PHOSPHOPROTEOMICS OF *Arabidopsis thaliana*

By

CAMILLE STRACHAN

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2005

This is dedicated to my parents, John and Marjorie Strachan, who always gave me their loving support and encouragement; to my second parents, Leon and Daphney Strachan, for their love and support throughout the years; and to my brothers, Wayne, Ken and Stephen, for their love and friendship.

ACKNOWLEDGMENTS

This work would not have been possible without the help of many individuals, namely, my three supervisory committee members: Dr. James Winefordner, Dr. Nancy Denslow, and Dr. Alice Harmon. I thank Dr. Winefordner for welcoming me into his group and allowing me to do a collaboration project even though it was outside of his field. Despite the fact that this project was not in his area, he always made the effort to gain an understanding of proteomics. Being a part of his group truly meant a lot to me and it was an inspiration to see his determination to keep his research group family by planning group trips.

I thank Dr. Denslow for taking the time to meet with me as I tried to find my ideal project (application of mass spectrometry to a biological project) which at the time seemed impossible to find. She discussed the various projects available and welcomed me into her lab to begin what seemed an impossible but exciting task. I am truly grateful for her invaluable guidance as I entered the world of research, an area with which I had very limited experience. Her never-ending enthusiasm and encouragement always gave me hope that anything is possible; you just have to figure out the right approach. How many approaches do I have to try, to figure out the right one? Her unending enthusiasm showed me that in research you have to keep trying until you find it; and once you do, you'll be guaranteed one of her famous handshakes in congratulations! Her passion for science was contagious and I will always try to emulate it. The knowledge she imparted to me in our weekly group meetings was invaluable and for that I am also truly grateful. I

am also thankful to her for starting me on the collaboration project with Dr. Harmon, to whom I am eternally grateful.

Dr. Harmon's unending patience with me (as I tried to grasp the biological concepts behind the projects) and with my never-ending questions about biological sample preparation was greatly appreciated. During my initial stages of working in her lab, she actually took the time to perform several experiments with me. I am eternally grateful to her for those learning experiences and the privilege of performing experiments with her. Working in her lab was a tremendous pleasure as I watched her enthusiasm for research and her patience with her students as she explained various approaches and also performed various experiments with them. She always made time to sit with me and discuss any results or approaches that I had questions about. She was an advisor and also a mentor, always lending a listening ear and giving advice as I vented my frustrations with research or daily life. She was a true inspiration to me and has encouraged me to continue my career in research and development. I am eternally indebted to her.

I also owe my gratitude to members of the three groups that I am associated with. I thank the Winefordner group for their friendship and for always making me feel welcomed even though I was not based in the Winefordner lab. I am grateful to Dr. Ben Smith and Dr. Nico Omenetto for their input and interest in group meetings. I thank the Protein Core Facility for the use of its mass spectrometers and for always making me feel welcomed in the lab. Most especially I thank Dr. Stanley Stevens and Marjorie Chow. I thank Dr. Stevens for sharing his knowledge of mass spectrometry and database searching, as well as his renditions of catchy tunes during a frustrating day of lab work. I am truly grateful to Marjorie Chow for her tremendous help and support with sample-

preparation techniques, and for always giving me a listening ear as I vented about one thing or another. I thank Scott McMillen for MALDI-TOF-MS training. I extend my gratitude to Alfred Chung for giving me insight into synthesis and for always encouraging me to choose the best career path. I am truly grateful to Scott McLung for always ensuring that the QITMS was running optimally. Our friendly banter in the lab always brought a smile to my face, especially with the constant chocolate supply that Scott always provided. I also thank Margaret Joyner from the Harmon lab for her constant encouragement and reminiscing conversations, and for always ensuring that I had anything I needed in the lab to achieve my goals.

Last but not least, I would like to extend my gratitude to my family and friends for their unending love, support, encouragement, and patience. Each day I give thanks for my good fortune of having such a wonderful family. My mother's unending strength and determination are only a few of the qualities that I try to emulate. My father's organizational qualities and attention to detail I hope to one day be able to achieve. I wish he could be here to see what I have achieved. I know that he would be proud. The presence of newly made friends and colleagues made my time here pleasurable. Finally, I thank my Heavenly Father for giving me the ability to achieve all that I have achieved so far. Without Him I would never have gotten this far. He has given me the strength and perseverance to strive for excellence even when times were rough. If there is anyone I have left out, I thank them.

TABLE OF CONTENTS

page

LIST OF TABLES

LIST OF FIGURES

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

PHOSPHOPROTEOMICS OF *Arabidopsis thaliana*

By

Camille Nicola Strachan

August 2005

Chair: James D. Winefordner
Cochair: Alice C. Harmon
Major Department: Chemistry

Reversible protein phosphorylation on serine, threonine and tyrosine residues is one of the most common and important regulatory modifications of intracellular proteins, playing a role in many biological and biomedical phenomena such as cellular signal transduction, cell growth, cell differentiation, cell division, metabolism and cancer. Mass spectrometry has emerged as the method of choice for identifying phosphorylation sites in phosphopeptides because of its advantages over previous methods (high performance liquid chromatography separation of radiolabelled proteins with ^{32}P or ^{33}P followed by Edman degradation) including its increased sensitivity and speed, and because it eliminates the need for protein radiolabelling.

This research focused on the application of mass spectrometry to phosphoproteomic analyses of *Arabidopsis thaliana*. We demonstrated the application of mass spectrometry to four phosphoproteomic projects. Before working on these projects, we examined the development of methods for phosphorylation enrichment and analysis.

These methods were then applied to the various projects. Project 1, identifying autophosphorylation sites of a calcium-dependent protein kinase, demonstrated the use of several complementary methods for identifying numerous autophosphorylation sites of the protein. Project 2, identifying substrates of a calcium-dependent protein kinase from *Arabidopsis thaliana*, demonstrated the application of several newer technologies for identifying numerous substrates of the kinase. Project 3, identifying 14-3-3 interactors from *Arabidopsis thaliana*, examined the identification of numerous protein interactors, several of which were proven to be phosphorylated. These interactors were then shown to overlap with the substrates identified for the kinase, possibly indicating interaction between the two families of proteins. Finally, project 4, the application of robotic instrumentation was demonstrated as a means for high-throughput phosphoproteomic analysis, which resulted in identifying several phosphorylated proteins.

CHAPTER 1
INTRODUCTION

Reversible protein phosphorylation on serine, threonine and tyrosine residues is one of the most common and important regulatory modifications of intracellular proteins, playing a role in many biological and biomedical phenomena such as cellular signal transduction, cell growth, cell differentiation, cell division, metabolism, and cancer.[1, 2] Highlighting its importance is the fact that up to one-third of all proteins in a cell are phosphorylated at any given time, and as much as 5% of all the genes in a vertebrate genome code for enzymes involved in phosphorylation (kinases) or dephosphorylation (phosphatases).[2] Due to its importance, research has been initiated in many areas of biomedical research towards the understanding of the regulatory properties of protein phosphorylation. Included in these studies are the investigation of the function of protein phosphorylation in cell cycle regulation, enzyme activation/deactivation, and protein-protein association.

