runs and

therefore did not result from HPLC column carryover. Those proteins found only in anti-Beta-catenin column eluates were similarly confirmed to be absent in the preceding blanks.

A total of 528 proteins met the 99% peptide confidence threshold in all experimental samples, whereas 797 were identified at or above 95% confidence from the same data. Similarly 336 and 526 proteins were identified within the rabbit serum controls at 99% and 95% confidence respectively. The dataset generated at 95% confidence contained more peptide identifications per protein and more proteins per sample relative to that obtained at 99% confidence. The 95% confidence interval was used in later stages of data processing because it provided higher sensitivity in peptide detection and therefore superior selectivity. For example, based on the 99% cutoff Nuclear cap binding protein subunit 1 appeared in Beta-catenin immunoprecipitates from both experimental states but not in control immunoprecipitates, however under relaxed tolerance the protein appeared in two of four control samples. This protein was confirmed to be present in the controls by manual interpretation of tandem mass spectra of its peptides. Identical peptides were also observed in the experimental samples with 99% confidence at equivalent retention times, strengthening the credibility of sequence assignments from the controls. Similar findings for the proteins AT rich interactive domain 1A, Fragile X mental retardation protein 1 homolog isoform 2 and La-related protein 4 (LARP4) indicated that the 99% confidence threshold yielded false positive interacting partners. Of the 526 proteins identified at 95% peptide confidence from rabbit serum column eluates, 226 were found in a single sample, indicating they were retained randomly. That more than half of the non-specifically bound fraction was found in multiple control samples could be attributed to their abundance in the lysate and this conclusion was supported by the presence of several isoforms of actin, histones and heterogeneous nuclear ribonucleoproteins in controls from all biological replicates.

Known and novel Beta-catenin binding partners present in immunoprecipitates

Fourteen proteins were exclusively found in Beta-catenin immunoprecipitates, including Beta-catenin itself (Table 4.1). Among these were alpha-catenin and CBP, both of which have been demonstrated to form heteromers with Beta-catenin at the plasma membrane (32). The known Beta-catenin binding transcription activator brahma-related gene 1 (BRG1), which also regulates Frizzled expression, was found in Beta-catenin immunoprecipitates as well as three of the four rabbit serum controls indicating that it was not selectively enriched by the Beta-catenin antibody (33,34). An additional 49 proteins were found in at least two experimental samples as well as at least one control using the 99% peptide confidence threshold and like BRG1 some of these may have been false negative interacting partners; however this was considered acceptable given that the data handling strategy appeared to minimize false positives.

The Contaminant Repository for Affinity Purification (CRAPome), a public database of non-specifically bound proteins in immunoprecipitation experiments, was queried for the gene products found exclusively in Beta-catenin immunoprecipitates (35). Human homologues were referenced since *Rattus norvegicus* data was unavailable. Of the 411 experimental datasets making up the repository, BRG1 appeared in 59 (14%) while Beta-catenin appeared in one (0.2%), suggesting that BRG1 is retained by immunoaffinity media and was therefore prone to false negative interacting partner assignment. Nine of the putative binding partners had fewer entries in the database than BRG1 while 14-3-3 epsilon was most frequently found (45.3%), suggesting its association with Beta-catenin was the most difficult to prove by immunoaffinity chromatography (Table 4.1).

Interrogation of the CRAPome revealed 14-3-3 epsilon to be the most prevalent contaminant of all the species unique to the Beta-catenin immunoprecipitate. Given its apparent tendency toward non-specific binding, the absence of 14-3-3 epsilon from the controls was conspicuous. Nuclear Beta-catenin is shuttled to the cytoplasm by the protein Chibby in combination with 14-3-3 epsilon, which accounts for the observation of the latter (36). Neither of the two Chibby isoforms were found by MS/MS.

Protein name	Abbreviation	Uniprot ID	pCDNA3 replicates	pCDNA3-MKK6/p38 replicates	Subcellular localization	Function	Uniprot ID of Human isoform	CRAPome Occurance (n = 411)
14-3-3 protein epsilon		P62260	2	2	multiple (non-nuclear)	signalling	P62258	186
alpha-catenin		Q5U302	3	3	cytoskeleton	cytoskeletal anchoring	P35221	34
beta-catenin		Q9WU82	3	3	cytoskeleton, cytoplasm, nucleus	cell adhesion, transcriptional regulation	Q9NSA3	1
Cyclic AMP response element binding protein	CBP	Q6JHU9	3	2	cytoplasm, nucleus	transcriptional regulation	Q92793	1
Glutamate rich WD repeat containing protein	GRWD1	Q5XI13	2	2	nucleus	transcriptional regulation	A0MNN5	42
Histone deacetylase 1	HDAC1	Q4QQW4	1	1	nucleus	transcriptional regulation	Q13547	108
KH domain-containing, RNA-binding, signal transduction-associated protein 1	Sam68	Q91V33	3	1	nucleus	transcriptional regulation	Q07666	138
La-related protein 1	LARP1	F1M062	3	3	cytoskeleton, cytoplasm	translational regulation	Q6PKG0	78
PDZ and LIM domain protein 7	LMP-1	Q9Z1Z9	2	1	cytoskeleton, cytoplasm	transcriptional regulation	Q9NR12	16
Poly(A) binding protein, nuclear 1	Pabpn1	G3V7Z8	1	2	nucleus	transcriptional regulation	Q86U42	50
Polymerase (RNA) II (DNA directed) polypeptide C	POLR2C	Q5EB90	1	2	nucleus	transcription	Q6FGR6	24
purine-rich element binding protein alpha	PURA or ssCRE-BP	P86252	2	1	nucleus	transcriptional regulation	Q00577	27
purine-rich element binding protein beta	PURB	Q68A21	2	1	nucleus	transcriptional regulation	Q96QR8	23
WD repeat containing protein 5	WDR5	Q498M4	2	3	nucleus	transcriptional regulation	P51532	59

Table 4.1: Proteins specific to anti-beta catenin immunoprecipitates of A10 lysates observed in two or more biological replicates at or exceeding the 95% peptide confidence limit. Uniprot IDs associated with the sequence assignments are indicated in the third column. The fourth and fifth columns show the number of biological replicates in which each protein was observed from pCDNA3-transfected and pCDNA3-MKK6-EE/p38C-transfected cultures respectively. Uniprot IDs of homologous proteins from *Homo sapiens* used to interrogate the CRAPome are displayed in the eighth column and the frequency of their occurrence is indicated in the ninth column.

The majority of co-purified gene products were previously reported to have nuclear functions, possibly because their association with Beta-catenin affects transcriptional regulation. Consistent with the similarity in polypeptide distribution observed by SDS-PAGE, all proteins which co-immunoprecipitated with Beta-catenin from p38 hyperactivated cells were also found in at least one sample collected from cells with basal p38 activity (Figure 4.2 and Table 4.1). This might imply that binding partners of Beta-catenin were unaffected by p38 activation alone or that the immunoprecipitates contained a mixture of p38-dependent and p38-independent binding partners. Several aspects of the experimental design could account for this trend. The purity of p38 activated cells was limited by transfection efficiency while immunoprecipitation of whole cell extracts allowed the simultaneous capture of nuclear and cytoplasmic Beta-catenin. The liberation of organellar contents during lysis potentially exposed free Beta-catenin to binding partners from which it had been isolated during cell culture. These aspects of sample preparation could have increased the uniformity of immunoprecipitate components from the two biological conditions tested. The isolation of basal state interacting partners from those directed by p38 could be expected to require subcellular fractionation since they recruit Beta-catenin simultaneously *in vivo*. Alpha-catenin interacts with cadherin anchored-Beta-catenin at the plasma membrane and its presence in Beta-catenin pull-downs from p38-stimulated cells could be attributed to a membrane bound Beta-catenin population remaining in the absence of a functional degradation complex (37). This is reasonable given that some Beta-catenin remains bound to cell-cell adherence junctions during Wnt and PDGF signalling (38).

The dataset from which the interacting partners were initially identified was processed for relative quantification in order to distinguish Beta-catenin interacting proteins in the two metabolic states. Extracted ion chromatograms for peptides of alpha-catenin, Beta-catenin and BRG1 were obtained, however the areas under the curves for each precursor ion were insufficiently reproducible among biological replicates to estimate their relative quantity. Quantification by spectral counting was not feasible given the small number of

spectra assigned to each protein. These findings were realistic given the stochastic nature of ESI sampling and the untargeted acquisition settings used (37,24,38).

Biological roles of candidate Beta-catenin binding partners

Data dependent acquisition yielded five unique WDR5 peptides with 99% confidence (Figure 4.3). Though it has not been previously shown to interact directly with Beta-catenin, WDR5 is a known enhancer of canonical Wnt signalling which binds the promoter of the Wnt target gene c-myc as well as that of Wnt1 and whose expression affects nuclear Beta-catenin concentration in osteoblasts (39). It is conceivable that WDR5 recruits Beta-catenin to the nucleus, enhancing WDR5-directed transcription. A FLAG-tagged WDR5 construct was overexpressed in A10 and Beta-catenin immunoprecipitation was performed on the cell lysates. Western blot analysis confirmed the interaction of WDR5 and Beta-catenin (Figure 4.4).

Two unique LARP1 peptides were found with 99% confidence and an additional peptide exceeded 95% confidence (Figure 4.3). LARP1 is an mRNA-binding protein controlling translation which has been found to bind cytoskeletal and cell-cell junction components (40,41). It is therefore conceivable that LARP1 and Beta-catenin associate at the inner face of the cell membrane. LARP1 has been shown to interact with a member or members of the polyadenylate binding protein family, so the presence of Pabpn1 in Beta-catenin immunoprecipitates may be due to a Beta-catenin-LARP1-Pabpn1 complex (40,42,43). RNA polymerase II polypeptide C (POLR2C), of which two unique peptides were sequenced, may also comprise part of this complex since Pabpn1 was shown to bind RNA polymerase II during transcription (44). Pabpn1 has been implicated in myogenesis but little is known about its regulation during this process (45).

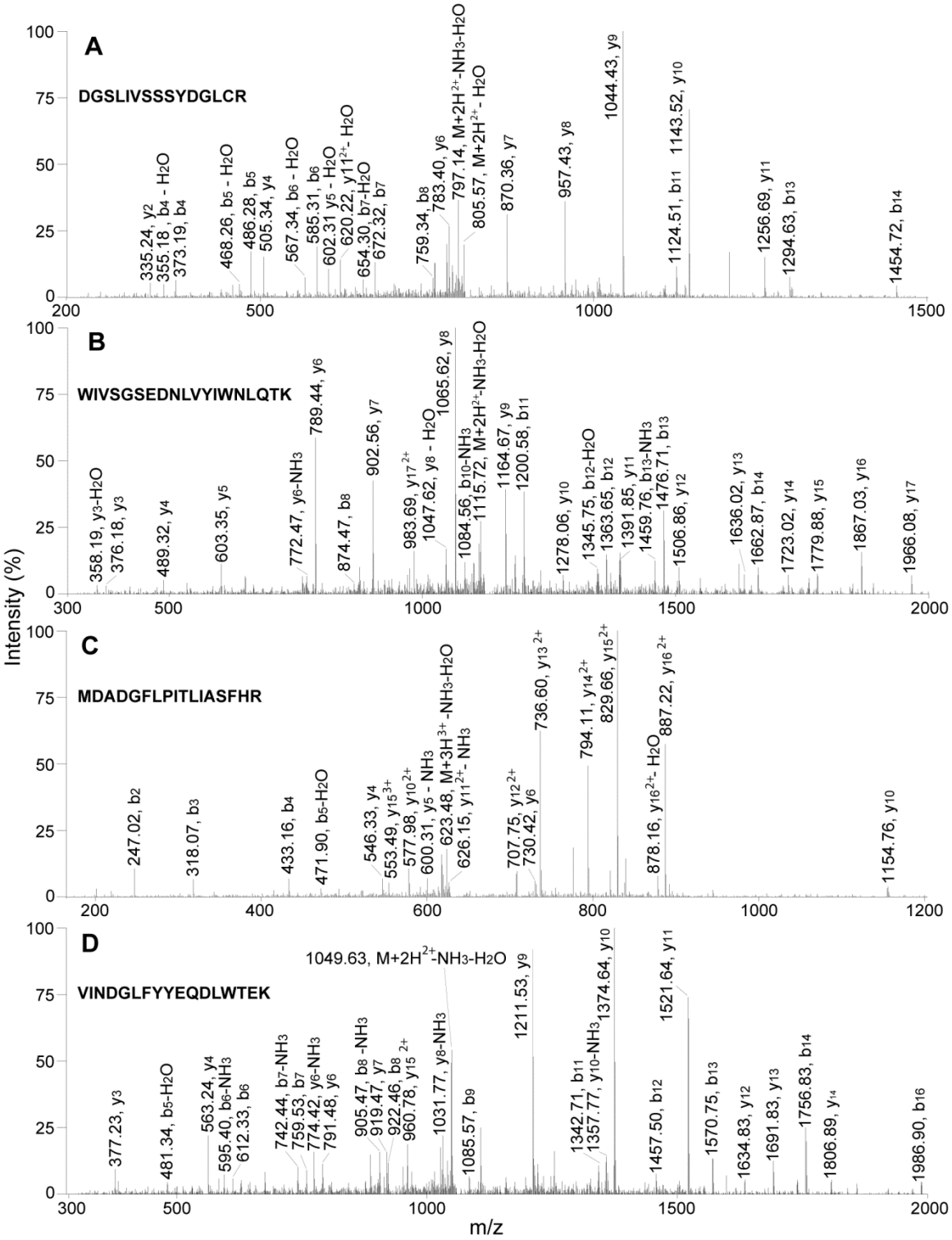

A
DGSLIVSSSYDGLCR
335.24, y2
355.18, b4 - H2O
373.19, b4
468.26, b5 - H2O
486.28, b5
505.34, y4
567.34, b6 - H2O
585.31, b6
602.31 y5 - H2O
620.22, y11 2+ - H2O
654.30, b7-H2O
672.32, b7
759.34, b8
783.40, y6
797.14, M+2H 2+-NH3-H2O
805.57, M+2H 2+ - H2O
870.36, y7
957.43, y8
1044.43, y9
1124.51, b11
1143.52, y10
1256.69, y11
1294.63, b13
1454.72, b14
B
WIVSGSEDNLVYIWNLQTK
358.19, y3-H2O
376.18, y3
489.32, y4
603.35, y5
772.47, y6-NH3
789.44, y6
874.47, b8
902.56, y7
983.69, y17 2+
1047.62, y8 - H2O
1065.62, y8
1084.56, b10-NH3
1115.72, M+2H 2+-NH3-H2O
1164.67, y9
1200.58, b11
1278.06, y10
1345.75, b12-H2O
1363.65, b12
1391.85, y11
1459.76, b13-NH3
1476.71, b13
1506.86, y12
1636.02, y13
1662.87, b14
1723.02, y14
1779.88, y15
1867.03, y16
1966.08, y17
C
MDADGFLPITLIASFHR
247.02, b2
318.07, b3
433.16, b4
471.90, b5-H2O
546.33, y4
553.49, y15 3+
577.98, y10 2+
600.31, y5 - NH3
623.48, M+3H 3+ -NH3-H2O
626.15, y11 2+ - NH3
707.75, y12 2+
730.42, y6
736.60, y13 2+
794.11, y14 2+
829.66, y15 2+
878.16, y16 2+ - H2O
887.22, y16 2+
1154.76, y10
D
VINDGLFYYEQDLWTEK
377.23, y3
481.34, b5-H2O
563.24, y4
595.40, b6-NH3
612.33, b6
742.44, b7-NH3
759.53, b7
774.42, y6-NH3
791.48, y6
905.47, b8 -NH3
919.47, y7
922.46, b8 2+
960.78, y15 2+
1031.77, y8-NH3
1049.63, M+2H 2+-NH3-H2O
1085.57, b9
1211.53, y9
1342.71, b11
1357.77, y10-NH3
1374.64, y10
1457.50, b12
1521.64, y11
1570.75, b13
1634.83, y12
1691.83, y13
1756.83, b14
1806.89, y14
1986.90, b16
Intensity (%)
m/z

Figure 4.3: Representative tandem mass spectra of WDR5 (A and B) and LARP1 (C and D) from Beta-catenin immunoprecipitates. A: Doubly protonated, singly carboxymethylated DGSLIVSSSYDGLCR with a m/z of 814.88202 Th (-0.00088 Th mass error) and a retention time of 39.32 minutes from p38 hyperactive A10 culture. B: Doubly protonated WIVSGSEDNLVYIWNLQTK with a m/z of 1133.08777 Th (0.00659 Th mass error) and a retention time of 49.54 minutes from pCDNA3-transfected A10 culture. C: Triply protonated MDADGFLPITLIASFHR with a m/z of 635.33130 Th (0.00192 Th mass error) and a retention time of 51.60 minutes from pCDNA3-transfected A10 culture. D: Doubly protonated VINDGLFYYEQDLWTEK with a m/z of 1067.01453 (0.0021 Th mass error) and a retention time of 47.29 minutes. Sequence assignments for the spectra shown were given more than 99% confidence.