Calcium-Dependent Protein Kinases

In eukaryotes, protein kinases regulate key aspects of cellular function (such as metabolism) and responses to external signals by catalyzing the transfer of the terminal group of ATP to seryl or threonyl residues in a variety of protein substrates.[3] Recent mapping of the Arabidopsis genome provides the first opportunity to identify all the protein kinases present in a plant model and to begin to understand their physiological roles. The Arabidopsis genome encodes 1085 typical protein kinases, which is about 4% of the predicted 25,500 genes.[4] These plant kinases differ from animal kinases, since in

plants they phosphorylate only serine and threonine residues; while in animals, tyrosine is the predominant residue that is phosphorylated. Moreover, a number of kinase families in plants are either not found in animals or yeast, or are highly divergent. Some of these are calcium-dependent protein kinases (CDPK**s)** found in vascular and nonvascular plants, green algae, and certain protozoa (ciliates and apicomplexans).[5] These enzymes are proposed to be involved in all aspects of plant development and physiology, and participate in the coupling of cellular responses to environmental and developmental signals.[6]

Regulation of the CDPK kinase activity depends on calcium signaling and possibly autophosphorylation on Ser/Thr residues of the kinase itself; however, the regulatory effects of autophosphorylation still remain unclear.[7-9] Autophosphorylation of a CDPK from groundnut was suggested to be a prerequisite for its activation,[10] while inhibition of activity was seen after autophosphorylation of a winged bean.[11] On the other hand, preautophosphorylation of a CDPK from sandalwood had no effect on kinase activity.[12] Also, conflicting results were seen for a CDPK from ice plant whereby mutation of either one of the two autophosphorylation sites increased activity, but mutations at both residues dramatically decreased activity.[9] This regulatory process needs further study to learn if common autophosphorylation sites exist among this family of kinases that will eventually lead to a greater understanding in the role of this mechanism.

Another aspect of CDPKs that is not understood and requires further investigation is their recognition of substrate proteins. It is expected that CDPKs have access to hundreds of potential substrates in the cytosol and nucleus since they are found as both soluble and membrane-anchored isoforms.[13] The expectation is that most isoforms will be

found associated with membranes, such as the plasma membrane,[14] peroxisomes,[14] endoplasmic reticulum,[15] seed oil bodies,[16] and mitochondria.[17] In addition to their widespread subcellular distribution, there is evidence that some CDPKs can change locations in response to a stress treatment.[9] This was seen when an isoform McCPK1 from ice plant (*Mesembryanthemum crystallinum*) was tagged with a fluorescent protein and transiently expressed in leaves. The tagged McCPK1 showed a pronounced shift in localization from the plasma membrane to the nucleus in response to a salt or dehydration stress, indicating that CDPK targeting is dynamic.

Again, very few substrates are known. Knowledge of the mechanism that these enzymes use to recognize their diverse substrate proteins is even more limited. Typically, the sites phosphorylated by a particular protein kinase share a set of common sequence elements (its consensus sequence) whose existence is necessary and sufficient for recognition by that enzyme.[18] These common sequence elements refer to the sequence elements immediately surrounding the site(s) phosphorylated by the kinase, generally taking the form of a short linear sequence of amino acids. According to Kennelly and Krebs,[18] several assumptions are implied in the formulation of a consensus sequence: 1) The existence of a consensus sequence on a protein is essential and adequate for its recognition as a substrate by a particular protein kinase. 2) The specificity-determining feature of the phosphorylation site is contained in a neighboring sequence of amino acids around the phosphoacceptor, not including elements from different polypeptide chains or from widely scattered portions of a single polypeptide chain. 3) Not all sequence positions surrounding the phosphoacceptor group carry equal weight in determining the recognition code.[18]

Summarizing the complexities of the substrate-recognition process as a set of short recognition sequences has its usefulness in its simplicity, which has facilitating the evaluation and application of a large body of observations. However, this sequence can be an oversimplification that can lead one to think that the primary sequence alone controls recognition; when in fact, factors such as secondary/tertiary structure or distant secondary recognition sites play a significant role in substrate recognition.[18, 19] The secondary/tertiary structure of the protein may actually determine substrate specificity by denying access to potential phosphoacceptor groups. This means that the existence of intricate secondary/tertiary structures could be an important key to substrate recognition. That is, the more complex the determinants, the more discriminating the kinase. The presence of a consensus sequence does not assure that a protein is a substrate of the kinase, but instead functions as a guide whose implications must be confirmed.

In early studies of CDPK substrate specificity, two simple phosphorylation motifs were recognized; Basic$_{-3}$-x-x-[S/T]$_0$ and S$_0$-x-Basic$_{+2}$. To date, in-depth analyses have resulted in the reporting of four consensus sequences for CDPKs with some differences apparent among the isoforms: 1) φ_{-5}-x-Basic$_{-3}$-x-x-S$_0$-x-x-x-φ_{+4} (minimal) or Basic$_{-6}$-φ_{-5}-x-Basic$_{-3}$-x-x-S$_0$-x-x-x-φ_{+4}-Basic $_{-6}$ (optimal), 2) [Basic$_{-9}$-Basic$_{-8}$-x$_{-7}$-Basic$_{-6}$]-φ_{-5}-x-x-x-x-[S/T]$_0$-x-Basic$_{+2}$, where the exact ordering of residues within brackets is not specified, 3) φ_{-1}-[ST]$_0$-φ_{+1}-X-Basic$_{+3}$-Basic$_{+4,}$ and 4) [AL]$_{-5}$-X$_{-4}$-R$_{-3}$-X$_{-2}$-X$_{-1}$-S$_0$-X$_{+1}$-R$_{+2}$-Z$_{+3}$-R$_{+4}$, where φ is a hydrophobic residue, x is any amino acid, and Basic is a basic amino acid residue (K or R), and Z is any residue but R or K.[20-26] Because individual CDPK isoforms can, in general, recognize all four motifs, it appears that CDPKs may have a series of

overlapping but non-identical polypeptide binding grooves that can accommodate the different sequences.

Studies to determine these CDPK motifs have largely been performed by using synthetic peptides. However, although these peptides have represented powerful investigative tools, their small size and random conformation significantly limit their abilities to mimic the proteins they are intended to model. Therefore, using proteins to identify substrates for new phosphorylation sequences could help define primary structural determinants of protein kinase specificity.

Protein extracts are usually in the denatured form, making it more difficult to interpret physiological relevance of results. Bylund and Krebs[19] showed that phosphorylation may increase with the unfolding of the protein substrate. Native lysozyme (which was not a substrate for the cyclic AMP-dependent protein kinase of rabbit skeletal muscle) became susceptible to phosphorylation by the enzyme once the protein was denatured by heating.[19] Therefore, many proteins may contain sites that can be phosphorylated once they have been exposed. Careful interpretation is therefore needed for protein phosphorylation reactions observed *in vitro*, since denatured proteins may become protein kinase substrates even though they were not substrates in their native state. Additionally, the mixing of proteins and kinases from different subcellular compartments may lead to phosphorylation of proteins that would not occur *in vivo*.

Despite these complications, the identification of phosphorylation motifs is of fundamental importance and may be useful for functional genomics and prediction of phosphoproteins. There are at least 34 CDPK isoforms in Arabidopsis, some of which have been implicated in drought stress, pathogen response pathways, and the regulation

of metabolic enzymes, transport proteins, and cell structure.[7] Consequently, it is of fundamental importance to understand how this important group of Ser/Thr-kinases targets their substrate proteins. So even though these results might not signify physiological occurrences, they are important in terms of fundamental information, as the identification of new motifs may aid in the understanding of how these kinases target their substrates.

14-3-3s

Another family of proteins associated with phosphorylation is the 14-3-3 proteins. These proteins were first identified as abundant brain proteins that were isolated as soluble, cytosolic, and acidic proteins.[27, 28] Naming of these proteins was given according to their particular migration pattern on two-dimensional diethylaminoethyl cellulose (DEAE-cellulose) chromatography and starch gel electrophoresis. The proteins were then named by Greek letters according to their respective elution positions on HPLC. Further studies showed that these proteins were also present in all eukaryotic organisms examined to date, existing as protein families that contain highly conserved, but individually distinct isoforms.[29-31] Of the organisms characterized, Arabidopsis has the largest family (10 distinct 14-3-3 proteins of 248 to 268 amino acids: GF14ψ, GF14χ, GF14φ GFl4ω, GF14υ, GF14τ, GF14ε, GF14κ, GF14μ, and GF14ν).[28]

Members of the 14-3-3 family are homo- and heterodimers, whose L-shaped monomers come together to form a broad central groove that contains two binding sites for target proteins. Phosphorylation of these binding partners (and possibly also the 14-3-3 proteins themselves) may be important in regulating these interactions.[32, 33] To date, two different binding motifs have been identified in nearly all known 14-3-3 binding proteins, RSXpSXP and RXY/FXpSXP (where ‘X’ denotes ‘any amino acid

residue', R represents a basic residue, and pS denotes phosphorylated serine); and other binding motifs have been discovered, including unphosphorylated sites on a few proteins.[32, 34, 35]

Binding of these phosphoserine-containing proteins with 14-3-3 proteins implicated 14-3-3s as proteins that mediate interaction among diverse components of many biological activities. In mammals, most of the known 14-3-3 binding proteins are components of intracellular signalling pathways. In contrast, 14-3-3s in plants have emerged as important regulators of phosphorylated enzymes of biosynthetic metabolism, ion channels and regulators of plant growth.[36] While many proteins have been identified to bind with 14-3-3s, little is known about how many targets exist in plants. Also, since there are only two currently known phosphoserine-containing binding motifs among target proteins, it would be of interest to discover whether there are more 14-3-3 binding motifs, and to determine if some of these targets are also targets of CDPKs. As mentioned earlier, the common use of denatured proteins would actually be of benefit for these studies since crystal structures of a 14-3-3ζ:phosphopeptide complex showed that the phosphopeptides bind in an extended conformation, thus resulting in a greater chance of finding novel binding motifs.[32]

Detection and Analysis of Protein Phosphorylation

To understand processes regulated by phosphorylation on the molecular level, one must first determine which proteins are phosphorylated, and second identify the exact sites of phosphorylation. However, as Mann et al.[37] mentions, analysis of phosphoproteins is not straightforward for several reasons. First, the stoichiometry of phosphorylation is generally relatively low; that is, only a small fraction of the available intracellular pool of a protein is phosphorylated at any given time as a result of a given

stimulus. Second, the phosphorylated sites on proteins might vary, implying that any given phosphoprotein is heterogeneous; that is, it exists in several different phosphorylated forms. Third, many of the signaling molecules are present at low abundance within cells and in these cases, enrichment is a prerequisite. Fourth, most analytical techniques used for studying protein phosphorylation have a limited dynamic range; which means that, although major phosphorylation sites might be located easily, minor sites might be difficult to identify. Fifth, phosphatases present in cell lysates could dephosphorylate residues unless precautions are taken to inhibit their activity during preparation and purification steps. Finally, proteins can also be phosphorylated by kinases which are already present during extraction and purification, so this needs to be prevented.

The analysis of phosphorylation usually proceeds in a certain order, with the detection of the phosphorylated protein coming first. This may be done by using either radiolabeling, antibodies, or fluorescent labeling; and analysis is done by using chromatographic methods [high-performance liquid chromatography (HPLC), thin-layer chromatography TLC)], electrophoresis, western blotting, autoradiography, scintillation counting, or mass spectrometry (MS).[1] Recently, a method for selectively staining phosphorylated proteins separated by polyacrylamide gel electrophoresis was developed. This proprietary fluorescent stain (Pro-Q® Diamond Phosphoprotein Gel Stain) developed by Molecular Probes allows direct, in-gel detection of phosphate groups attached to serine, threonine, and tyrosine residues, without the need for radiolabelling or antibodies.

Another way to detect phosphoproteins is to screen for the presence of individual phosphopeptides. This can be done by first partially hydrolyzing the labeled or unlabelled protein by partial acid, enzymatic, or alkaline hydrolysis of the amide bonds of the protein to release the phosphopeptides. The resulting phosphopeptides then must be separated from other peptides by using either thin-layer chromatography (TLC), electrophoresis (one or two dimensional), or high-performance liquid chromatography (HPLC). Retention time, mass spectrometry, antibody recognition, or amino acid sequencing (Edman, MS/MS) can then be used for identification.

Recently, mass spectrometry-based methods have become increasingly popular for analyzing phosphopeptides resulting from proteolytic digests of phosphorylated proteins because of their increased sensitivity and speed, and because they remove the need for protein radiolabelling.