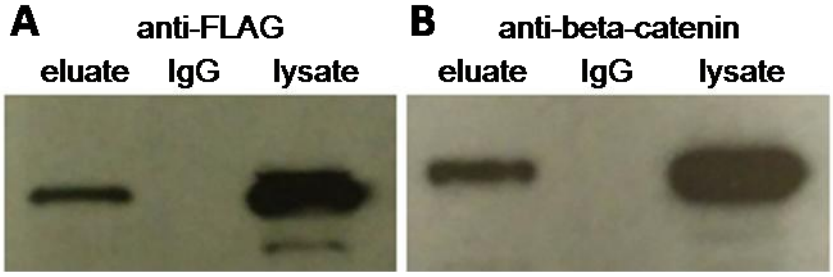


Figure 4.4: Western blot analysis of Beta-catenin immunoprecipitates from FLAG-WDR5-overexpressing A10 lysates. Blots with anti-FLAG (A) and anti-Beta-catenin (B) as the primary antibodies are shown. Anti-beta catenin column eluates (eluate) appear in the left lanes, rabbit serum control column eluates (IgG) appear in the center lanes and crude cell lysates (lysate) appear in the right lanes of each blot. The bands shown in panel A appeared at approximately 40 kDa while those shown in panel B appeared at approximately 90 kDa.

An intricate relationship between histone deacetylase 1 (HDAC1), CBP and Beta-catenin emerging through the results of parallel studies could account for their simultaneous observation in Beta-catenin immunoprecipitates. CBP modulates the Wnt pathway by acetylating Beta-catenin and it is conceivable that HDAC1 reverses this reaction since

histone deacetylase inhibition results in CBP-dependent Beta-catenin mediated transcription (20, 46). Alternately, the opposing roles of CBP and HDAC1 in transcriptional control may be manifested in their reactions with other proteins while Beta-catenin serves to enhance the effect of each separately. HDAC1 is known to bind reptin, which itself complexes with Beta-catenin, and all three coordinate to negatively regulate expression of the tumor metastasis repressor p83 (47, 48). A single unique reptin peptide was observed with more than 99% confidence in a single beta catenin immunoprecipitate from one biological replicate but not in any of the controls and its presence could account for the appearance of HDAC1 in the immunoprecipitates. Detection of KH domain-containing RNA-binding signal transduction-associated protein 1 (Sam68) provided further evidence that Beta-catenin and CBP function in concert. CBP and Sam68 colocalize in the nucleus and their interaction stimulates Sam68-regulated transcription (49). Beta-catenin and Sam68 have both been co-purified with protein tyrosine kinase 6, of which they are both substrates (50).

In the nucleus, CBP is coupled to the DNA by its binding partner cyclic AMP response element binding protein (CREB) (51). CREB was not found in the immunoprecipitates though two similar DNA binding proteins, PURA and PURB, were identified by two unique peptides above 99% confidence. PURA, also known as single-stranded CRE binding protein (ssCRE-BP), has affinity for the same DNA element as CREB and both are activated by calmodulin so each may connect Beta-catenin-containing transcription machinery to their respective target genes in response to calcium or other stimuli (52, 53). PURA and PURB form a heterodimer, which could explain their simultaneous observation (54). The target genes and primary structures of PURA/PURB and CREB differ. Only seven amino acids in the region spanning residues 66 to 84 of PURA are also present in CREB and this is the region of highest sequence homology between the two. CREB residues 256-299 and PURB residues 103-141 have the highest homology at 36% sequence identity. CREB residues 130 to 141 including the calmodulin-dependent phosphorylation site at Serine-133 were demonstrated to affect CBP binding and are absent in PURA and PURB (55). Given the dissimilarity of their primary structures, it is

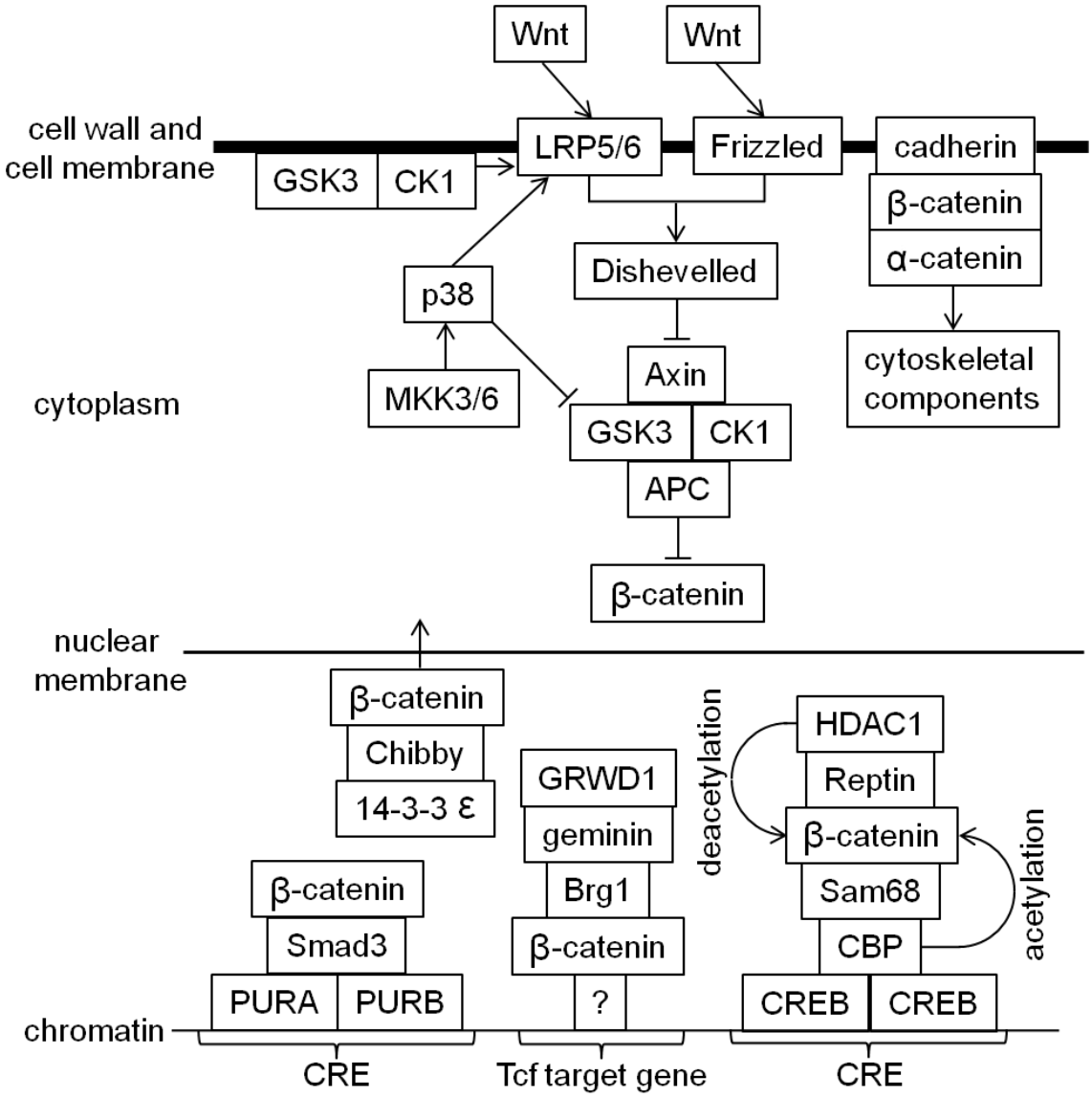
likely that PURA/PURB connects to the transcription machinery through an entirely different motif than that employed by CREB. There is subtle evidence to suggest that Smad3 and PURA/PURB associate in myoblasts upon transforming growth factor beta 1 stimulation while the same growth factor has been demonstrated to induce interaction between Beta-catenin and Smad3 necessary for epithelial-mesenchymal transition and myogenensis (56, 57, 58). It is therefore possible that Smad3 connects Beta-catenin to PURA/PURB in a multisubunit complex, though Smad3 was absent from the SEQUEST output.

Little is known about the function of GRWD1, however there is indirect evidence for its involvement with Beta-catenin. Initially shown to affect ribosome biosynthesis and differentiation, GRWD1 was also found in a protein complex with the cell cycle regulators Cdc10-dependent transcript 1 (Cdt1) and geminin (59, 60). In embryonic stem cells and early embryos, geminin overexpression decreased cadherin expression and increased Wnt pathway activation through an unknown mechanism in which Beta-catenin may participate (61, 62). BRG1, which itself enhances TCF target gene transcription, interacts directly with geminin, providing a link to Beta-catenin (63, 33). It is therefore possible that GRWD1 contributes to a geminin-BRG1-Beta-catenin functionality. Neither Cdt1 nor geminin were sequenced by MS/MS of the A10 immunoprecipitates.

PDZ and LIM domain protein 7 (LMP-1) expression is associated with bone formation *in vivo* and causes the C2C12 myoblast cell line to differentiate into osteoblasts (64). No relationship between Wnt signalling and LMP-1 has been established however the differentiation of progenitor cells into osteoblasts and the maintenance of osteogenic lineage are dependent on Wnt signalling through Beta-catenin (65, 66). In the context of the known role of LMP-1, the data presented here are consistent with Beta-catenin enhancing its function during osteogenic stimuli.

Figure 4.5: Model of Beta-catenin function including known roles of p38 in Wnt signalling. Pointed arrows indicate activation and blunt ended arrows indicate inactivation. Three hypothetical DNA-binding nuclear protein complexes containing

Beta-catenin are shown, two associating with cyclic AMP response elements (CRE) and one associating with a Tcf target gene. Possible opposing roles of CBP and HDAC1 in controlling the CREB complex through Beta-catenin acetylation and deacetylation are shown.



4.4 Conclusion

Both p38 and Beta-catenin are essential to muscle development, however the actions of each in facilitating differentiation and proliferation remain incompletely described.

Direct phosphorylation of Beta-catenin by p38 was investigated here. The apparent failure of the active kinase to modify the only Beta-catenin residues matching its highly specific substrate sequence motif indicated that p38 affects Beta-catenin localization and Wnt signalling through other substrates such as LRP6 and GSK3B. Though Beta-catenin-associated proteins in a cellular model of high p38 activity persisted without p38 stimulation, novel interacting partners were indicated.

The co-purification of proteins known to bind other Beta-catenin-binding species by immunoprecipitation could be accounted for by multisubunit transcription complexes. HDAC1, PURA and PURB are each associated with Beta-catenin through an additional protein. Nuclear proteins identified here could represent parts of at least two distinct complexes, one binding the DNA through CREB and another through PURA/PURB. The implication of Beta-catenin in the regulation of PURA/PURB target genes is without precedent. A possible novel contributor to the canonical Wnt signalling pathway, GRWD1 was also identified. Consequences of the associations shown to the fate of muscle progenitors are unclear and require further *in vitro* and *in vivo* study to elucidate.

4.5 Contributors

This author contributed to the experimental design, silver staining of SDS-PAGE gels, LC-MS sample preparation, LC-MS analysis as well as data interpretation. The remainder of the experimental work, including cloning, cell culture, *in vitro* protein phosphorylation, immunoprecipitation and Western blotting were performed by Saviz Ehyai (B.Sc.) as part of his ongoing Ph.D dissertation work.

Chapter 5

Identification of isobaric peptide ions and product ions from sequence information: Implications for selective reaction monitoring

5.1 Introduction

Selected reaction monitoring (SRM) is a sensitive technique allowing the quantification of multiple peptide analytes in complex biological matrices including tissue homogenates, serum and plasma (1-3). SRM has drawbacks arising from two phenomena; transitions selected to represent an analyte are sometimes produced by other species and transition signal intensity is sometimes distorted by matrix effects. Both problems are significant in proteomic samples containing thousands of polypeptides and other biomolecules as well as contaminants of sample preparation. Around half of all transitions may be insufficiently reproducible in their signal abundance to permit accurate quantification, with some being more reproducible than others from the same peptide (4). The limited resolution of quadrupoles contributes to ambiguity in quantification since signals from the analyte and contaminants cannot be isolated, however reducing transmission window width to increase specificity simultaneously decreases sensitivity. Practically all transitions are subject to isobaric interference. It follows that SRM specificity increases with the number of transitions-per-peptide, though no method for establishing an acceptable minimum number has been established (5).

Even isobaric species with no sequence homology to the analyte can cause interference in an SRM assay. One example of this occurred during optimization of an assay for chaperonin 10 in human serum. Despite having no major product ions in common, the

non-tryptic albumin peptide LVAASQAALGL was associated with four peaks corresponding to transitions for the tryptic chaperonin 10 peptide GGEIQPVSVK and having the same relative abundance (Figure 5.1). Though they had distinct retention times over a 90 minute separation, the difference would be reduced by compression of the chromatographic gradient, an important consideration in minimizing run time.

Enhanced product ion scanning, an MS/MS mode in which fragments are accumulated in the ion trap, automatically triggered when the transitions exceeded a user defined intensity threshold allowed both the expected chaperonin 10 peptide and the albumin peptide to be identified by a Mascot search. It should be noted that incorporating MS/MS scans into an SRM method for analyte identity verification reduces the number of data points in the SRM experiment so that peak integration suffers. MS/MS spectra of the two species were highly dissimilar. The product ions of the albumin peptide isobaric to those of the chaperonin peptide transitions could only be attributed to low level noise. This implies that peptide mass alone can cause isobaric interference in SRM as well as SIM. Being one of the most abundant plasma proteins, albumin could remain in concentrations equivalent to those of a low abundance analyte even in a purified blood sample. Transition selection may therefore be assisted by knowledge of isobaric peptides and their proteins of origin.

Every proteomic sample contains a finite number of sequences determined by the biological material from which it was derived and how it was processed so that protein sequence databases can in theory be used to describe all polypeptides contributing to isobaric interference. A variety of tools supporting the selection of specific transitions for SRM assay development are available. One optimization approach is to find all unique hypothetical transitions within a whole proteome. This entails calculating the precursor and product ion m/z values of all proteotypic peptides within a sequence library then comparing them within some tolerance and eliminating any matches (6). The concept was extended to include retention time and spectral library information (7). With these approaches the majority of peptides within selected proteomes were associated with sets of two to three transitions which were unique identifiers; however the identities of

interfering peptides were not discussed in detail. While these studies probed the proteome for transition sets devoid interference, we worked from the assumption that the end user of SRM has specific proteins of interest and must accept some degree of interference which can nevertheless be minimized by avoiding the effects of prominent endogenous matrix proteins as well as those transitions which are generally redundant.

The determination of transition redundancy using a novel automated algorithm (SRMass), designed by this author and written by Olga Krakovska at York University, is described below. The SRMass program is similar in function to SRMCollider in that the modifications considered, precursor and product ion mass tolerance and proteomic library are all user defined (7). Relationships between redundancy, mass deviation and peptide sequence were examined using prostate-specific antigen (PSA), a current biomarker of prostate disease, as a model protein. The focus on specific peptides revealed distinct trends in redundancy. For example, the number of isobaric transitions increases at different rates with expanding precursor ion mass tolerance for different product ions. There was a linear relationship between precursor resolution setting and the number of potential interfering species for many product ions however no clear relationship between product ion size and transition specificity could be discerned. These observations have implications for the transition selection process and practical mass transmission ranges of SRM assays.

5.2 Methods

Isobaric interference was evaluated by determining the number of peptides within a large sequence database which were isobaric to a given analyte which also produced fragment ions isobaric to those of the analyte. This approach provides perspective on the extent and nature of SRM interference caused by the biological background for given a protein of interest. Precursor/product ion m/z pairs corresponding to tryptic peptides of human PSA, a prostate health biomarker used in the diagnosis of prostate cancer were queried against a RefSeq non-redundant human proteome FASTA format library (downloaded September 20, 2010) representing 19,127,252 amino acids in 32208 distinct proteins.