Mass Spectrometric-based Methods

Mass spectrometry is becoming the method of choice for analyzing complex protein mixtures. Routine analysis of biomolecules, particularly proteins and peptides, was made possible by the advent of two mass spectrometry ionization tools: matrix-assisted laser desorption/ionization (MALDI) and electrospray ionization (ESI). Since 1988 when MALDI and ESI mass spectrometry were first proven useful for analyzing peptides, proteins, carbohydrates, and oligonucleotides, they have become the MS methods of choice for biopolymer analysis. Our study used three mass spectrometers: a matrix-assisted laser desorption/ionization time-of-flight mass spectrometer (MALDI-TOF-MS), a nanospray quadrupole ion trap mass spectrometer (nanoESI-QIT-MS), and a tandem quadrupole time-of-flight mass spectrometer with interchangeable MALDI and nanospray sources (MALDI or ESI-QqTOF-MS).

Ionization techniques

Matrix-assisted laser desorption/ionization (MALDI), a "soft ionization" technique, was first described in the late 1980s as a technique used with mass spectrometry for analyzing large, polar, nonvolatile molecules.[38] In this technique, a solid organic matrix compound that is strongly UV-absorbing at the designated wavelength, is dissolved in an appropriate solvent and mixed with the solution of the sample of interest (the analyte). A 0.5-3 μL aliquot of this solution is then placed on a stainless-steel target plate and allowed to dry. On drying, the analyte is co-crystallized with a large (10^4) molar excess of the solid matrix material. Once in the mass spectrometer (typically time-of-flight), the sample is then irradiated with a pulsed laser beam [usually a nitrogen laser in the ultraviolet range (337 nm)] for desorption and ionization of the analyte molecules.

Even though the mechanism by which MALDI operates is still unclear, it is agreed that the matrix is critical and fills several roles. First, using the large excess of matrix helps to isolate analyte molecules from each other, thereby reducing intramolecular interactions. Second, the matrix absorbs large amounts of energy from the incoming photons of the pulsed laser beam, resulting in an explosive breakdown of the matrix-analyte lattice, sending both matrix and analyte molecules into the gas phase. Third, the matrix is necessary for ionization of the analyte because it transfers protons to the analyte via gas-phase reactions in the dense cloud that forms.[39] The desorption/ionization process is shown in Figure 1-1.[39] Matrix choice depends on the irradiance wavelength and the type of sample being analyzed. Common matrices that are used with N_2 lasers operating in the UV at 337 nm are 3,5-dimethoxy-4-hydroxy-*trans*-cinnamic acid (sinapinic acid),

2,5-dihydroxy-benzoic acid (DHB), and α-cyano-4-hydroxy-trans-cinnamic acid (α-cyano) (Figure 1-2),[40, 41] with the latter two being best for peptides.

Electrospray ionization, another "soft ionization" technique, was first introduced in the late 1980s. In this technique, ions are formed from peptides and proteins by spraying a dilute solution of the analyte (typically dissolved in a mixture of water, an organic modifier such as acetonitrile, and a few percent by volume of a volatile acid) from a fine tip at atmospheric pressure. Generally, a high electric field is created by applying a high voltage to either the spray tip or the counter electrode, resulting in a fine mist of droplets that are highly charged. This generated electric field (E) between the spray tip and counter electrode is expressed by the equation:

$$E = (2V/r) \ln(4d/r)$$

where V is the voltage applied, r is the radius of the needle, and d is the distance between the spray tip and counter electrode. This imposed electric field will also penetrate to the liquid flowing through the needle, causing ions in solution to move toward the liquid surface. Accumulation of these charges at the surface then leads to destabilization of the surface because the ions at the surface are drawn to the counter electrode yet can't escape, resulting in the formation of a Taylor cone (Figure 1-3). Due to the cone's instability (influenced by the surface tension of the fluid γ), charged droplets are emitted. The onset voltage (V_{on}) required to initiate charged-droplet emission is related to surface tension by the equation:

$$V_{on} = 2 \times 10^5 (\gamma r)^{0.5} n (4d/r)$$

The radius (R) of the emitted droplets will depend on the fluid density (ρ), flow rate (V_f), and surface tension (γ), given by the relationship:

$$R \propto (\rho V_f^2 \gamma)^{1/3}$$

Thus, the higher the flow rate (V_f), the larger the initial droplet size which leads to lower ionization efficiency because the droplets are not so close in size to the Rayleigh limit (will be discussed later). On the same note, the lower the flow rate (as in nanospray), the smaller the droplet size, the higher the surface-to-volume ratios leading to a larger amount of the analyte available for ionization, thus a higher ionization efficiency.

Once these small droplets are formed and are accelerated toward the counter-electrode, solvent rapidly evaporates and the analytes (peptides or proteins) in the droplets pick up one, two, or more protons from the solvent to form singly or more frequently, multiply charged ions (for example, $[M+H]^{+1}$, $[M+2H]^{+2}$, $[M+3H]^{+3}$, etc.). As the solvent evaporates and the droplet shrinks, the charge density on the surface increases to the point where Coulombic charge repulsion overcomes the forces holding the droplet and solvated ions together (Rayleigh Limit), leading to disintegration of the droplet into smaller droplets (Figure 1-4). This limit where the Coulombic explosions begin is given by:

$$Q^2 = 64\pi^2 \varepsilon_0 \gamma R^3$$

where ε_0 is the permittivity of vacuum. Once the ions are "emitted" or "evaporated" from the droplet surface, the ions are then sampled into the high-vacuum region of the mass spectrometer for mass analysis and detection, most often using a quadrupole mass analyzer.

Overall, the widespread acceptance of ESI-MS and MALDI-MS are accountable for several reasons. Both methods are usually combined with relatively low-cost, easily operated mass spectrometers. They both offer high sensitivity (picomole-femtomolar

range), accuracy (± 0.01%), and capability of analyzing molecules with a wide molecular weight range.[39] Even though both techniques have many common capabilities, they do have their own unique capabilities that are of practical significance. Although both techniques work best with clean (salt- and detergent-free) samples, MALDI is more tolerant of many of the common biological buffers, that is, information can be obtained by MALDI directly from a relatively dirty sample. Also, even though both methods can provide molecular weight information for large proteins, MALDI-MS is more sensitive and provides the information more easily. However, in the lower molecular weight range, ESI-MS usually provides more accurate molecular weight measurements as well as better mass resolution. Additionally, since MALDI produces predominantly singly charged molecular ions from peptides and proteins, analysis of the resulting MALDI-MS spectrum is very straightforward. Finally, since ions in electrospray are produced at atmospheric pressure from flowing liquid streams, ESI is ideally suited for on-line coupling to high-performance liquid chromatography, making it possible to analyze mixtures of peptides and proteins. With all of these considerations in mind, it can be said that MALDI and ESI are methods that can be used to complement each other.

Mass analysis

With the advent of MALDI and ESI, peptide and protein analysis have been a large focus of efforts in mass spectrometry over the last 17 years, bringing many types of mass analyzers into use in this area of research, namely time-of-flight, quadrupole ion trap, and triple quadrupole mass spectrometers, primarily because of their cost and ease of use.