Figure 5.2 depicts the SRMass user interface. The following search terms are used: Mass mode (calculates either monoisotopic or average precursor and product ion masses from the FASTA database), precursor ion m/z (in Th), precursor ion m/z tolerance (twice the absolute mass deviation limit for the precursor ion), product ion m/z (in Th), product ion m/z tolerance, possible modifications, and the maximum number of modifications on any one peptide (0-20). Full tryptic cleavage is the only allowed proteolytic condition. Precursor ions with charges of 1 to 6 and product ions with charges of 1 to 2 are considered in the search.

In order to test the effect of the precursor ion mass window on selectivity, otherwise identical searches were conducted at precursor mass tolerances between 0 and 2 Th, a range selected to span the practical resolution settings used in SRM acquisition. The following constant parameters were used: product ion tolerance, 0.1 Da; mass mode, monoisotopic; modifications, cysteine carbamidomethylation and methionine oxidation; maximum number of modifications considered, 3. The input data set included precursor/product ion pairs from nine fully tryptic and one semi-tryptic peptide, representing 122 of the 261 amino acids from PSA isoform 1 preproprotein (GI 4502173). The semi-tryptic sequence WIKDTIVANP covers the protein C-terminus. Transitions for the peptide LQCVDLHVISNDVCAQVHPQK carbamidomethylated at position 14 but unmodified at the third cysteine were searched as were transitions for the doubly carmidomethylated version. All sequences were between 6 and 21 residues long. The product ion masses used corresponded to all y-ions and in some cases b-ions from every peptide with those containing fewer than three amino acid residues excluded. The query list totalled 132 precursor/product ion pairs.

5.3 Results

Trends in transition redundancy were inferred from results generated using the relatively small input set of precursor/product ion pairs. The universality of isobaric interference was apparent in that virtually all precursor/product ion pairs tested had tens to hundreds of database matches at m/z widths typical of triple quadrupole instruments (Figures 5.3

and 5.4). Even mass windows associated with decreased sensitivity resulted in tens to hundreds of matches for most transitions. Transmission ranges below 1 Da are associated with signal attenuation, however at precursor and product ion mass windows of 0.1 Da all transitions tested matched non-sequence precursor/product ion pairs from the FASTA database (Figures 5.3 and 5.4 and Table 5.1). Given this result, the mass accuracy required to identify analyte ions with certainty from a few transitions is beyond the capability of a triple quadrupole but possible with a contemporary Qq-TOF.

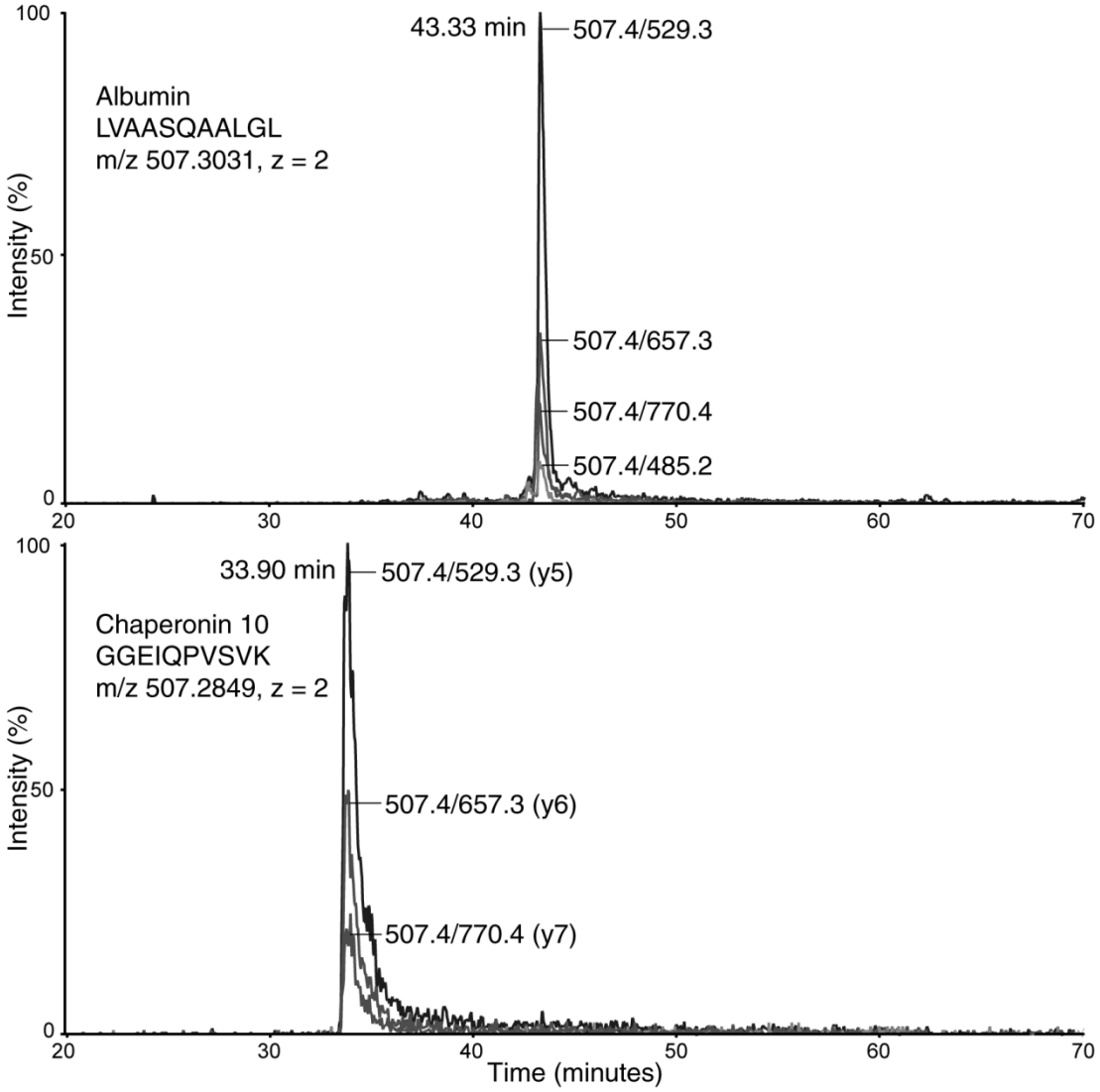


Figure 5.1: Isobaric interference in an SRM assay. Transitions designed to be specific for the unique, proteotypic chaperonin 10 peptide GGEIQPVSVK also resulted in signals from the non-tryptic albumin peptide LVAASQAALGL, a contaminant originating in the sample. Top panel: Signals from the albumin peptide. Bottom panel: Signals from the chaperonin 10 peptide. The transition m/z 507.4/485.2 corresponds to a b_5 ion from GGEIQPVSVK. The data was acquired on a 4000QTRAP.

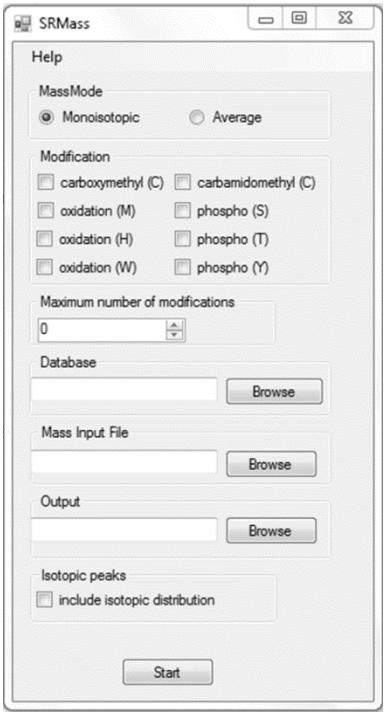


Figure 5.2: The SRMass user interface.

A complex relationship between product ion identity, allowable m/z deviation and transition specificity was evident in results for the PSA peptides. At precursor ion and product ion tolerances of 0.01 and 0.1 Da respectively, the average number of matches to the database for the transitions tested was less than 5, with many returning no hits. Beyond a precursor tolerance of 1 Da, almost all transitions exceeded 100 matches with many having hundreds (Figures 5.3 and 5.4). This showed that a set of peptide ions can produce many isobaric product ions and neither high precursor ion tolerance nor high

product ion tolerance alone can be expected to produce unique transitions. Though SRMass does not report peptide ion matches lacking matching product ions, searches against y_1 ions can provide similar data because a high proportion of isobaric peptides contain the same C-terminal residue as the query. The y_1 ions of SVILLGR and carbamidomethylated FMLCAGR returned 5220 and 4136 results respectively at 1.0 Da precursor and 0.1 Da product tolerances, values one to two orders of magnitude higher than for transitions including larger product ions. This result is consistent with SRM being more selective than SIM but also indicates that a large percentage of all isobaric peptides share isobaric b-ions or y-ions.

For some of the peptides tested, a small number of product ions had substantially higher m/z redundancy than others of the same ion series at all mass tolerances tested. For example, b_6 of doubly protonated SVILLGR had 407 matches at precursor and product ion tolerances of 0.1 Da each, while the average number of matches for y_3 to y_6 and b_3 to b_5 from the peptide was 64.4 under this condition (Figure 5.5). This suggests that among transitions for a given peptide, some are significantly more prone to misinterpretation than others. Even b_2 and y_2 from SVILLGR had fewer matches than b_6 at the aforementioned tolerances, with 111 and 87 hits respectively. It could be reasoned that a longer fragment ion would share its m/z value with fewer species than would a shorter one, since length tends to increase sequence uniqueness. Large product ions may be favoured in transition design for this reason. Findings for HSQPWQVLVASR and VTKFMLCAGR support this line of reasoning; with the slopes of the redundancy plots decreasing from y_3 to y_{11} in the former and from y_3 to y_6 and remaining low for y_7, y_8 and y_9 in the latter (Figure 5.3A and B). For doubly charged FLRPGDDSSHDLMLLR, y_3 and y_4 had the most matches, however these ions contain two leucine residues and were therefore given to interference from both leucine-containing and isoleucine-containing fragments (Figure 5.5). In contrast, large fragments from certain peptides had more isobaric ions than did smaller fragments in the same series, suggesting the absence of a direct correlation between product ion mass and specificity. Notably, doubly charged y_{20} from doubly carbamidomethylated LQCVDLHVISNDVCAQVHPQK at m/z 1174.0677

had more than eight times as many matches as the average for the remaining 18 ions investigated from this peptide at low mass tolerances (Figure 5.5). Doubly protonated y_{13} at m/z 748.3697 was the next most redundant product ion from this peptide with 24 hits whereas y_5 at m/z 608.3515 returned none. The y_8 ion of IVGGWECEK, the largest in the y-ion series, had 113 transition matches at precursor/product ion mass deviation limits of 0.1 Da whereas the series from y_3 to y_7 had 24.8 matches on average (Figure 5.3C). The SVILLGR b_6 ion discussed above shows this also occurs with short sequences. One interpretation of these results is that a few product ions from a given series are outliers with respect to the probability of interference. As such, the product ion containing the most residues will not necessarily make the most accurate transition.

For most precursor/product ion masses tested, the number of database matches increased linearly with the precursor mass window. All queried transitions of HSQPWQVLVASR and the semi-tryptic VTKFMLCAGR obeyed this rule with redundancy increasing by as many as 466 and as few as 10 matches per Da in the case of y_{11} and y_3 from the former peptide (Figure 5.3A and B). Two alternate relationships between precursor tolerance and specificity were also observed. Matches to the y_5 and y_8 transitions of IVGGWECEK as well as the y_{11} transition of LSEPAELTDAVK increased according to natural logarithmic functions due to large differences in the number of matches over short precursor tolerance ranges, especially around 0.01 and 0.1 Da (Figure 5.3 C and D). The y_6 and b_6 transitions of carbamidomethylated FMLCAGR showed similar trends. In comparison to those with a simple linear relationship, transitions of this type are more susceptible to noise at narrow transmission window widths and are therefore likely to be the least reliable in quantification. Findings at higher tolerances support this conclusion, since each logarithmic functions lay above most others from 0.5 Da and beyond. Interestingly, the FMLCAGR b_6 ion was by far the most redundant at tolerances of 0.01 and 0.1 Da while the y_6 ion was most redundant from 0.5 to 1.5 Da, demonstrating the potential for inversions of transition specificity at higher sensitivity (Figure 5.4). The slope of the redundancy curve for transition y_3 from IVGGWECEK was linear from 0.01 to 1.5 Da but increased with precursor tolerance beyond that point in a manner best

described by a polynomial. Examples of each of the three function types were found among the six transitions from IVGGWECEK queried; with y_5 and y_8 following natural logarithmic patterns and the remaining three being linear (Figure 5.3C). Collectively, these data suggest that trends among transitions from the same peptide are independent of one another. For the SRM experiment this could translate to highly variable susceptibilities to noise among related transitions at increasing sensitivity. Using a single precursor transmission window for all transitions would have different consequences for the reliability of each one. Furthermore, those transitions with many matches at low m/z tolerance, such as y_8 from IVGGWECEK y_{11} from LSEPAELTDAVK and b_6 from FMLCAGR also have above average number of hits at m/z widths more realistic for SRM acquisition.

The sequences discussed above all have tryptic cleavage sites at their termini and are without unusual modifications. Two exceptional sequences were investigated with SRMass in order to evaluate the effect of m/z bias in the input. WIKDTIVANP with an internal lysine residue and lacking a C-terminal residue and LQCVDLHVISNDVCAQVHPQK with incomplete carbamidomethylation, gave results similar to other peptides. Searching the sequence library for fully tryptic peptides against the m/z values of a non-tryptic peptide might be expected to result in a low number of matches, especially to y-ions lacking lysine, but y_3 to y_7 of WIKDTIVANP had an average of 222 hits at 1 Da precursor and 0.1 Da product ion tolerances while y_8 and y_9 had 388 and 46 respectively. Similarly, introducing a mass shift of -57 Da to a cysteine residue in the query relative to those in the database might decrease the number of matches. The number of matches per transition from doubly and singly modified LQCVDLHVISNDVCAQVHPQK differed, however the values were on the same order of magnitude. For the 18 identical y-ions the two species have in common, the largest difference in redundancy at precursor and product ion tolerances of 0.1 Da was that of y_9, for which the singly alkylated version received 18 matches where the doubly alkylated version had 9. The average numbers of hits among the y-ions queried for both variants of LQCVDLHVISNDVCAQVHPQK were comparable at all tolerances tested. From

searches of these unusual sequences it is apparent that any proteomic library of sufficient size contains corresponding transitions from tryptic peptides.

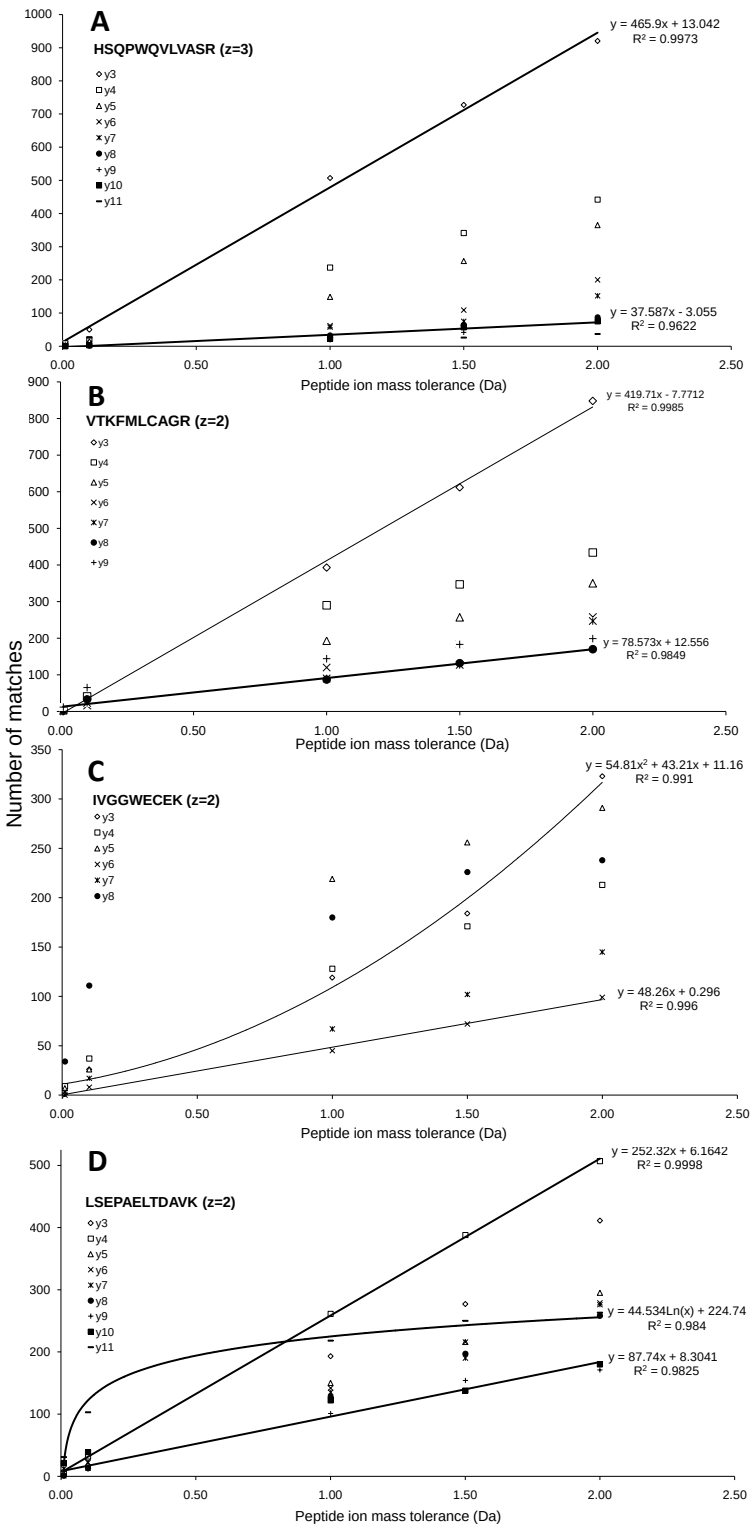

A
HSQPWQVLVASR (z=3)
y = 465.9x + 13.042
R² = 0.9973
y = 37.587x - 3.055
R² = 0.9622
B
VTKFMLCAGR (z=2)
y = 419.71x - 7.7712
R² = 0.9985
y = 78.573x + 12.556
R² = 0.9849
C
IVGGWECEK (z=2)
y = 54.81x² + 43.21x + 11.16
R² = 0.991
y = 48.26x + 0.296
R² = 0.996
D
LSEPAELTDAVK (z=2)
y = 252.32x + 6.1642
R² = 0.9998
y = 44.534Ln(x) + 224.74
R² = 0.984
y = 87.74x + 8.3041
R² = 0.9825
Number of matches
Peptide ion mass tolerance (Da)

Figure 5.3: The number of precursor/product ion pairs within a non-redundant human protein database matching transitions as a function of precursor ion mass tolerance. The human FASTA sequence database was queried with transitions for PSA peptides at a constant product ion mass tolerance limit of 0.1 Da and precursor ion mass ranges from 0 to 2 Da. A: Triply protonated HSQPWQVLVASR. B: Doubly protonated semi-tryptic VTKFMLCAGR. C: Doubly protonated IVGGWECEK. D: Doubly protonated LSEPAELTDAVK. All cysteine residue masses were adjusted for carbamidomethylation.

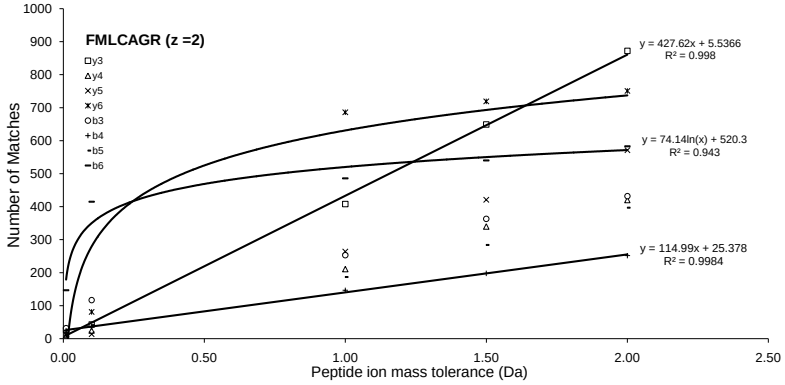


Figure 5.4: Relationships between precursor ion tolerance and specificity among transitions of carbamidomethylated FMLCAGR.

An analyte sharing single transitions with several unrelated peptides might lead to inconsistent quantification and necessitate a change in the set of product ions monitored. A peptide unrelated to the analyte but sharing multiple transitions with it presents a different problem, namely that the alternate ion chromatogram could be mistaken for that of the analyte and integrated in its place. Recognition of such cases could improve the efficiency of method development, as problematic peptide analytes could then be avoided altogether.

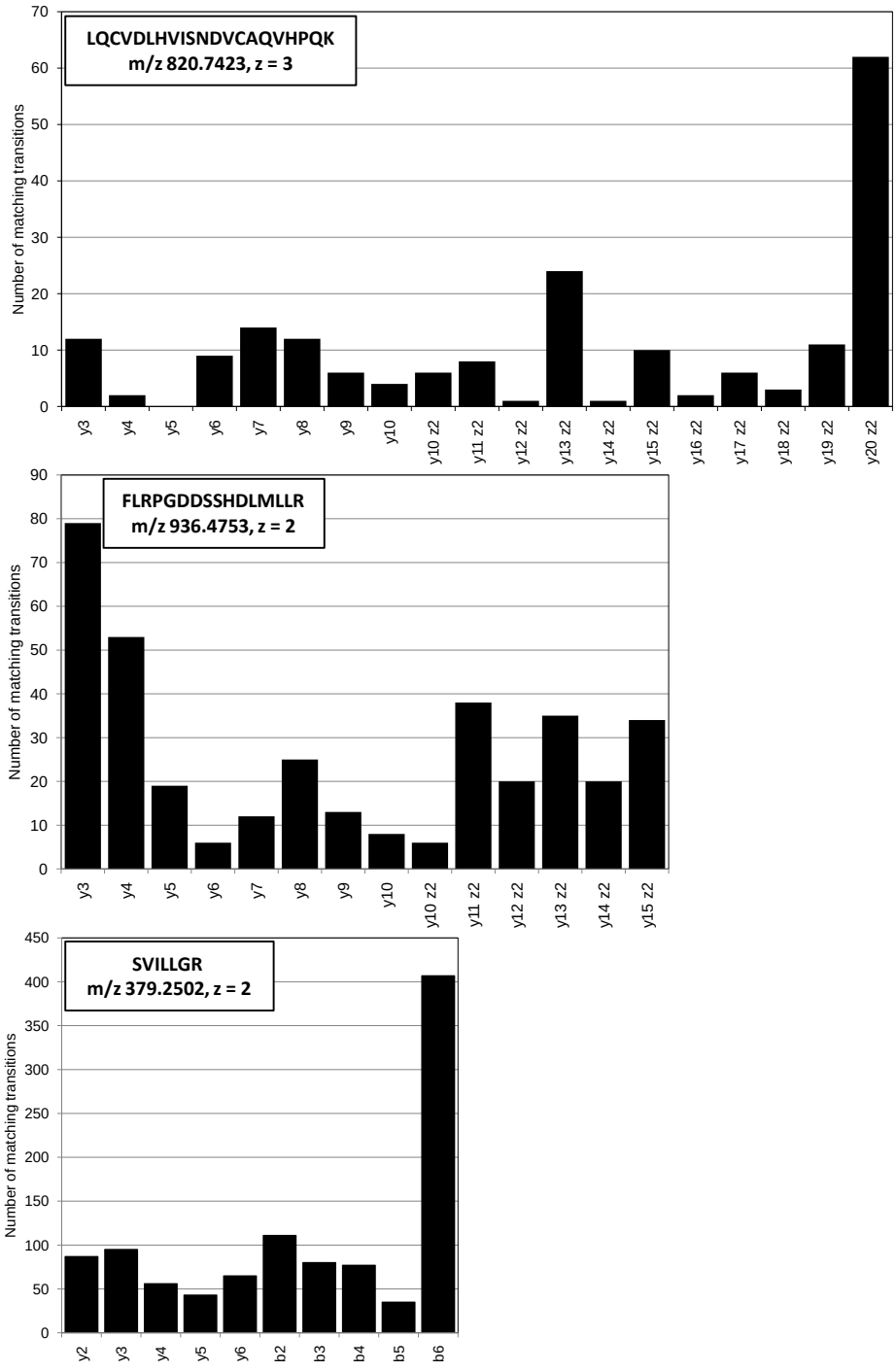


Figure 5.5: Transition redundancy for three tryptic PSA peptide ions. The number of precursor/product ion pairs within the human proteome matching m/z values of each transition is plotted. Matches to peptide sequences derived from PSA isoforms within the

database were excluded from the results shown. Precursor ion mass tolerance and product ion mass tolerance were both 0.1 Da. All cysteine residue masses were adjusted for carbamidomethylation.

SRMass results for three PSA peptides were interrogated for matches having more than one isobaric product ion. The maximum allowable m/z deviations were set to $\pm$0.05 Th to mimic instances in which the contaminant and the analyte could not be clearly discriminated by MS/MS scans at quadrupole resolution. Under these conditions IVGGWECEK shared between two and three transitions with 11 tryptic peptides, LSEPAELTDAVK shared between two and four transitions with 29 peptides and HSQPWQVLVASR from two to four transitions with 15 peptides from other proteins. Conserved domains and homologous proteins were underrepresented in this group; rather the matches had few sequence elements in common with the queries and came from proteins of various functions. Thus the basis of the phenomenon for these three peptides was not protein families but the distribution of amino acids throughout the human proteome. The differences in mass between query and match were typically well within tolerances applied to both the precursor and product ion search space (Table 5.1). Larger query sequences also returned multiple hits from single peptides with low mass deviations, for instance triply charged TGIAFIYAVFNLVLTTVMITVVHER from Ceramide cholinephosphotransferase 2 and the doubly charged PSA ion FLRPGDDSSHDLMLLR were within 0.05 Da of each other as were their y_8 and y_{15}^{2+} ions. The obvious difference in the masses of these two sequences might be overlooked during quantification since charge state is not assessed in the SRM assay. Arginine and lysine differ in mass by 28 Da and therefore contribute different masses to y-ions, however, regardless of the C-terminal residue of the query sequences, matches to multiple y-ions from peptides with both C-terminal arginine and lysine were found (Table 5.1). Aside from showing the discrepancy between sequence and mass, these data indicate that SRM at low resolution is not sufficiently specific for certain peptides.

Query precursor and product ions	Matching fragment m/z (Th)	Matching sequence	Charge of matching fragment	Matching fragment type	Matching peptide location start	Matching peptide location end	Precursor m/z deviation (Th)	Product m/z deviation (Th)
HSQPWQVLVASR, 2+		**EQPLPAQDPVSVK, 2+**						
y3	333.2132	SVK	1	y3	100	112	-0.0002	0.0232
y4	432.2817	VSVK	1	y4	100	112	-0.0002	0.0217
y6	644.3614	DPVSVK	1	y6	100	112	-0.0002	-0.0486
y7	772.4199	QDPVSVK	1	y7	100	112	-0.0002	-0.0501
IVGGWECEK, 2+		**DPNEWAMAK, 2+**						
y3	436.2224	AMAK	1	y4	830	838	-0.0182	0.0364
y5	751.3443	EWAMAK	1	y6	830	838	-0.0182	0.0363
y7	865.3873	NEWAMAK	1	y7	830	838	-0.0182	0.0364
LSEPAELTDAVK, 2+		**EALLAEMGVAIR, 2+**						
y6	646.3705	MGVAIR	1	y6	460	471	0.0147	-0.0095
y7	775.4131	EMGVAIR	1	y7	460	471	0.0147	-0.0069
y8	846.4502	AEMGVAIR	1	y8	460	471	0.0147	-0.0098
y10	1072.6183	LLAEMGVAIR	1	y10	460	471	0.0147	0.0683

Table 5.1: Pairs of peptides having low sequence homology and multiple isobaric transitions. Matching peptide ions shown for each PSA sequence included EQPLPAQDPVSVK from homeobox protein DLX-3, DPNEWAMAK from prune homolog 2 protein, and EALLAEMGVAIR (m/z 636.8527) from kinesin-like protein KIF1B isoform alpha. The positions of the first and last amino acids of each peptide within their respective full length proteins are shown in the sixth and seventh columns.

5.4 Conclusion

The selection of peptides on which to base protein quantification can be informed by variables such as response factor and chromatographic resolution. Another consideration in this process is the fact that non-redundant tryptic sequences have m/z values of varying uniqueness and CAD products thereof similarly vary in their mass specificity. SRMass allows the categorization of transitions on the basis of specificity at discrete instrument settings. Furthermore the program can be used to predict those sample components most likely to interfere with the quantification of a given analyte.

The suitability of a peptide as an SRM target may be measured as the total number of database matches for all its transitions at fixed precursor/product ion tolerances, an estimate of the overall likelihood of isobaric interference. Based on this metric,

LSEPAELTDAVK would be a better target than FMLCAGR in an MRM-based PSA assay. Those peptide ions for which all transitions have redundancy functions with low slopes, such as IVGGWECEK could be less prone to interference at increasing transmission window, and so may be preferred targets. A set of product ions from a given precursor differ substantially in their specificity between mass accuracies of 0.01 and 2 Th, suggesting that random isobaric interference affects some more than others. Some m/z pairs are common in proteomic samples but no simple rule defines them. This finding is in agreement with studies employing global approaches to the identification of accurate transitions. Though rare, the mimicry of several analyte transition signals by a single contaminant is especially troublesome as this decreases confidence in the identity of the target compound. Grouping matches by peptide revealed those matrix proteins most incompatible with the assay and could guide proactive method design.

It should be noted that the restricted proteolytic digestion and PTM conditions simulated in searches reported here resulted in conservative estimates of redundancy. Isotopic distribution of the elements is another important contributing factor to isobaric interference which was not explored. The isotopic envelopes of analytes differing in mass by several Daltons overlap so ions several mass units lower than the query as reported by other groups is reasonable (6, 7). This effect was approximated at precursor ion mass tolerances of 1 to 2 Th. Even with the simple model used here, all transitions tested at mass windows typical of SRM experiments returned tens to hundreds of hits to the human protein database, showing the proteome to be a major contributor to chemical noise common in SRM assays.

5.5 Contributors

All programming of the algorithm used in this study was performed by Olga Krakovska (Ph.D.). The concept and input variable design of the algorithm as well as its final testing and application were carried out by this author.

Afterword

Studies described in the preceding chapters are ongoing and mass spectrometry continues to have a place in each of them. Findings presented in this dissertation are summarized below as are lines of scientific inquiry arising from them.

Chapter 2 details the assignment of phosphosites within two *S. aureus* transcription factors. In the case of VraR, the location of the phosphosite was previously assumed to be D55 based on several lines of evidence. While unmodified and phosphorylated D55 were differentiated in crude cell lysates by the method developed here, the latter form could only be observed when phospho-VraR made up around 0.1% of the total protein content of the sample, a concentration apparently higher than that of the endogenous protein. Initial attempts to detect endogenous phospho-VraR following its purification from antibiotic challenged *S. aureus* cultures were unsuccessful. It is unclear whether the absence of signal was due to inefficient purification or to low phospho-VraR concentration, since VraR autodephosphorylation appears to occur rapidly. Because VraR and phospho-VraR have similar affinity for the VraSR promoter region, VraR-dependent transcription may be maintained when the intracellular concentration of the unmodified protein is elevated. VraR notably regulates its own promoter, so antibiotic response may be activated by VraR phosphorylation then maintained by an excess of VraR. To determine if long-term antibiotic stress response is decoupled from VraR phosphorylation *in vivo*, endogenous phospho-VraR could be isolated, reduced then measured following brief duration oxacillin exposure. If VraR purification must be improved in order to detect endogenous phospho-VraR, a novel antibody may be required for this purpose. Quantification of VraR at various periods of antibiotic exposure could reveal the relative contributions of phospho-VraR and VraR to survival over time. This

could be accomplished with an MRM assay standardized with stable isotope-labelled synthetic VraR peptides.

VraR is one of many bacterial proteins controlled through aspartate phosphorylation, though phosphoanhydride reduction remains an underutilized technique in the study of this type of PTM. The methodology presented here could in principle be applied to many other proteins in *S. aureus* and other species. There are seventeen recognized two-component systems within *S. aureus* and the phosphorylation of each RR could be characterized by LC-MS.

The third chapter of this dissertation details the identification of three phosphosites which presumably modulate *Z. mays* SbeIIb function. Specific roles of each phosphosite in the interaction of SbeIIb with other proteins and in starch branching catalysis remain to be demonstrated. SbeIIb phosphosite point mutants generated by site-directed mutagenesis could be used to construct protein-protein interaction models by testing their affinities for other members of the starch biosynthesis complex. The kinases and phosphatases which modify SbeIIb are also unknown, though several candidates were recognized. Each candidate kinase could be cloned and tested in purified form for activity on each SbeIIb phosphosite. If the kinases responsible for K1 and K2 activity were established in this manner, recombinant kinases in combination with wild-type and point mutant SbeIIb could be used to find the effects of each PTM on starch affinity and starch branching.