Time-of-flight mass analyzers are among the simplest of the mass analyzers.[42] The principle of time-of-flight mass spectrometry involves measuring the time required for an

ion to travel from the ionization source to the detector.[43] For a simple TOF-MS, there are three components, an ionization source (typically MALDI), a field-free drift region, and a detector (Figure 1-5). Upon ionization, ions are accelerated out of the source under the influence of a strong electric field. Even though all the ions receive the same kinetic energy during acceleration at the ionization source, as they traverse the field-free region they separate into groups according to velocity because they may have different *m/z* values. These ions then sequentially strike the detector in order of increasing *m/z* value (lighter ions arriving first), upon which the time-of-flight analyzer converts the time-of-flight of the ions to a mass-to-charge ratio using the equations:

$$E = \tfrac{1}{2} mv^2$$

$$t = L/v = L\,[m / 2zeV]^{1/2}$$

therefore, $m/z = 2t^2eV/L^2$

where t = time of flight (seconds), L = length of flight tube (m), v = velocity (m/s), m = mass (kg), and z = charge.

Simple linear mass spectrometers as described above are somewhat limited due to their rather low resolution. This low resolution is partly due to the initial kinetic energy spread of individual ion populations, that is, various members of the same ion population will arrive at the detector at slightly different times. An effective way to correct for this energy spread is through the use of a reflectron (ion mirror) which acts as an energy-focusing device. When an ion reaches the reflectron it is slowed down until it is stopped by a voltage that is applied to the back end of the reflectron, the ion is turned around, and then reaccelerated out to a second detector at a slightly different angle so the flight path of the reflected ions does not cross with the ions entering the reflectron (Figure 1-6). In

the case of an ion population with a spread of slightly different kinetic energies, ions with a slightly lower energy will not penetrate the reflectron as deeply, therefore turning around more quickly and catching up to those ions with full kinetic energy. While ions with slightly greater kinetic energies will penetrate more deeply, turn around more slowly, and have their flight times retarded, allowing the other ions to catch up. The addition of this reflectron will then cause ions of a given *m/z* to be spacially focused into packets with flight times that are closer together. The addition of the reflectron also increases the flight path for an ion without increasing the size of the flight tube, also resulting in an improvement in resolution by enhancing the time dispersion of ions of different *m/z*. Another improvement that has been made for achieving better resolution on MALDI-TOF instruments is time-lag focusing or "delayed extraction" in the ionization source. In this source, ions are created in a field-free region and allowed to spread out before extraction and acceleration voltages are applied.

Quadrupole ion trap mass spectrometers are mass analyzers that operate by trapping ions in a three-dimensional electric field consisting of two end-cap electrodes and a ring electrode, each having a hyperbolic geometry (Figure 1-7).[42, 44, 45] In the normal mode of use, the end-cap electrodes have an auxiliary oscillating potential of low amplitude applied while the ring electrode has an RF oscillating drive potential of 1 MHz, resulting in the creation of a potential well (quadrupolar trapping field). This field can be described as having a saddle shape that is constantly spinning whereby the field at any particular point in time will possess this saddle shape. For an ideal quadrupole field to be generated, the mathematical relationship presented below has to be fulfilled:

$$r_0^2 = 2z_0^2$$

where r_0 is the radius of the ring electrode in the central horizontal plane and z_0 is the separation of the two end-cap electrodes measured along the axis of the ion trap. Typically, once the magnitude of r_0 is given, the sizes of all the electrodes and electrode spacings are fixed. It should be noted that the majority of commercial ion traps in use today have r_0 at either 1.00 or 0.707 cm.

An ion's stability in this quadrupolar trapping field is dependent upon its *m/z*, the potentials applied to the electrodes, and the internal dimensions of the ion trap electrodes. An ion that is stable in this field will possess a trajectory that has the appearance of a Lissajous figure, allowing it be trapped within the specific electric field of the ion trap (Figure 1-8). Unstable ions will have trajectories that increase in magnitude as they near the ring of the endcaps, resulting in their collision with the electrodes. Determination of whether the trajectory of an ion will be stable or unstable under defined conditions of the electric field may be calculated with the Mathieu equations:

$$a_z = -2a_r = \frac{-16eU}{m\Omega^2(r_0^2 + 2z_0^2)} \qquad (1)$$

$$q_z = -2q_r = \frac{8eV}{m\Omega^2(r_0^2 + 2z_0^2)} \qquad (2)$$

where a_z and q_z are two reduced parameters, r symbolizes the radial direction, z symbolizes the axial direction, e is the charge of an ion, U is the DC amplitude applied to the ring electrode, V is the RF amplitude, m is the mass of an ion, Ω is the angular drive frequency ($2\pi f_{rf}$), r_0 is the radius, and z_0 is distance from the center to the end-cap. It should be noted that an ion has to be stable in both the r and z directions to be confined within the trap, thus, a_r and q_r parameters also have to be considered:

$$a_r = \frac{8eU}{m\Omega^2(r_0^2 + 2z_0^2)} \qquad (3)$$

$$q_r = \frac{-4eV}{m\Omega^2(r_0^2 + 2z_0^2)} \qquad (4)$$

The resulting stable trajectories that are calculated from the operating parameters $a_{r,z}$ and $q_{r,z}$ can be displayed graphically as the Mathieu stability diagram (Figure 1-9), whereby the region of stability is defined by the boundaries at $\beta_z=0$, $\beta_z=1$, $\beta_r=0$, and $\beta_r=1$. This means that if an ion has an a_z and q_z within this region, it will be stable in both the r and z directions and will be trapped in the ion trap.

Typically, ions generated by electrospray ionization (external source) are focused into the ion trap using electrostatic lenses. Once in the ion trap, collisions with helium buffer gas at a pressure of 1 mTorr dampen the kinetic energy of the ions and contract trajectories toward the center of the trap, where a range of m/z values are held in stable orbits by the RF potential. As the RF potential on the ring electrode is increased, the ions become more energetic and develop unstable trajectories along the axis of symmetry (the z-axis), then in order of increasing m/z value, ions exit the trap through holes in the end-cap electrodes to a detector. As the RF amplitude is ramped and ions are ejected to the detector one at a time, a mass spectrum is generated; usually several such spectra (microscans) are obtained in succession and are then summed prior to display and recorded as a macroscan. A detailed description of quadrupole ion traps has been discussed by March if additional information is necessary.[44, 45]

Along with measuring the m/z values of ions introduced to the mass spectrometer, quadrupole ion trap mass spectrometers can also be used to obtain detailed structural information from these ions. This information is obtained by performing multiple mass-selective operations, one after another, that is, tandem mass spectrometry (MS/MS). The first mass-selective operation is used for the isolation of the ion species of interest

(designated as the parent ion), and the second is used to determine the mass/charge ratios of the fragment ions (product) formed by collision-induced dissocation (CID) of the isolated ion of interest. CID refers to the process whereby the kinetic energy of the selected ion population is increased by applying a voltage resonant with the frequency of the precursor ion, causing more energetic collisions with the He bath gas. Subjecting the ions to many hundreds of low-energy collisions will ultimately increase the internal energy of the ion until fragmentation occurs.

Another mass spectrometer also possessing the ability to perform MS/MS analysis that was available for this project was a tandem quadrupole time-of-flight mass spectrometer (QqTOF-MS), where Q refers to a mass-resolving quadrupole and q refers to an r.f.-only quadrupole or hexapole collision cell. This configuration can be regarded as either the addition of a mass-resolving quadrupole and collision cell to an ESI-TOF, or the replacement of the third quadrupole (Q3) in a triple quadrupole by a TOF mass spectrometer,[46] with the latter being the simplest description for the purpose of describing the basic principles. A thorough review of triple quadrupole instruments has been described by Yost and Boyd.[47]

A typical QqTOF configuration consists of three quadrupoles, Q0, Q1 and Q2, followed by a reflecting TOF mass analyzer with orthogonal injection of ions (Figure 1-10). Q0 provides collisional damping, Q1 acts as a mass filter, and Q2 is a collision cell. In the case when ions are provided by a high-pressure electrospray source, Q0 acts as an ion guide with collisional cooling and focusing of the ions as they enter the instrument. When single MS (or TOF-MS) measurements are required, the mass filter Q1 is operated in the r.f.-only mode serving as only a transmission element while the TOF analyzer

records spectra. For MS/MS, Q1 is operated in the mass filter mode to only transmit the parent ion of interest which then gets accelerated to an energy of between 20 and 200 eV before it enters the collision cell Q2, where it is subjected to CID and subsequently collisionally cooled and focused before analysis by the TOF mass analyzer.

Mass analysis of peptides and proteins

In a generic mass spectrometry-based experiment, the typical proteomics experiment consists of 5 stages. In stage 1, proteins are isolated from cell lysates or tissues with gel electrophoresis typically used as a method for biochemical fractionation. Since MS of whole proteins are less sensitive than MS of peptides and the mass of the intact protein is usually insufficient for protein identification, proteins from stage 1 are typically enzymatically digested with trypsin for the formation of C-terminally protonated amino acid peptides (stage 2). These peptides are then separated by high-performance liquid chromatography followed by ESI (stage 3) whereby ions are introduced to the mass spectometer for mass analysis, producing an MS spectrum (stage 4). The computer then generates a prioritized list of these peptides for fragmentation and a series of MS/MS experiments ensues (stage 4). These MS and MS/MS spectra are then used for matching against a known protein database for the identification of the proteins (stage 5).[48]

Stage 1, the sample extraction and preparation step, may be considered the most critical step in any proteomics study.[49] In this regard, proteomic analysis of plant tissues is even more problematic than other organisms due to the involvement of several challenges. These include the fact that plant tissues typically have low amounts of proteins, they are often rich in proteases and materials that may severely interfere with downstream protein separation and analysis, including cell wall and storage

polysaccharides, lipids, phenolic compounds and a broad array of secondary metabolites.[49] These contaminants pose a serious problem for one of the most commonly used separation techniques in proteomics, 2-dimensional gel electrophoresis. Their presence can result in horizontal and vertical streaking, smearing, and a reduction in the number of distinctly resolved protein spots. In order to alleviate this problem, several protein extraction techniques for plant tissues have been compared by Saravanan and Rose, whereby the quantitative and qualitative characteristics of the protein extracts were examined.[49] From this study, it was demonstrated that the phenol-based method gave the greatest protein yield and the least contaminants.

Once protein extraction from the plant tissue has been performed, the proteins are typically separated by 1- or 2-dimensional gel electrophoresis (GE), which are techniques that separate proteins by the application of an electric field. SDS-PAGE (sodium dodecyl sulfate-polyacrylamide gel electrophoresis) is a 1D-GE technique that separates proteins according to their molecular weights. 2D-GE involves the separation of proteins according to their isoelectric point (the pH at which a protein carries no net electric charge) in the first dimension, and separation according to MW in the second dimension. Once the proteins are immobilized in the gel, they are typically visualized by a gel staining method.[50, 51] Some gel stains are visible in visible light, others which are fluorescent stains that require visualization by UV light. Typical fluorescent imaging devices are CCD camera-based systems or laser scanner systems.[52, 53]

Once the proteins are visualized and the proteins of interest are determined, the gel bands/spots are excised for enzymatic digestion.[54] Prior to digestion, protein stains have to be removed to prevent later interference of enzyme activity or mass spectrometric

analysis. Also, the presence of sulfhydryl-containing amino acids in a protein may result in the formation of disulfides bonds along the protein backbone resulting in the formation of a three-dimensional shape. As a result, inner portions of the 3D protein structure may be inaccessible to the enzyme, thus unraveling of this structure is necessary for complete enzyme cleavage of the protein. This is achieved by the reduction of the sulfhydryl groups with a reducing agent such as dithiothreitol (DTT), and subsequent alkylation to prevent reforming of the disulfide bond, resulting in the linearization of the protein. At this point, an enzyme is added for protein cleavage. Trypsin, a commonly used enzyme, cuts the protein at the C-terminal end of lysine and arginine residues, resulting in the formation of C-terminally protonated amino acid peptides for mass spectrometric analysis.

As mentioned earlier, the routine analysis of proteins and peptides has become possible due to the introduction of MALDI and ESI techniques. Typically, MALDI-MS is used for the measurement of peptide masses in an enzymatic digest. ESI-MS is typically coupled with on-line reversed-phase liquid chromatography systems for the separation of the peptides prior to introduction to the mass spectrometer, upon which the peptide masses are measured and tandem MS of the peptides is performed for structural information.

The tandem MS utility available on two of the mass spectrometers described above, the QIT-MS and QqTOF-MS, enables immense structural information to be obtained on proteins and peptides. Advancing this utility even more is the incorporation of decision-making algorithms that can automatically perform MS/MS experiments on precursor ions from a previously acquired full scan, enabling the instrument to make real-time decisions

concerning the experiment at hand. The use of this algorithm is called data-dependent analysis. A typical example of its use would be for peptides eluting from an HPLC separation. As peptides enter the ion trap mass spectrometer, a full scan is obtained. Once an ion is detected above a preset threshold, the mass spectrometer will automatically switch from full scan mode to MS/MS mode on that ion. If there are coeluting peptides, the mass spectrometer will perform MS/MS on the most intense ion, this ion would then be placed on an exclusion list, and the second most intense ion from the full scan would then be subjected to MS/MS. So as each ion gets subjected to MS/MS analysis it is placed on an exclusion list and will not be removed from this list until after a user-defined length of time.