The ultimate objective of the study of amyloplast protein phosphorylation is the complete description of starch synthesis complex function from extracellular signalling to the completion of the mature granule. GBSS, SSIIa, and SP undergo PTM at unknown residues and the discovery of additional modification sites within the complex is of importance to describing its workings. As with SbeIIb, the definition of novel GBSS, SSIIa, and SP phosphosites could be followed by biochemical examinations of each, leading to a model of protein complex modulation.

In chapter 4, the hypothesis that Beta-catenin is a substrate of p38 was tested with an *in vitro* phosphorylation assay. No phosphorylation of Beta-catenin was observed,

supporting the conclusion that the activating effect of p38 was produced indirectly. Co-immunoprecipitation of a muscle cell lysate in which Beta-catenin was the antigen failed to differentiate a p38-hyperactivated state from a state of basal p38 activity. Repetition of the immunoprecipitation experiment with subcellular fractionation preceding affinity purification is warranted as the division of compartment-specific Beta-catenin complexes may allow the effects of p38 on each to be revealed by LC-MS.

Several proteins which co-immunoprecipitated with Beta-catenin are novel interaction partners, and in the case of WDR5 the interaction was confirmed by Western blot. Studies on the interplay of Beta-catenin and WDR5 could follow from this initial discovery. Spatial and temporal components of WDR-Beta-catenin association could both be examined using LC-MS in conjunction with microscopic and biochemical techniques. LARP1 is a putative Beta-catenin binding protein; however any continuing investigation of their interaction should begin with further biochemical confirmation of its occurrence.

Results presented in chapter 5 showed that the m/z redundancy of a peptide and its product ions can be tested considering the biological matrix in which the peptide appears. Furthermore, the redundancy of a particular transition is not strictly dictated by precursor ion length or product ion length. A comparison of mass tolerance and transition redundancy revealed that selectivity varies to different degrees for each precursor/product ion pair. This implies that two or more isobaric peptides of highly dissimilar sequence may not be distinguishable based on an MS/MS spectrum containing a small number of product ions as well as the precursor ion mass, because such peptides may still have several product ion masses in common.

The SRMass program was tested on a small set of transitions and extending this study to a larger set would permit statistical analysis of transition redundancy. Predicted redundancy trends for specific transitions could also be related to actual mass spectrometry data, to show how quantification is affected by mass selection windows as well as transition selection.

References

Chapter 1

1. Edman P, Högfeldt E, Sillén LG, Kinell P. Method for determination of the amino acid sequence in peptides. Acta Chemica Scandinavica. 1950;4:283-93.

2. Amino acid sequences can be determined by automated edman degradation - biochemistry - NCBI bookshelf [Internet]; cited 1/30/2013. Available from: http://www.ncbi.nlm.nih.gov/books/NBK22571/.

3. Carr SA, Hemling ME, Bean MF, Roberts GD. Integration of mass spectrometry in analytical biotechnology. Analytical Chemistry. 1991;63(24):2802-24.

4. Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM. Electrospray ionization for mass spectrometry of large biomolecules. Science. 1989;246(4926):64-71.

5. Hillenkamp F, Karas M. Mass spectrometry of peptides and proteins by matrix-assisted ultraviolet laser desorption/ionization. Methods Enzymol. 1990;193:280-95.

6. Karas M, Hillenkamp F. Laser desorption ionization of proteins with molecular masses exceeding 10,000 Daltons. Analytical Chemistry. 1988;60(20):2299-301.

7. McLafferty FW, Bockhoff FM. Separation/identification system for complex mixtures using mass separation and mass spectral characterization. Analytical Chemistry. 1978;50(1):69-76.

8. Johnson RS, Biemann K. The primary structure of thioredoxin from chromatium vinosum determined by high-performance tandem mass spectrometry. Biochemistry. 1987;26(5):1209-14.

9. Cui W, Rohrs HW, Gross ML. Top-down mass spectrometry: Recent developments, applications and perspectives. Analyst. 2011;136(19):3854-64.

10. Keil B. Specificity of proteolysis. Berlin Heidelberg: Springer-Verlag. 1992: 66-73.

11. Olsen JV, Ong SE, Mann M. Trypsin cleaves exclusively C-terminal to arginine and lysine residues. Molecular & Cellular Proteomics. 2004;3(6):608-14.

12. Schagger H, von Jagow G. Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. Analytical Biochemistry. 1987;166:368-79.

13. Laemmli UK. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature. 1970;227:680-5.

14. McNaught AD, Wilkinson AM. IUPAC. compendium of chemical terminology, 2nd ed. (the "gold book"). XML on-line corrected version: Http://goldbook.iupac.org (2006) created by M. Nic, J. Jirat, B. Kosata; updates compiled by A. Jenkins; Blackwell Scientific Publications. 1997.

15. Skoog DA, West DM, Holler FJ, editors. Fundamentals of analytical chemistry. 7th ed. Orlando, Florida, USA: Saunders College; 1997.

16. van Deemter JJ, Zuiderweg FJ, Klinkenberg A. Longitudinal diffusion and resistance to mass transfer as causes of nonideality in chromatography. Chemical Engineering Science. 1956;5(6):271-89.

17. Garcia MC. The effect of the mobile phase additives on sensitivity in the analysis of peptides and proteins by high-performance liquid chromatography-electrospray mass spectrometry. Journal of Chromatography B. 2005;825(2):111-23.

18. Garcia MC, Hogenboom AC, Zappey H, Irth H. Effect of the mobile phase composition on the separation and detection of intact proteins by reversed-phase liquid chromatography-electrospray mass spectrometry. Journal of Chromatography A. 2002;957(2):187-99.

19. Chowdhury SK, Chait BT. Method for the electrospray ionization of highly conductive aqueous solutions. Analytical Chemistry. 1991;63(15):1660-1664.

20. Di Palma S, Hennrich ML, Heck AJ, Mohammed S. Recent advances in peptide separation by multidimensional liquid chromatography for proteome analysis. Journal of Proteomics. 2012;75(13):3791-813.

21. Wolters DA, Washburn MP, Yates JR,3rd. An automated multidimensional protein identification technology for shotgun proteomics. Analytical Chemistry. 2001;73(23):5683-90.

22. Stensballe A, Andersen S, Jensen ON. Characterization of phosphoproteins from electrophoretic gels by nanoscale fe(III) affinity chromatography with off-line mass spectrometry analysis. Proteomics. 2001;1(2):207-22.

23. Andersson L, Porath J. Isolation of phosphoproteins by immobilized metal (Fe3+) affinity chromatography. Analytical Biochemistry. 1986;154(1):250-254.

24. Posewitz MC, Tempst P. Immobilized gallium(III) affinity chromatography of phosphopeptides. Analytical Chemistry. 1999;71(14):2883-2892.

25. Zhou H, Tian R, Ye M, Xu S, Feng S, Pan C, et al. Highly specific enrichment of phosphopeptides by zirconium dioxide nanoparticles for phosphoproteome analysis. Electrophoresis. 2007;28(13):2201-15.

26. Mann M, Ong SE, Gronborg M, Steen H, Jensen ON, Pandey A. Analysis of protein phosphorylation using mass spectrometry: Deciphering the phosphoproteome. Trends in Biotechnology. 2002;20(6):261-8.

27. Ficarro SB, McCleland ML, Stukenberg PT, Burke DJ, Ross MM, Shabanowitz J, et al. Phosphoproteome analysis by mass spectrometry and its application to Saccharomyces cerevisiae. Nature Biotechnology. 2002;20(3):301-5.

28. Sugiyama N, Masuda T, Shinoda K, Nakamura A, Tomita M, Ishihama Y. Phosphopeptide enrichment by aliphatic hydroxy acid-modified metal oxide chromatography for nano-LC-MS/MS in proteomics applications. Molecular & Cellular Proteomics. 2007;6(6):1103-9.

29. Wilm M, Mann M. Analytical properties of the nanoelectrospray ion source. Analytical Chemistry. 1996;68(1):1-8.

30. El-Faramawy A, Siu KWM, Thomson BA. Efficiency of nano-electrospray ionization. Journal of the American Society for Mass Spectrometry. 2005;16(10):1702-1707.

31. Peschke M, Verkerk U, Kebarle P. Features of the ESI mechanism that affect the observation of multiply charged noncovalent protein complexes and the determination of the association constant by the titration method. Journal of the American Society for Mass Spectrometry. 2004;15(10):1424-34.

32. Rayleigh L. On the equilibrium of liquid conducting masses charged with electricity. Philosophical Magazine Series 5. 1882;14(87):184-6.

33. Allen MH, Vestal ML. Design and performance of a novel electrospray interface. Journal of the American Society for Mass Spectrometry. 1992; 3(1):18-26.

34. Kebarle P, Verkerk UH. On the mechanism of electrospray ionization mass spectrometry (ESIMS). In: Electrospray and MALDI Mass Spectrometry: Fundamentals, Instrumentation, Practicalities, and Biological Applications John Wiley & Sons, Inc. 2010. Pg 1-48.

35. Grimm RL, Beauchamp JL. Field-induced droplet ionization mass spectrometry. Journal of Physical Chemistry B. 2003;107(51):14161-3.

36. Iribarne J, Thomson B. On the evaporation of small ions from charged droplets. Journal of Chemical Physics. 1976;64(6):2287.

37. Schmelzeisen-Redeker G, Bütfering L, Röllgen FW. Desolvation of ions and molecules in thermospray mass spectrometry. International Journal of Mass Spectrometry and Ion Processes. 1989;90(2):139-50.

38. Fenn JB. Mass spectrometric implications of high-pressure ion sources. International Journal of Mass Spectrometry. 2000;200(1-3):459-478.

39. Jugroot M, Groth CPT, Thomson BA, Baranov V, Collings BA. Numerical investigation of interface region flows in mass spectrometers: Neutral gas transport. Journal of Physics D. 2004;37(8):1289-1300.

40. Jugroot M, Groth CPT, Thomson BA, Baranov V, Collings BA. Numerical investigation of interface region flows in mass spectrometers: Ion transport. Journal of Physics D. 2004;37(4):550-559.

41. Smith RD, Loo JA, Barinaga CJ, Edmonds CG, Udseth HR. Collisional activation and collision-activated dissociation of large multiply charged polypeptides and proteins produced by electrospray ionization. Journal of the American Society for Mass Spectrometry. 1990;1(1):53-65.

42. Wujcik CE, Kadar EP. Reduction of in-source collision-induced dissociation and thermolysis of sulopenem prodrugs for quantitative liquid chromatography/electrospray ionization mass spectrometric analysis by promoting sodium adduct formation. Rapid Communications in Mass Spectrometry. 2008;22(20):3195-206.

43. Cobb JS, Easterling ML, Agar JN. Structural characterization of intact proteins is enhanced by prevalent fragmentation pathways rarely observed for peptides. Journal of the American Society for Mass Spectrometry. 2010;21(6):949-59.

44. Kelly RT, Tolmachev AV, Page JS, Tang K, Smith RD. The ion funnel: Theory, implementations, and applications. Mass Spectrometry Reviews. 2010;29(2):294-312.

45. March R, Hughes R, Todd JFJ. Quadrupole storage mass spectrometry. New York: Wiley; 1989.

46. Douglas DJ. Linear quadrupoles in mass spectrometry. Mass Spectrometry Reviews. 2009;28(6):937-960.

47. Pedler RE, editor. Practical quadrupole theory: Graphical theory. Pittsburg: Extrel CMS; 2001.

48. Makarov A. Electrostatic axially harmonic orbital trapping: A high-performance technique of mass analysis. Analytical Chemistry. 2000;72(6):1156-62.

49. Guilhaus M. Special feature: Tutorial. principles and instrumentation in time-of-flight mass spectrometry. physical and instrumental concepts. Journal of Mass Spectrometry. 1995;30(11):1519-1532.

50. Cotter RJ, Griffith W, Jelinek C. Tandem time-of-flight (TOF/TOF) mass spectrometry and the curved-field reflectron. Journal of Chromatography B. 2007;855(1):2-13.

51. Hu Q, Noll RJ, Li H, Makarov A, Hardman M, Graham Cooks R. The orbitrap: A new mass spectrometer. Journal of Mass Spectrometry. 2005;40(4):430-43.

52. Roepstorff P, Fohlman J. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. Biomedical Mass Spectrometry. 1984;11(11):601.

53. Wysocki VH, Tsaprailis G, Smith LL, Breci LA. Mobile and localized protons: A framework for understanding peptide dissociation. Journal of Mass Spectrometry. 2000;35(12):1399-406.

54. Harrison AG. To b or not to b: The ongoing saga of peptide b ions. Mass Spectrometry Reviews. 2009;28(4):640-54.

55. Paizs B, Suhai S. Fragmentation pathways of protonated peptides. Mass Spectrometry Reviews. 2005;24(4):508-48.

56. Bouchoux G. Gas phase basicities of polyfunctional molecules. part 3: Amino acids. Mass Spectrometry Reviews. 2011;31(3):391-435.

57. Schlosser A, Lehmann WD. Five-membered ring formation in unimolecular reactions of peptides: A key structural element controlling low-energy collision-induced dissociation of peptides. Journal of Mass Spectrometry. 2000;35(12):1382-90.

58. Papayannopoulos IA. The interpretation of collision-induced dissociation tandem mass spectra of peptides. Mass Spectrometry Reviews. 1995;14(1):49-73.

59. Bleiholder C, Suhai S, Harrison AG, Paizs B. Towards understanding the tandem mass spectra of protonated oligopeptides. 2: The proline effect in collision-induced dissociation of protonated ala-ala-xxx-pro-ala (xxx = ala, ser, leu, val, phe, and trp). Journal of the American Society for Mass Spectrometry. 2011;22(6):1032-9.

60. Schwartz BL, Bursey MM. Some proline substituent effects in the tandem mass spectrum of protonated pentaalanine. Biological Mass Spectrometry. 1992;21:92-6.

61. Fragmentation of protonated tripeptides: The proline effect revisited. Journal of Physical Chemistry B. 2004;8(15): 4899–4908.

62. Harrison AG, Young AB. Fragmentation reactions of deprotonated peptides containing proline. the proline effect. Journal of Mass Spectrometry. 2005;40(9):1173-86.

63. Gu C, Tsaprailis G, Breci L, Wysocki VH. Selective gas-phase cleavage at the peptide bond C-terminal to aspartic acid in fixed-charge derivatives of asp-containing peptides Analytical Chemistry. 2000;72(23):5804-5813.

64. Qin J, Chait BT. Preferential fragmentation of protonated gas-phase peptide ions adjacent to acidic amino acid residues. Journal of the American Chemical Society. 1995;117(19):5411-5412.

65. Maux D, Enjalbal C, Martinez J, Aubagnac J. New example of proline-induced fragmentation in electrospray ionization mass spectrometry of peptides. Rapid Communications in Mass Spectrometry. 2002;16(15):1470-1475.

66. Syka JE, Coon JJ, Schroeder MJ, Shabanowitz J, Hunt DF. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. Proceedings of the National Academy of Sciences of the United States of America. 2004;101(26):9528-33.

67. Wodrich MD, Zhurov KO, Vorobyev A, Ben Hamidane H, Corminboeuf C, Tsybin YO. Heterolytic N-C alpha bond cleavage in electron capture and transfer dissociation of peptide cations. Journal of Physical Chemistry B. 2012;116(35):10807-15.

68. Cook SL, Zimmermann CM, Singer D, Fedorova M, Hoffmann R, Jackson GP. Comparison of CID, ETD and metastable atom-activated dissociation (MAD) of doubly and triply charged phosphorylated tau peptides. Journal of Mass Spectrometry. 2012;47(6):786-94.

69. Li X, Li Z, Xie B, Sharp JS. Improved identification and relative quantification of sites of peptide and protein oxidation for hydroxyl radical footprinting. Journal of the American Society for Mass Spectrometry. 2013;24(11):1767-1776.

70. Swaney DL, McAlister GC, Wirtala M, Schwartz JC, Syka JE, Coon JJ. Supplemental activation method for high-efficiency electron-transfer dissociation of doubly protonated peptide precursors. Analytical Chemistry. 2007;79(2):477-85.

71. Ong S, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, et al. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Molecular & Cellular Proteomics. 2002;1(5):376-86.