As peptides undergo fragmentation by low-energy gas phase collisions during MS/MS, they undergo cleavage at the amide bonds (-CO-NH-) that join pairs of amino acid residues, generating a ladder of sequence ions.[55] If the charge is retained on the N-terminal end after cleavage of the amide bond, b-type ions are formed. However, if the charge is retained on the C-terminal end, y-type ions are formed. The most commonly observed fragment ions and their nomenclature are shown in Figure 1-11A. The determination of the amino acid sequence becomes possible if a complete series of either one or both ion types are present for the subtraction of the masses of adjacent sequence ions.

If a phosphate group is present, as with phosphorylated peptides, fragmentation occurs differently, that is, very little cleavage occurs along the peptide backbone (amide bonds) as would be expected. Instead, β-elimination (removal of the phosphate moiety) occurs as the primary fragmentation with the phosphate moiety being lost as either

H_3PO_4, HPO_3, $H_2PO_4^-$, HPO_4^{2-}, PO_4^{3-}, or PO_3^-, depending on the ionization mode (positive or negative) being used (Figure 1-11B). These signature losses from phosphopeptides can be used as an indicator to determine whether a peptide is phosphorylated. On the QIT-MS, a reconstructed chromatogram can be created by the software after MS/MS analysis to locate all peptides in a chromatogram that lost a neutral fragment (H_3PO_4 or HPO_3) (a reconstructed neutral fragment chromatogram). However, sequencing of the peptide is generally difficult due to insufficient fragmentation. A similar feature available on the QqTOF-MS is precursor ion scanning, whereby peptides losing the precursor ion ($H_2PO_4^-$, HPO_4^{2-}, PO_4^{3-}, or PO_3^-) are recorded by the TOF-MS in the negative ion mode. The drawback of this method is that the experiment has to be repeated in the positive ion mode so that peptide sequencing may be obtained because fragments created in the negative ion mode are very difficult to interpret.

Data interpretation

Once mass spectra are obtained, protein identification is performed according to the method of analysis. For MALDI-MS, peptide mapping, often referred to as peptide-mass mapping or peptide-mass fingerprinting is performed (Figure 1-12). In this method, proteins are identified by matching the list of experimental peptide masses to a list of predicted peptide masses that would occur after digestion with a specific enzyme of all entries in the protein database.[42, 48] A match is generally found if a sufficient number of peptide ions are matched and there are not more than two proteins present. Typically, as the size of the database being searched is increased, the level of uncertainty also increases due to the possibility of more proteins being present that could generate peptides with similar masses.

If tandem MS was performed, identification of the protein is more clear-cut because not only is the mass of the peptide known, but the peptide sequence information is also available from the peak pattern of the CID spectrum (Figure 1-13). The CID spectra are then scanned against a comprehensive protein sequence database using one of a number of different algorithms, each with its own strengths and weaknesses.[48, 56-60] Two of the most commonly used methods are cross-correlation and probability-based matching.

In the cross-correlation method used in the SEQUEST search engine, theoretical tandem mass spectra are constructed for all peptide sequences in the protein database matching the mass of the experimental peptide and the overlap or cross-correlation of the predicted spectra with the experimental spectra is used to determine the best match.[42, 56-58] The quality of the match between the sequence and the spectrum is indicated by the magnitude of the cross-correlation value and the quality of the match versus all the other top ranking sequences in the database is shown by the difference between the normalized cross-correlation score to the second ranked sequence. In the probability-based matching method used in the MASCOT search engine, calculated fragments of peptides in the database are compared with observed peaks and a score is calculated that reflects the statistical significance of the match between the spectrum and the sequences contained in a database. There are three advantages to this approach:[60] 1) A simple rule can be used to judge whether a result is significant or not, guarding against false positives. 2) Scores can be compared with those from other types of searches. 3) Search parameters can be readily optimized by iteration. For each of these methods, the identified peptides are compiled into a protein hit list with corresponding scores or statistics.

The introduction of these search engines have provided a means for handling the enormous amounts of CID spectra that can be produced from data-dependent scans, resulting in high-throughput proteomics. However, because protein identifications rely on the matches with sequence databases, high-throughput proteomics is limited to those species having a comprehensive sequence database available.

Unfortunately, identification and mapping of phosphorylated peptides by tandem mass spectrometry is not as straightforward as described above for several reasons.[61] First, cleavage of the protein by trypsin can be inhibited due to the negatively charged modifications of the phosphate group, resulting in incomplete peptide coverage. Secondly, phosphorylation is often sub-stoichiometric, resulting in the phosphopeptide being present in much lower abundance than the other unphosphorylated peptides. This will result in suppression of the phosphopeptide relative to its unphosphorylated counterpart during the mass spectrometric analysis. Suppression of the phosphopeptide is even more pronounced in the presence of many other unphosphorylated peptides such as that found in the protein digest. Reducing the number of unphosphorylated peptides present in the sample by enriching for the phosphopeptides will therefore enhance mass spectrometric mapping of the phosphorylation site. Finally, performing tandem MS on the phosphopeptide to determine the location of the phosphate group can be difficult due to the instability of the phosphate moiety whereby the phosphate moiety may be eliminated before the peptide can even undergo fragmentation, making it difficult to locate the phosphorylation site. Also, since fragmentation of the phosphopeptide may result in only the elimination of the phosphate moiety and very little fragmentation along the backbone, database searching will not identify the phosphopeptide.

Isolation and enrichment techniques

Isolation and enrichment techniques have been developed to enhance mass spectrometric analysis of phosphopeptides even further, eliminating the need for radiolabeling, antibodies or fluorescent labeling. The most common enrichment methods are immobilized metal-ion affinity chromatography (IMAC),[62-67] and chemical modification methods.[68-73] Once isolated, the resulting phosphopeptides can then be identified by mass spectrometry, however, this can be very difficult as will be discussed later.