72. Ross PL, Huang YN, Marchese JN, Williamson B, Parker K, Hattan S, et al. Multiplexed protein quantitation in saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. Molecular & Cellular Proteomics. 2004;3(12):1154-69.

73. Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nature Biotechnology. 1999;17(10):994-9.

74. Liu H, Sadygov RG, Yates JR,3rd. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. Analytical Chemistry. 2004;76(14):4193-201.

75. Old WM, Meyer-Arendt K, Aveline-Wolf L, Pierce KG, Mendoza A, Sevinsky JR, et al. Comparison of label-free methods for quantifying human proteins by shotgun proteomics. Molecular & Cellular Proteomics. 2005;4(10):1487-502.

76. Finney GL, Blackler AR, Hoopmann MR, Canterbury JD, Wu CC, MacCoss MJ. Label-free comparative analysis of proteomics mixtures using chromatographic alignment of high-resolution microLC-MS data. Analytical Chemistry. 2008;80(4):961-71.

77. Domon B, Aebersold R. Mass spectrometry and protein analysis. Science. 2006;312(5771):212-7.

78. Gillet LC, Navarro P, Tate S, Röst H, Selevsek N, Reiter L, et al. Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: A new concept for consistent and accurate proteome analysis. Molecular & Cellular Proteomics. 2012;11(6): O111.016717.

79. Liu Y, Huttenhain R, Surinova S, Gillet LCJ, Mouritsen J, Brunner R, et al. Quantitative measurements of N-linked glycoproteins in human plasma by SWATH-MS. Proteomics. 2013;13(8):1247-56.

80. Tabb DL, Smith LL, Breci LA, Wysocki VH, Lin D, Yates JR,3rd. Statistical characterization of ion trap tandem mass spectra from doubly charged tryptic peptides. Analytical Chemistry. 2003;75(5):1155-63.

81. Frank AM. Predicting intensity ranks of peptide fragment ions. Journal of Proteome Research. 2009;8(5):2226-40.

82. Frank AM. A ranking-based scoring function for peptide-spectrum matches. Journal of Proteome Research. 2009;8(5):2241-52.

83. Kapp EA, Schütz F, Connolly LM, Chakel JA, Meza JE, Miller CA, et al. An evaluation, comparison, and accurate benchmarking of several publicly available MS/MS search algorithms: Sensitivity and specificity analysis. Proteomics. 2005;5(13):3475-90.

84. Wu X, Tseng CW, Edwards N. HMMatch: Peptide identification by spectral matching of tandem mass spectra using hidden markov models. Journal of Computational Biology. 2007;14(8):1025-43.

85. Ma B, Zhang K, Hendrie C, Liang C, Li M, Doherty-Kirby A, et al. PEAKS: Powerful software for peptide de novo sequencing by tandem mass spectrometry. Rapid Communications in Mass Spectrometry. 2003;17(20):2337-42.

86. Lam H, Aebersold R. Building and searching tandem mass (MS/MS) spectral libraries for peptide identification in proteomics. Methods. 2011 Aug;54(4):424-31.

87. Houel S, Abernathy R, Renganathan K, Meyer-Arendt K, Ahn NG, Old WM. Quantifying the impact of chimera MS/MS spectra on peptide identification in large-scale proteomics studies. Journal of Proteome Research. 2010;9(8):4152-60.

88. Perkins DN, Pappin DJ, Creasy DM, Cottrell JS. Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis. 1999;20(18):3551-67.

89. Mann M, Wilm M. Error-tolerant identification of peptides in sequence databases by peptide sequence tags. Analytical Chemistry. 1994;66(24):4390-9.

90. Eng JK, McCormack AL, Yates JR. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. Journal of the American Society for Mass Spectrometry. 1994;5(11):976-89.

91. Shilov IV, Seymour SL, Patel AA, Loboda A, Tang WH, Keating SP, et al. The paragon algorithm, a next generation search engine that uses sequence temperature values and feature probabilities to identify peptides from tandem mass spectra. Molecular & Cellular Proteomics. 2007;6(9):1638-55.

92. Elias JE, Gygi SP. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. Nature Methods. 2007;4(3):207-14.

93. Reidegeld KA, Eisenacher M, Kohl M, Chamrad D, Körting G, Blüggel M, et al. An easy-to-use decoy database builder software tool, implementing different decoy strategies for false discovery rate calculation in automated MS/MS protein identifications. Proteomics. 2008;8(6):1129-37.

94. Zhu P, Bowden P, Tucholska M, Zhang D, Marshall JG. Peptide-to-protein distribution versus a competition for significance to estimate error rate in blood protein identification. Analytical Biochemistry. 2011;411(2):241-53.

95. Fenyo D, Beavis RC. A method for assessing the statistical significance of mass spectrometry-based protein identifications using general scoring schemes. Analytical Chemistry. 2003;75(4):768-74.

96. Kapp EA, Schutz F. Overview of tandem mass spectrometry (MS/MS) database search algorithms. Current Protocols in Protein Science. 2007;25(2):1-19.

97. Brosch M, Swamy S, Hubbard T, Choudhary J. Comparison of mascot and X!tandem performance for low and high accuracy mass spectrometry and the development of an adjusted mascot threshold. Molecular & Cellular Proteomics. 2008;7(5):962-70.

98. Zhu P, Bowden P, Zhang D, Marshall JG. Mass spectrometry of peptides and proteins from human blood. Mass Spectrometry Reviews. 2010; 30(5): 685-732.

99. Tang L, Kebarle P. Dependence of ion intensity in electrospray mass spectrometry on the concentration of the analytes in the electrosprayed solution. Analytical Chemistry. 1993;65(24):3654-68.

100. Liu K, Zhang J, Wang J, Zhao L, Peng X, Jia W, et al. Relationship between sample loading amount and peptide identification and its effects on quantitative proteomics. Analytical Chemistry. 2009;81(4):1307-14.

101. Van Eeckhaut A, Lanckmans K, Sarre S, Smolders I, Michotte Y. Validation of bioanalytical LC-MS/MS assays: Evaluation of matrix effects. Journal of Chromatography B. 2009;877(23):2198-207.

102. Zhang N, Li L. Effects of common surfactants on protein digestion and matrix-assisted laser desorption/ionization mass spectrometric analysis of the digested peptides using two-layer sample preparation. Rapid Communications in Mass Spectrometry. 2004;18(8):889-96.

103. Hirabayashi A, Ishimaru M, Manri N, Yokosuka T, Hanzawa H. Detection of potential ion suppression for peptide analysis in nanoflow liquid chromatography/mass spectrometry. Rapid Communications in Mass Spectrometry. 2007;21(17):2860-6.

104. Matuszewski BK, Constanzer ML, Chavez-Eng CM. Matrix effect in quantitative LC/MS/MS analyses of biological fluids: A method for determination of finasteride in human plasma at picogram per milliliter concentrations. Analytical Chemistry. 1998;70(5):882-9.

105. Buhrman DL, Price PI, Rudewicz PJ. Quantitation of SR 27417 in human plasma using electrospray liquid chromatography-tandem mass spectrometry: A study of ion suppression. Journal of the American Society for Mass Spectrometry. 1996;7(11):1099-105.

106. Liang HR, Foltz RL, Meng M, Bennett P. Ionization enhancement in atmospheric pressure chemical ionization and suppression in electrospray ionization between target drugs and stable-isotope-labeled internal standards in quantitative liquid chromatography/tandem mass spectrometry. Rapid Communications in Mass Spectrometry. 2003;17(24):2815-21.

107. Jemal M, Schuster A, Whigan DB. Liquid chromatography/tandem mass spectrometry methods for quantitation of mevalonic acid in human plasma and urine: Method validation, demonstration of using a surrogate analyte, and demonstration of

unacceptable matrix effect in spite of use of a stable isotope analog internal standard. Rapid Communications in Mass Spectrometry. 2003;17(15):1723-34.

108. Matuszewski BK, Constanzer ML, Chavez-Eng C. Strategies for the assessment of matrix effect in quantitative bioanalytical methods based on HPLC-MS/MS. Analytical Chemistry. 2003;75(13):3019-30.

109. Yeung YG, Nieves E, Angeletti RH, Stanley ER. Removal of detergents from protein digests for mass spectrometry analysis. Analytical Biochemistry. 2008;382(2):135-7.

110. Antharavally BS, Mallia KA, Rosenblatt MM, Salunkhe AM, Rogers JC, Haney P, et al. Efficient removal of detergents from proteins and peptides in a spin column format. Analytical Biochemistry. 2011;416(1):39-44.

111. Yeung YG, Stanley ER. Rapid detergent removal from peptide samples with ethyl acetate for mass spectrometry analysis. Current Protocols in Protein Science. 2010;Chapter 16: Unit 16.12.

112. Rey M, Mrazek H, Pompach P, Novak P, Pelosi L, Brandolin G, et al. Effective removal of nonionic detergents in protein mass spectrometry, hydrogen/deuterium exchange, and proteomics. Analytical Chemistry. 2010;82(12):5107-16.

113. King R, Bonfiglio R, Fernandez-Metzler C, Miller-Stein C, Olah T. Mechanistic investigation of ionization suppression in electrospray ionization. Journal of the American Society for Mass Spectrometry. 2000;11(11):942-950.

114. Mitulovic • G, Stingl C, Steinmacher I, Hudecz O, Hutchins JRA, Peters J, et al. Preventing carryover of peptides and proteins in nano LC-MS separations. Analytical Chemistry. 2009;81(14):5955-60.

115. Lowenthal MS, Liang Y, Phinney KW, Stein SE. Quantitative bottom-up proteomics depends on digestion conditions. Analytical Chemistry. 2014;86(1):551-8.

116. Proc JL, Kuzyk MA, Hardie DB, Yang J, Smith DS, Jackson AM, et al. A quantitative study of the effects of chaotropic agents, surfactants, and solvents on the digestion efficiency of human plasma proteins by trypsin. Journal of Proteome Research. 2010;9(10):5422-37.

117. Strader MB, Tabb DL, Hervey WJ, Pan C, Hurst GB. Efficient and specific trypsin digestion of microgram to nanogram quantities of proteins in Organic/Aqueous solvent systems. Analytical Chemistry. 2006;78(1):125-34.

118. Li F, Schmerberg CM, Ji QC. Accelerated tryptic digestion of proteins in plasma for absolute quantitation using a protein internal standard by liquid chromatography/tandem mass spectrometry. Rapid Communications in Mass Spectrometry. 2009;23(5):729-32.

119. Arsene CG, Ohlendorf R, Burkitt W, Pritchard C, Henrion A, O'Connor G, et al. Protein quantification by isotope dilution mass spectrometry of proteolytic fragments: Cleavage rate and accuracy. Analytical Chemistry. 2008;80(11):4154-60.

120. Juan H, Chang S, Huang H, Chen S. A new application of microwave technology to proteomics. Proteomics. 2005;5(4):840-2.

121. Hart GW, Copeland RJ. Glycomics hits the big time. Cell. 2010;143(5):672-6.

122. Gagneux P, Varki A. Evolutionary considerations in relating oligosaccharide diversity to biological function. Glycobiology. 1999;9(8):747-55.

123. Rademaker GJ, Pergantis SA, Blok-Tip L, Langridge JI, Kleen A, Thomas-Oates JE. Mass spectrometric determination of the sites of O-glycan attachment with low picomolar sensitivity. Analytical Biochemistry. 1998;257(2):149-60.

124. Maley F, Trimble RB, Tarentino AL, Plummer Jr. TH. Characterization of glycoproteins and their associated oligosaccharides through the use of endoglycosidases. Analytical Biochemistry. 1989;180(2):195-204.

125. Creskey MC, Smith DGS, Cyr TD. Strain identification of commercial influenza vaccines by mass spectrometry. Analytical Biochemistry. 2010;406(2):193-203.

126. R. Alley Jr. W, Mechref Y, Novotny M. Use of activated graphitized carbon chips for liquid chromatography/mass spectrometric and tandem mass spectrometric analysis of tryptic glycopeptides. Rapid Communications in Mass Spectrometry. 2009;23(4):495-505.

127. Song E, Pyreddy S, Mechref Y. Quantification of glycopeptides by multiple reaction monitoring liquid chromatography/tandem mass spectrometry. Rapid Communications in Mass Spectrometry. 2012;26:1941-54.

128. McCleary WR, Stock JB, Ninfa AJ. Is acetyl phosphate a global signal in *Escherichia coli*? Journal of Bacteriology. 1993;175(10):2793-8.

129. Zolnierowicz S, Bollen M. Protein phosphorylation and protein phosphatases de panne, belgium, september 19-24, 1999. EMBO Journal. 2000;19(4):483-8.

130. Attwood PV, Besant PG, Piggott MJ. Focus on phosphoaspartate and phosphoglutamate. Amino Acids. 2011;44(4):1035-51.

131. Trumbore MW, Wang RH, Enkemann SA, Berger SL. Prothymosin alpha in vivo contains phosphorylated glutamic acid residues. Journal of Biological Chemistry. 1997;272(42):26394-404.

132. Attwood PV, Piggott MJ, Zu XL, Besant PG. Focus on phosphohistidine. Amino Acids. 2007;32(1):145-56.

133. Palumbo AM, Tepe JJ, Reid GE. Mechanistic insights into the multistage gas-phase fragmentation behavior of phosphoserine- and phosphothreonine-containing peptides. Journal of Proteome Research. 2008;7(2):771-9.

134. Resing KA, Ahn NG. Protein phosphorylation analysis by electrospray ionization-mass spectrometry. Methods in Enzymology. Academic Press. 1997;283: 29-44.

135. Tichy A, Salovska B, Rehulka P, Klimentova J, Vavrova J, Stulik J, et al. Phosphoproteomics: Searching for a needle in a haystack. Journal of Proteomics. 2011;74(12):2786-97.

136. Palumbo AM, Reid GE. Evaluation of gas-phase rearrangement and competing fragmentation reactions on protein phosphorylation site assignment using collision induced dissociation-MS/MS and MS3. Analytical Chemistry. 2008;80(24):9735-47.

137. Arrigoni G, Resjö S, Levander F, Nilsson R, Degerman E, Quadroni M, et al. Chemical derivatization of phosphoserine and phosphothreonine containing peptides to increase sensitivity for MALDI-based analysis and for selectivity of MS/MS analysis. Proteomics. 2006;6(3):757-766.

138. Lehmann WD, Kruger R, Salek M, Hung C, Wolschin F, Weckwerth W. Neutral loss-based phosphopeptide recognition: A collection of caveats. Journal of Proteome Research. 2007;6(7):2866-73.

139. Geiger T, Clarke S. Deamidation, isomerization, and racemization at asparaginyl and aspartyl residues in peptides. succinimide-linked reactions that contribute to protein degradation. Journal of Biological Chemistry. 1987;262(2):785-94.

140. Robinson NE, Robinson ZW, Robinson BR, Robinson AL, Robinson JA, Robinson ML, et al. Structure-dependent nonenzymatic deamidation of glutaminyl and asparaginyl pentapeptides. Journal of Peptide Research. 2004;63(5):426-436.

141. Yang H, Zubarev RA. Mass spectrometric analysis of asparagine deamidation and aspartate isomerization in polypeptides. Electrophoresis. 2010;31(11):1764-72.

142. Krokhin OV, Antonovici M, Ens W, Wilkins JA, Standing KG. Deamidation of -asn-gly- sequences during sample preparation for proteomics: Consequences for MALDI and HPLC-MALDI analysis. Analytical Chemistry. 2006;78(18):6645-50.

143. Reissnera KJ, Aswad DW. Deamidation and isoaspartate formation in proteins: Unwanted alterations or surreptitious signals? Cellular and Molecular Life Sciences. 2003;60:1281-95.

144. Joshi AB, Kirsch LE. The relative rates of glutamine and asparagine deamidation in glucagon fragment 22–29 under acidic conditions. Journal of Pharmaceutical Sciences. 2002;91(11):2331-45.

145. Hao P, Ren Y, Alpert AJ, Sze SK. Detection, evaluation and minimization of nonenzymatic deamidation in proteomic sample preparation. Molecular & Cellular Proteomics. 2011;10(10): O111.009381.

146. Perdivara I, Deterding LJ, Przybylski M, Tomer KB. Mass spectrometric identification of oxidative modifications of tryptophan residues in proteins: Chemical artifact or post-translational modification? Journal of the American Society for Mass Spectrometry. 2010;21(7):1114-7.

147. Boja ES, Fales HM. Overalkylation of a protein digest with iodoacetamide. Analytical Chemistry. 2001;73(15):3576-82.

148. Bremang M, Cuomo A, Agresta AM, Stugiewicz M, Spadotto V, Bonaldi T. Mass spectrometry-based identification and characterisation of lysine and arginine methylation in the human proteome. Molecular Biosystems. 2013;9(9):2231-47.

149. Ikeda F, Dikic I. Atypical ubiquitin chains: New molecular signals. 'protein modifications: Beyond the usual suspects' review series. EMBO Reports. 2008;9(6):536-42.

150. Dikic I, Wakatsuki S, Walters KJ. Ubiquitin-binding domains - from structures to functions. Nature Reviews Molecular Cell Biology. 2009;10(10):659-71.

151. Esteban Warren MR, Parker CE, Mocanu V, Klapper D, Borchers CH. Electrospray ionization tandem mass spectrometry of model peptides reveals diagnostic fragment ions

for protein ubiquitination. Rapid Communications in Mass Spectrometry. 2005;19(4):429-437.

152. Nielsen ML, Vermeulen M, Bonaldi T, Cox J, Moroder L, Mann M. Iodoacetamide-induced artifact mimics ubiquitination in mass spectrometry. Nature Methods. 2008;5(6):459-460.

153. Michalski A, Cox J, Mann M. More than 100,000 detectable peptide species elute in single shotgun proteomics runs but the majority is inaccessible to data-dependent LC−MS/MS. Journal of Proteome Research 2011;10(4):1785.

Chapter 2

1. Klein E, Smith DL, Laxminarayan R. Hospitalizations and deaths caused by methicillin-resistant *Staphylococcus aureus*. Emerging Infectious Diseases. 2007; 13(12):1840-6.

2. Ohki R, Giyanto, Tateno K, Masuyama W, Moriya S, Kobayashi K, et al. The BceRS two-component regulatory system induces expression of the bacitracin transporter, BceAB, in bacillus subtilis. Molecular Microbiology. 2003; 49(4):1135-44.

3. Kuroda M, Kuroda H, Oshima T, Takeuchi F, Mori H, Hiramatsu K. Two-component system VraSR positively modulates the regulation of cell-wall biosynthesis pathway in *Staphylococcus aureus*. Molecular Microbiology. 2003; 49(3):807-21.

4. Robinson VL, Buckler DR, Stock AM. A tale of two components: A novel kinase and a regulatory switch. Nature Structural and Molecular Biology. 2000; 7(8):626-33.

5. Otto M, Kawada-Matsuo M, Yoshida Y, Zendo T, Nagao J, Oogai Y, et al. Three distinct two-component systems are involved in resistance to the class I bacteriocins, nukacin ISK-1 and nisin A, in *Staphylococcus aureus*. PLoS ONE. 2013;8(7).

6. Kawada-Matsuo M, Yoshida Y, Nakamura N, Komatsuzawa H. Role of two-component systems in the resistance of *Staphylococcus aureus* to antibacterial agents. Virulence. 2011;2(5):427-30.

7. Kuroda M, Ohta T, Uchiyama I, Baba T, Yuzawa H, Kobayashi I, et al. Whole genome sequencing of meticillin-resistant *Staphylococcus aureus*. The Lancet. 2001; 357(9264): 1225-40.

8. Thomas SA, Brewster JA, Bourret RB. Two variable active site residues modulate response regulator phosphoryl group stability. Molecular Microbiology. 2008;69(2):453-65.

9. Attwood PV, Besant PG, Piggott MJ. Focus on phosphoaspartate and phosphoglutamate. Amino Acids. 2011; 44(4):1035-51.

10. Yan D, Cho HS, Hastings CA, Igo MM, Lee SY, Pelton JG, et al. Beryllofluoride mimics phosphorylation of NtrC and other bacterial response regulators. Proceedings of the National Academy of Sciences of the United States of America. 1999;96(26):14789-94.

11. Chabre M. Aluminofluoride and beryllofluoride complexes: New phosphate analogs in enzymology. Trends in Biochemical Sciences. 1990;15(1):6-10.

12. Degani C, Boyer PD. A borohydride reduction method for characterization of the acyl phosphate linkage in proteins and its application to sarcoplasmic reticulum adenosine triphosphatase. Journal of Biological Chemistry. 1973;248(23):8222-6.

13. Sanders DA, Gillece-Castro BL, Stock AM, Burlingame AL, Koshland DE,Jr. Identification of the site of phosphorylation of the chemotaxis response regulator protein, CheY. Journal of Biological Chemistry. 1989;264(36):21770-8.

14. Appleby JL, Bourret RB. Activation of CheY mutant D57N by phosphorylation at an alternative site, ser-56. Molecular Microbiology. 1999;34(5):915-25.

15. Gardete S, Wu SW, Gill S, Tomasz A. Role of VraSR in antibiotic resistance and antibiotic-induced stress response in *Staphylococcus aureus*. Antimicrobial Agents and Chemotherapy. 2006;50(10):3424-34.

16. Lukat GS, McCleary WR, Stock AM, Stock JB. Phosphorylation of bacterial response regulator proteins by low molecular weight phospho-donors. Proceedings of the National Academy of Sciences of the United States of America. 1992;89(2):718-22.

17. Belcheva A, Golemi-Kotra D. A close-up view of the VraSR two-component system. A mediator of *Staphylococcus aureus* response to cell wall damage. Journal of Biological Chemistry. 2008;283(18):12354-64.

18. Leonard PG, Golemi-Kotra D, Stock AM. Phosphorylation-dependent conformational changes and domain rearrangements in *Staphylococcus aureus* VraR activation. Proceedings of the National Academy of Sciences of the United States of America. 2013;110(21):8525-30.

19. Liu YH, Belcheva A, Konermann L, Golemi-Kotra D. Phosphorylation-induced activation of the response regulator VraR from *Staphylococcus aureus*: Insights from hydrogen exchange mass spectrometry. Journal of Molecular Biology. 2009;391(1):149-63.

20. Belcheva A, Verma V, Golemi-Kotra D. DNA-binding activity of the vancomycin resistance associated regulator protein VraR and the role of phosphorylation in transcriptional regulation of the vraSR operon. Biochemistry. 2009;48(24):5592-601.

21. Hessling B, Bonn F, Otto A, Herbst F, Rappen G, Bernhardt J, et al. Global proteome analysis of vancomycin stress in *Staphylococcus aureus*. International Journal of Medical Microbiology. 2013;303(8):624-34.

22. Meehl M, Herbert S, Gotz F, Cheung A. Interaction of the GraRS two-component system with the VraFG ABC transporter to support vancomycin-intermediate resistance in *Staphylococcus aureus*. Antimicrobial Agents and Chemotherapy. 2007;51(8):2679-89.

23. Li M, Cha DJ, Lai Y, Villaruz AE, Sturdevant DE, Otto M. The antimicrobial peptide-sensing system aps of *Staphylococcus aureus*. Molecular Microbiology. 2007;66(5):1136-47.

24. Falord M, Mader U, Hiron A, Debarbouille M, Msadek T. Investigation of the *Staphylococcus aureus* GraSR regulon reveals novel links to virulence, stress response and cell wall signal transduction pathways. PLoS One. 2011;6(7).

25. Fridman M, Williams GD, Muzamal U, Hunter H, Siu KWM, Golemi-Kotra D. Two unique phosphorylation-driven signaling pathways crosstalk in *Staphylococcus aureus* to modulate the cell-wall charge: Stk1/Stp1 meets GraSR. Biochemistry. 2013;52(45):7975-86.

26. Debarbouille M, Dramsi S, Dussurget O, Nahori MA, Vaganay E, Jouvion G, et al. Characterization of a serine/threonine kinase involved in virulence of *Staphylococcus aureus*. Journal of Bacteriology. 2009;191(13):4070-81.

27. Donat S, Streker K, Schirmeister T, Rakette S, Stehle T, Liebeke M, et al. Transcriptome and functional analysis of the eukaryotic-type serine/threonine kinase PknB in *Staphylococcus aureus*. Journal of Bacteriology. 2009;191(13):4056-69.

28. Beltramini AM, Mukhopadhyay CD, Pancholi V. Modulation of cell wall structure and antimicrobial susceptibility by a *Staphylococcus aureus* eukaryote-like serine/threonine kinase and phosphatase. Infection and Immunity. 2009;77(4):1406-16.

29. Burnside K, Rajagopal L. Regulation of prokaryotic gene expression by eukaryotic-like enzymes. Current Opinion in Microbiology. 2012;15(2):125-31.

30. UniProt Consortium. Reorganizing the protein space at the universal protein resource (UniProt). Nucleic Acids Research. 2012; 40(Database issue):D71-5.

31. Mann M, Ong SE, Gronborg M, Steen H, Jensen ON, Pandey A. Analysis of protein phosphorylation using mass spectrometry: Deciphering the phosphoproteome. Trends in Biotechnology. 2002; 20(6):261-8.

32. Palumbo AM, Reid GE. Evaluation of gas-phase rearrangement and competing fragmentation reactions on protein phosphorylation site assignment using collision induced dissociation-MS/MS and MS3. Analytical Chemistry. 2008; 80(24):9735-47.

33. Krokhin OV, Antonovici M, Ens W, Wilkins JA, Standing KG. Deamidation of -asn-gly- sequences during sample preparation for proteomics: Consequences for MALDI and HPLC-MALDI analysis. Analytical Chemistry. 2006; 78(18):6645-50.

34. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. Journal of Molecular Biology. 1982; 157(1):105-32.

Chapter 3

1. Pérez S, Bertoft E. The molecular structures of starch components and their contribution to the architecture of starch granules: A comprehensive review. Starch - Stärke. 2010;62(8):389-420.

2. Klucinec JD, Thompson DB. Structure of amylopectins from ae-containing maize starches. Cereal Chemistry Journal. 2002;79(1):19-23.

3. Yamamori M, Fujita S, Hayakawa K, Matsuki J, Yasui T. Genetic elimination of a starch granule protein, SGP-1, of wheat generates an altered starch with apparent high amylose. Theoretical and Applied Genetics. 2000;101(1):21-9.

4. Shu X, Jia L, Gao J, Song Y, Zhao H, Nakamura Y, et al. The influences of chain length of amylopectin on resistant starch in rice (oryza sativa L.). Starch - Stärke. 2007;59(10):504-9.

5. Hanashiro I, Itoh K, Kuratomi Y, Yamazaki M, Igarashi T, Matsugasako J-, et al. Granule-bound starch synthase I is responsible for biosynthesis of extra-long unit chains of amylopectin in rice. Plant and Cell Physiology. 2008;49(6):925-33.

6. Tetlow IJ. Starch: Origins, structure and metabolism. Essential Reviews In Experimental Biology. 2012;5:141-78.

7. Lin Q, Huang B, Zhang M, Zhang X, Rivenbark J, Lappe RL, et al. Functional interactions between starch synthase III and isoamylase-type starch-debranching enzyme in maize endosperm. Plant Physiology. 2012;158(2):679-92.

8. Kuriki T, Guan HP, Preiss J. Structure-function relationships of starch branching enzyme. Nippon Nogeikagaku Kaishi. 1994;68:1581-4.

9. Takeda Y, Guan H, Preiss J. Branching of amylose by the branching isoenzymes of maize endosperm. Carbohydrate Research. 1993;240:253-63.

10. Guan HP, Preiss J. Differentiation of the properties of the branching isozymes from maize (*Zea mays*). Plant Physiology. 1993;102(4):1269-73.

11. Gao M, Fisher DK, Kim KN, Shannon JC, Guiltinan MJ. Independent genetic control of maize starch-branching enzymes IIa and IIb (isolation and characterization of a Sbe2a cDNA). Plant Physiology. 1997;114(1):69-78.

12. Kim KN, Fisher DK, Gao M, Guiltinan MJ. Molecular cloning and characterization of the amylose-extender gene encoding starch branching enzyme IIB in maize. Plant Molecular Biology. 1998;38(6):945-56.

13. Liu F, Ahmed Z, Lee EA, Donner E, Liu Q, Ahmed R, et al. Allelic variants of the amylose extender mutation of maize demonstrate phenotypic variation in starch structure resulting from modified protein-protein interactions. Journal of Experimental Botany. 2012;63(3):1167-83.

14. Hennen-Bierwagen TA, Liu F, Marsh RS, Kim S, Gan Q, Tetlow IJ, et al. Starch biosynthetic enzymes from developing maize endosperm associate in multisubunit complexes. Plant Physiology. 2008;146(4):1892-908.

15. Liu F, Romanova N, Lee EA, Ahmed R, Evans M, Gilbert EP, et al. Glucan affinity of starch synthase IIa determines binding of starch synthase I and starch-branching enzyme IIb to starch granules. Biochemical Journal. 2012;448(3):373-87.

16. Liu F, Makhmoudova A, Lee EA, Wait R, Emes MJ, Tetlow IJ. The amylose extender mutant of maize conditions novel protein–protein interactions between starch biosynthetic enzymes in amyloplasts. Journal of Experimental Botany. 2009;60(15):4423-40.

17. Tetlow IJ, Wait R, Lu Z, Akkasaeng R, Bowsher CG, Esposito S, et al. Protein phosphorylation in amyloplasts regulates starch branching enzyme activity and Protein–Protein interactions. The Plant Cell. 2004;16(3):694-708.

18. Yu Y, Mu HH, Mu-Forster C, Wasserman BP. Stromal localization of starch-biosynthetic enzymes and identification of an 81-kilodalton amyloplast stromal heat-shock cognate. Plant Physiology. 1998;116(4):1451-60.

19. Grimaud F, Rogniaux H, James MG, Myers AM, Planchot V. Proteome and phosphoproteome analysis of starch granule-associated proteins from normal maize and mutants affected in starch biosynthesis. Journal of Experimental Botany. 2008;59(12):3395-406.

20. Tetlow IJ, Blissett KJ, Emes MJ. Starch synthesis and carbohydrate oxidation in amyloplasts from developing wheat endosperm Planta. 1994;194(4):454-60.

21. Zhang Y. I-TASSER server for protein 3D structure prediction. BMC Bioinformatics. 2008;9:40.

22. Hansson T, Nordlund P, Aqvist J. Energetics of nucleophile activation in a protein tyrosine phosphatase. Journal of Molecular Biology. 1997;265(2):118-27.

23. Blom N, Gammeltoft S, Brunak S. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. Journal of Molecular Biology. 1999;294(5):1351-62.

24. Makhmoudova A, Williams D, Brewer D, Massey S, Patterson J, Silva A, et al. Identification of multiple phosphorylation sites on maize endosperm starch branching enzyme IIb, a key enzyme in amylopectin biosynthesis. Journal of Biological Chemistry. 2014; 289(13):9233-46.

25. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. Journal of Molecular Biology. 1982;157(1):105-32.

26. Lawless C, Hubbard SJ. Prediction of missed proteolytic cleavages for the selection of surrogate peptides for quantitative proteomics. OMICS. 2012 ;16(9):449-56.

27. Szczegielniak J, Klimecka M, Liwosz A, Ciesielski A, Kaczanowski S, Dobrowolska G, et al. A wound-responsive and phospholipid-regulated maize calcium-dependent protein kinase. Plant Physiology. 2005;139(4):1970-83.

28. Knetsch M, Wang M, Snaar-Jagalska BE, Heimovaara-Dijkstra S. Abscisic acid induces mitogen-activated protein kinase activation in barley aleurone protoplasts. The Plant Cell Online. 1996;8(6):1061-7.

29. Chen J, Huang B, Li Y, Du H, Gu Y, Liu H, et al. Synergistic influence of sucrose and abscisic acid on the genes involved in starch synthesis in maize endosperm. Carbohydrate Research. 2011;346(13):1684-91.

30. Park JH, Lee SY, Kim WY, Jung YJ, Chae HB, Jung HS, et al. Heat-induced chaperone activity of serine/threonine protein phosphatase 5 enhances thermotolerance in *Aarabidopsis thaliana*. New Phytologist. 2011;191(3):692-705.

Chapter 4

1. Trevisani F, Cantarini MC, Wands JR, Bernardi M. Recent advances in the natural history of hepatocellular carcinoma. Carcinogenesis. 2008;29(7):1299-305.

2. Mercola M, Ruiz-Lozano P, Schneider MD. Cardiac muscle regeneration: Lessons from development. Genes and Development. 2011;25(4):299-309.

3. Mill C, George SJ. Wnt signalling in smooth muscle cells and its role in cardiovascular disorders. Cardiovascular Research. 2012;95(2):233-40.

4. Nusse R, Brown A, Papkoff J, Scambler P, Shackleford G, McMahon A, et al. A new nomenclature for int-1 and related genes: The wnt gene family. Cell. 1991;64(2):231.

5. Wigmore PM, Evans DJ. Molecular and cellular mechanisms involved in the generation of fiber diversity during myogenesis. International Review of Cytology. 2002;216:175-232.

6. Cong F, Schweizer L, Varmus H. Wnt signals across the plasma membrane to activate the Beta-catenin pathway by forming oligomers containing its receptors, frizzled and LRP. Development. 2004;131(20):5103-15.

7. Clevers H. Wnt/Beta-catenin signaling in development and disease. Cell. 2006;127(3):469-80.

8. Itoh K, Antipova A, Ratcliffe MJ, Sokol S. Interaction of dishevelled and xenopus axin-related protein is required for wnt signal transduction. Molecular and Cellular Biology. 2000;20(6):2228-38.

9. Fagotto F. Looking beyond the wnt pathway for the deep nature of Beta-catenin. EMBO Reports. 2013;14(5):422-33.

10. Zeng X, Tamai K, Doble B, Li S, Huang H, Habas R, et al. A dual-kinase mechanism for wnt co-receptor phosphorylation and activation. Nature. 2005;438(7069):873-7.

11. Davidson G, Wu W, Shen J, Bilic J, Fenger U, Stannek P, et al. Casein kinase 1 gamma couples wnt receptor activation to cytoplasmic signal transduction. Nature. 2005;438(7069):867-72.

12. Arce L, Yokoyama NN, Waterman ML. Diversity of LEF/TCF action in development and disease. Oncogene. 2006;25(57):7492-504.

13. Donat S, Streker K, Schirmeister T, Rakette S, Stehle T, Liebeke M, et al. Transcriptome and functional analysis of the eukaryotic-type serine/threonine kinase PknB in *Staphylococcus aureus*. Journal of Bacteriology. 2009; 191(13): 4056-69.

14. Cadigan KM, Peifer M. Wnt signaling from development to disease: Insights from model systems. Cold Spring Harbor Perspectives in Biology. 2009;1(2):a002881.

15. Valenta T, Hausmann G, Basler K. The many faces and functions of Beta-catenin. EMBO Journal. 2012;31(12):2714-36.

16. Yang W, Xia Y, Ji H, Zheng Y, Liang J, Huang W, et al. Nuclear PKM2 regulates Beta-catenin transactivation upon EGFR activation. Nature. 2011;480(7375):118-22.

17. Fang D, Hawke D, Zheng Y, Xia Y, Meisenhelder J, Nika H, et al. Phosphorylation of Beta-catenin by AKT promotes Beta-catenin transcriptional activity. Journal of Biological Chemistry. 2007;282(15):11221-9.

18. Cross DA, Alessi DR, Cohen P, Andjelkovich M, Hemmings BA. Inhibition of glycogen synthase kinase-3 by insulin mediated by protein kinase B. Nature. 1995;378(6559):785-9.

19. Su Y, Fu C, Ishikawa S, Stella A, Kojima M, Shitoh K, et al. APC is essential for targeting phosphorylated Beta-catenin to the SCFbeta-TrCP ubiquitin ligase. Molecular Cell. 2008;32(5):652-61.

20. Wolf D, Rodova M, Miska EA, Calvet JP, Kouzarides T. Acetylation of Beta-catenin by CREB-binding protein (CBP). Journal of Biological Chemistry. 2002;277(28):25562-7.

21. Bikkavilli RK, Feigin ME, Malbon CC. p38 mitogen-activated protein kinase regulates canonical wnt-Beta-catenin signaling by inactivation of GSK3 beta. Journal of Cell Science. 2008;121(Pt 21):3598-607.

22. Thornton TM, Pedraza-Alva G, Deng B, Wood CD, Aronshtam A, Clements JL, et al. Phosphorylation by p38 MAPK as an alternative pathway for GSK3 beta inactivation. Science. 2008;320(5876):667-70.

23. Cervenka I, Wolf J, Masek J, Krejci P, Wilcox WR, Kozubik A, et al. Mitogen-activated protein kinases promote WNT/Beta-catenin signaling via phosphorylation of LRP6. Molecular and Cellular Biology. 2011;31(1):179-89.

24. Uglow EB, Slater S, Sala-Newby GB, Aguilera-Garcia CM, Angelini GD, Newby AC, et al. Dismantling of cadherin-mediated cell-cell contacts modulates smooth muscle cell proliferation. Circulation Research. 2003;92(12):1314-21.

25. Yamaguchi H, Igarashi M, Hirata A, Tsuchiya H, Susa S, Tominaga M, et al. Characterization of platelet-derived growth factor-induced p38 mitogen-activated protein kinase activation in vascular smooth muscle cells. European Journal of Clinical Investigation. 2001;31(8):672-80.

26. Moriguchi T, Kuroyanagi N, Yamaguchi K, Gotoh Y, Irie K, Kano T, et al. A novel kinase cascade mediated by mitogen-activated protein kinase kinase 6 and MKK3. Journal of Biological Chemistry. 1996;271(23):13675-9.

27. Zetser A, Gredinger E, Bengal E. p38 mitogen-activated protein kinase pathway promotes skeletal muscle differentiation. participation of the Mef2c transcription factor. Journal of Biological Chemistry. 1999;274(8):5193-200.

28. Rao RS, Miano JM, Olson EN, Seidel CL. The A10 cell line: A model for neonatal, neointimal, or differentiated vascular smooth muscle cells? Cardiovascular Research. 1997;36(1):118-26.

29. Blom N, Gammeltoft S, Brunak S. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. Journal of Molecular Biology. 1999;294(5):1351-62.

30. Ouwens DM, de Ruiter ND, van der Zon GC, Carter AP, Schouten J, van der Burgt C, et al. Growth factors can activate ATF2 via a two-step mechanism: Phosphorylation of Thr71 through the ras-MEK-ERK pathway and of Thr69 through RalGDS-src-p38. EMBO Journal. 2002;21(14):3782-93.

31. Bardwell L. Mechanisms of MAPK signalling specificity. Biochemical Society Transactions. 2006;34(Pt 5):837-41.

32. Takemaru KI, Moon RT. The transcriptional coactivator CBP interacts with Beta-catenin to activate gene expression. Journal of Cell Biology. 2000;149(2):249-54.

33. Barker N, Hurlstone A, Musisi H, Miles A, Bienz M, Clevers H. The chromatin remodelling factor brg-1 interacts with Beta-catenin to promote target gene activation. EMBO Journal. 2001;20(17):4935-43.

34. Griffin CT, Curtis CD, Davis RB, Muthukumar V, Magnuson T. The chromatin-remodeling enzyme BRG1 modulates vascular wnt signaling at two levels. Proceedings of the National Academy of Sciences of the United States of America. 2011;108(6):2282-7.

35. Mellacheruvu D, Wright Z, Couzens AL, Lambert JP, St-Denis NA, Li T, et al. The CRAPome: A contaminant repository for affinity purification-mass spectrometry data. Nature Methods. 2013;10(8):730-6.

36. Li FQ, Mofunanya A, Harris K, Takemaru K. Chibby cooperates with 14-3-3 to regulate Beta-catenin subcellular distribution and signaling activity. Journal of Cell Biology. 2008;181(7):1141-54.

37. Gottardi CJ, Gumbiner BM. Distinct molecular forms of Beta-catenin are targeted to adhesive or transcriptional complexes. Journal of Cell Biology. 2004;167(2):339-49.

38. Staal FJ, Noort Mv M, Strous GJ, Clevers HC. Wnt signals are transmitted through N-terminally dephosphorylated Beta-catenin. EMBO Reports. 2002;3(1):63-8.

39. Zhu ED, Demay MB, Gori F. Wdr5 is essential for osteoblast differentiation. Journal of Biological Chemistry. 2008;283(12):7361-7.

40. Burrows C, Abd Latip N, Lam SJ, Carpenter L, Sawicka K, Tzolovsky G, et al. The RNA binding protein Larp1 regulates cell division, apoptosis and cell migration. Nucleic Acids Research. 2010;38(16):5542-53.

41. Aoki K, Adachi S, Homoto M, Kusano H, Koike K, Natsume T. LARP1 specifically recognizes the 3' terminus of poly(A) mRNA. FEBS Letters. 2013;587(14):2173-8.

42. Tcherkezian J, Cargnello M, Romeo Y, Huttlin EL, Lavoie G, Gygi SP, et al. Proteomic analysis of cap-dependent translation identifies LARP1 as a key regulator of 5'TOP mRNA translation. Genes and Development. 2014;28(4):357-71.

43. Blagden SP, Gatt MK, Archambault V, Lada K, Ichihara K, Lilley KS, et al. Drosophila larp associates with poly(A)-binding protein and is required for male fertility and syncytial embryo development. Developmental Biology. 2009;334(1):186-97.

44. Bear DG, Fomproix N, Soop T, Bjorkroth B, Masich S, Daneholt B. Nuclear poly(A)-binding protein PABPN1 is associated with RNA polymerase II during transcription and accompanies the released transcript to the nuclear pore. Experimental Cell Research. 2003;286(2):332-44.

45. Apponi LH, Leung SW, Williams KR, Valentini SR, Corbett AH, Pavlath GK. Loss of nuclear poly(A)-binding protein 1 causes defects in myogenesis and mRNA biogenesis. Human Molecular Genetics. 2010;19(6):1058-65.

46. Lazarova DL, Chiaro C, Wong T, Drago E, Rainey A, O'Malley S, et al. CBP activity mediates effects of the histone deacetylase inhibitor butyrate on WNT activity and apoptosis in colon cancer cells. Journal of Cancer. 2013;4(6):481-90.

47. Kim JH, Kim B, Cai L, Choi HJ, Ohgi KA, Tran C, et al. Transcriptional regulation of a metastasis suppressor gene by Tip60 and Beta-catenin complexes. Nature. 2005;434(7035):921-6.

48. Rashid S, Pilecka I, Torun A, Olchowik M, Bielinska B, Miaczynska M. Endosomal adaptor proteins APPL1 and APPL2 are novel activators of Beta-catenin/TCF-mediated transcription. Journal of Biological Chemistry. 2009;284(27):18115-28.

49. Hong W, Resnick RJ, Rakowski C, Shalloway D, Taylor SJ, Blobel GA. Physical and functional interaction between the transcriptional cofactor CBP and the KH domain protein Sam68. Molecular Cancer Research. 2002;1(1):48-55.

50. Brauer PM, Zheng Y, Evans MD, Dominguez-Brauer C, Peehl DM, Tyner AL. The alternative splice variant of protein tyrosine kinase 6 negatively regulates growth and enhances PTK6-mediated inhibition of Beta-catenin. PLoS One. 2011;6(3):e14789.

51. Chrivia JC, Kwok RP, Lamb N, Hagiwara M, Montminy MR, Goodman RH. Phosphorylated CREB binds specifically to the nuclear protein CBP. Nature. 1993;365(6449):855-9.

52. Sheng M, Thompson MA, Greenberg ME. CREB: A Ca(2+)-regulated transcription factor phosphorylated by calmodulin-dependent kinases. Science. 1991;252(5011):1427-30.

53. Kuo CH, Nishikawa E, Ichikawa H, Sadakata T, Niu SY, Miki N. Calmodulin functions as an activator of pur alpha binding to single-stranded purine-rich DNA elements (PUR elements). Biochemical and Biophysical Research Communications. 1999;255(2):406-11.

54. Kelm RJ,Jr, Cogan JG, Elder PK, Strauch AR, Getz MJ. Molecular interactions between single-stranded DNA-binding proteins associated with an essential MCAT element in the mouse smooth muscle alpha-actin promoter. Journal of Biological Chemistry. 1999;274(20):14238-45.

55. Shih HM, Goldman PS, DeMaggio AJ, Hollenberg SM, Goodman RH, Hoekstra MF. A positive genetic selection for disrupting protein-protein interactions: Identification of CREB mutations that prevent association with the coactivator CBP. Proceedings of the National Academy of Sciences of the United States of America. 1996;93(24):13896-901.

56. Tian X, Zhang J, Tan TK, Lyons JG, Zhao H, Niu B, et al. Association of Beta-catenin with P-Smad3 but not LEF-1 dissociates in vitro profibrotic from anti-inflammatory effects of TGF-beta1. Journal of Cell Science. 2013;126(Pt 1):67-76.

57. Hariharan S, Kelm RJ,Jr, Strauch AR. The pur-alpha/Pur-beta single-strand DNA-binding proteins attenuate smooth-muscle actin gene transactivation in myofibroblasts. Journal of cellular physiology. 2014; 229:1256–71.

58. Charbonney E, Speight P, Masszi A, Nakano H, Kapus A. Beta-catenin and Smad3 regulate the activity and stability of myocardin-related transcription factor during epithelial-myofibroblast transition. Molecular Biology of the Cell. 2011;22(23):4472-85.

59. Gratenstein K, Heggestad AD, Fortun J, Notterpek L, Pestov DG, Fletcher BS. The WD-repeat protein GRWD1: Potential roles in myeloid differentiation and ribosome biogenesis. Genomics. 2005;85(6):762-73.

60. Sugimoto N, Kitabayashi I, Osano S, Tatsumi Y, Yugawa T, Narisawa-Saito M, et al. Identification of novel human Cdt1-binding proteins by a proteomics approach: Proteolytic regulation by APC/CCdh1. Molecular Biology of the Cell. 2008;19(3):1007-21.

61. Slawny L, O'Shea KS. Geminin promotes an epithelial-to-mesenchymal transition in an embryonic stem cell model of gastrulation. Stem Cells and Development. 2013;22(8):1177-89.

62. Emmett LS, O'Shea KS. Geminin is required for epithelial to mesenchymal transition at gastrulation. Stem Cells and Development. 2012;21(13):2395-409.

63. Seo S, Herr A, Lim JW, Richardson GA, Richardson H, Kroll KL. Geminin regulates neuronal differentiation by antagonizing Brg1 activity. Genes and Development. 2005;19(14):1723-34.

64. Fei Q, Boden SD, Sangadala S, Viggeswarapu M, Liu Y, Titus L. Truncated human LMP-1 triggers differentiation of C2C12 cells to an osteoblastic phenotype in vitro. Acta Biochimica et Biophysica Sinica. 2007;39(9):693-700.

65. Song L, Liu M, Ono N, Bringhurst FR, Kronenberg HM, Guo J. Loss of wnt/Beta-catenin signaling causes cell fate shift of preosteoblasts from osteoblasts to adipocytes. Journal of Bone and Mineral Research. 2012;27(11):2344-58.

66. Gambardella A, Nagaraju CK, O'Shea PJ, Mohanty ST, Kottam L, Pilling J, et al. Glycogen synthase kinase-3alpha/beta inhibition promotes in vivo amplification of endogenous mesenchymal progenitors with osteogenic and adipogenic potential and their differentiation to the osteogenic lineage. Journal of Bone and Mineral Research. 2011;26(4):811-21.

Chapter 5

1. Anderson L, Hunter CL. Quantitative mass spectrometric multiple reaction monitoring assays for major plasma proteins. Molecular and Cellular Proteomics. 2006;5:573-88.

2. Kuzyk MA, Smith D, Yang J, Cross TJ, Jackson AM, Hardie DB, et al. Multiple reaction monitoring-based, multiplexed, absolute quantitation of 45 proteins in human plasma. Molecular and Cellular Proteomics. 2009;8(8):1860-77.

3. DeSouza LV, Taylor AM, Li W, Minkoff MS, Romaschin AD, Colgan TJ, et al. Multiple reaction monitoring of mTRAQ-labeled peptides enables absolute quantification of endogenous levels of a potential cancer marker in cancerous and normal endometrial tissues. Journal of Proteome Research. 2008;7(8):3525-34.

4. Bertsch A, Jung S, Zerck A, Pfeifer N, Nahnsen S, Henneges C, et al. Optimal de novo design of MRM experiments for rapid assay development in targeted proteomics. Journal of Proteome Research. 2010;9(5):2696-704.

5. Sherman J, McKay MJ, Ashman K, Molloy MP. How specific is my SRM?: The issue of precursor and product ion redundancy. Proteomics. 2009;9(5):1120-3.

6. Sherman J, McKay MJ, Ashman K, Molloy MP. Unique ion signature mass spectrometry, a deterministic method to assign peptide identity. Molecular and Cellular Proteomics. 2009;8(9):2051-62.

7. Röst H, Malmström L, Aebersold R. A computational tool to detect and avoid redundancy in selected reaction monitoring. Molecular and Cellular Proteomics. 2012;11(8):540-9.