Immobilized metal-ion affinity chromatography

Immobilized metal-ion affinity chromatography, originally recognized by Andersson and Porath, is a method used to selectively isolate and enrich phosphopeptides from a peptide mixture via the interaction of the phosphate group on the peptide with the free coordination sites of metal ions immobilized (chelated) to a stationary phase[74] (Figure 1-14). Two of the more frequently used chelating ligands for IMAC are imino-diacetic acid (IDA) and nitrilo-triacetic acid (NTA), however, the majority of the published applications of IMAC for phosphoprotein and phosphopeptide characterization have used IDA. Townsend and coworkers reported that NTA sepharose was superior to IDA-sepharose for phosphopeptide purification by Fe(III)-IMAC because of its higher selectivity for phosphopeptides.[75] However, studies performed by other groups have not confirmed this.[67]

Since the first report utilizing Fe(III) by Andersson and Porath, the use of several other metal ions have been reported, including Al(III), Sc(III), Lu(III), Th(III),[74, 76] and Ga(III).[66] Of these metals, Fe(III) and Ga(III) seem to give best results for phosphopeptide isolation, with Ga(III) showing less overall suppression effect and the

ability to isolate multiply phosphorylated peptides, while Fe(III) shows better selectivity for monophosphorylated peptides.[66]

Although the IMAC methodology is for the isolation and enrichment of phosphopeptides, nonspecific binding could also occur for peptides possessing multiple carboxylic acid groups.[62] These peptides could suppress the signal from trace-level phosphopeptides in the mixture. One way to prevent this problem is to convert the peptides in the mixture to the corresponding peptide methyl esters by replacing the carboxylic acid groups. This is especially needed when acidic amino acids (aspartate, glutamate, and S-carboxymethylated cysteine) are present. By performing this methylation step prior to the IMAC isolation, it is expected that nonspecific binding through carboxyl groups will be prevented, resulting in the selective isolation of phosphopeptides only. The methylation step, however, may be a problem if there is moisture present due to the sensitivity of methanolic-HCl to water. This means that measures have to be taken to keep the reaction mixtures extremely dry, which can make the sample preparation prior to mass spectrometry more labor intensive taking approximately three and a half hours compared to an hour without the methylation step.

Another area of concern for the IMAC method is that the phosphate moiety may be lost during ionization or fragmentation due to the lability of the phosphate group. This means that once the phosphate group is removed, localization of the phosphorylation site will be difficult. Also, little sequence information may be obtained because most of the fragmentation energy is used to remove the phosphate moiety. One way of alleviating this problem would be to replace the phosphate group with a more stable label.

Chemical derivatization

The problem of the lability of the phosphate group has been addressed by several proposed chemical modification methods whereby H_3PO_4 is removed by β-elimination at high pH, a linker containing a thiol group is added by Michael addition, followed by the addition of a biotin-containing compound which acts as an affinity tag for purification of the phosphopeptides as well as a tag for phosphorylation site mapping[67-73, 77] (Figure 1-15). Several linkers have been used for modification, including ethanedithiol,[69, 71, 72, 77] and various lengths of alkanethiols.[73] Of these linkers, ethanedithiol seems to be the most popular due to the presence of a thiol group on both ends of the molecule, thus increasing the chances for it to be bound. Several types of biotinylated chemicals have been used, including iodoacetyl-PEO-biotin,[72, 77] biotin-HPDP,[71] and (+)-biotinyl,3-maleimidopropionamidyl-3,4-dioxoctanediamine.[69] The disadvantages of this technique are that it is labor-intensive, time-consuming (several hours), and the many sample handling steps involved could lead to significant sample loss. These problems can be eliminated if the label is synthesized and stored prior to sample preparation, however, not very many publications report doing this.[77]

Qiagen phosphoprotein purification kit

Within the past 2 years, the development of an enrichment method specifically for phosphorylated proteins from complex cell lysates was introduced by Qiagen Incorporated. Purification of the phosphoproteins is performed by a proprietary affinity chromatography method. The little information that is known about this column is that the phosphate groups on phosphoproteins are bound to the column with high specificity, while proteins without the phosphate groups will not bind to the column and therefore be found in the column flow-through. The bound phosphoproteins are then washed with a

phosphate-buffered saline buffer (PBS) and stored. The free phosphate in the elution buffer serves two functions: 1) displaces the phosphoproteins from the column, and 2) inhibits the phosphate activity in the cell lysate, therefore stabilizing the phosphorylation status of the proteins during downstream processing and storage.

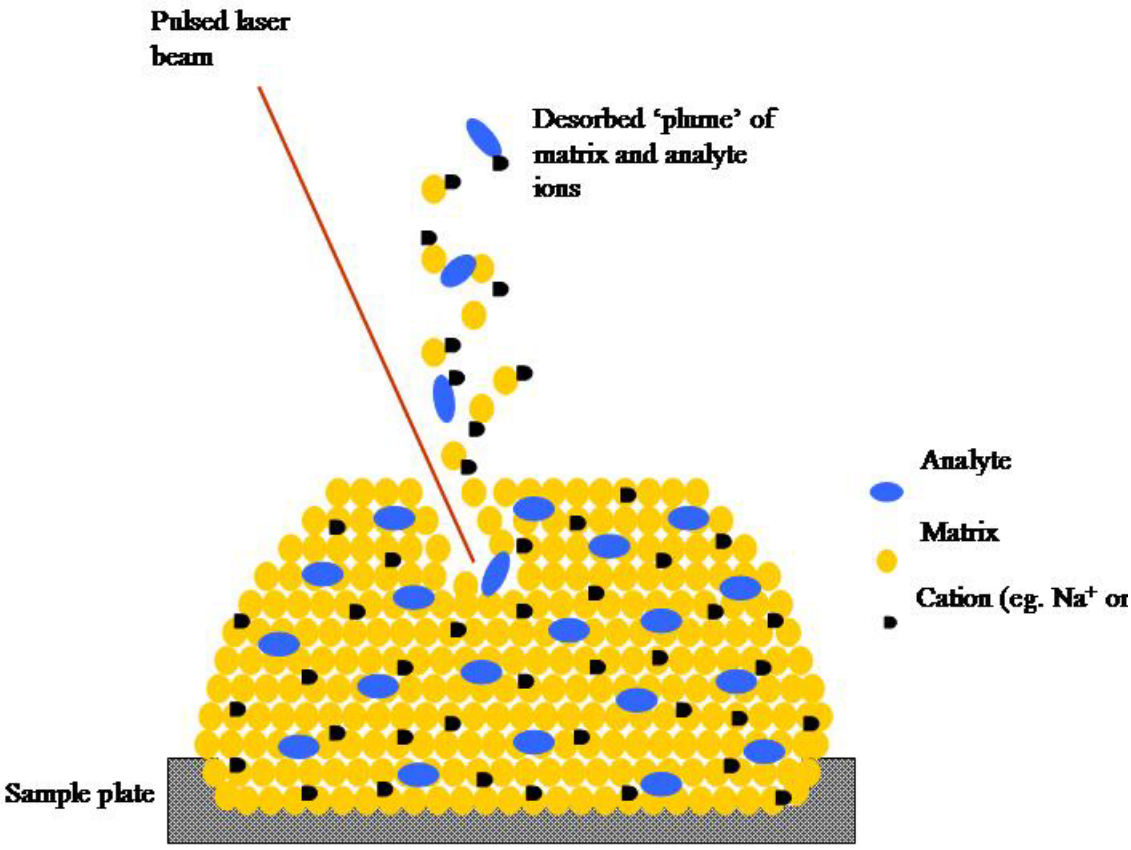

Figure 1-1. The MALDI process

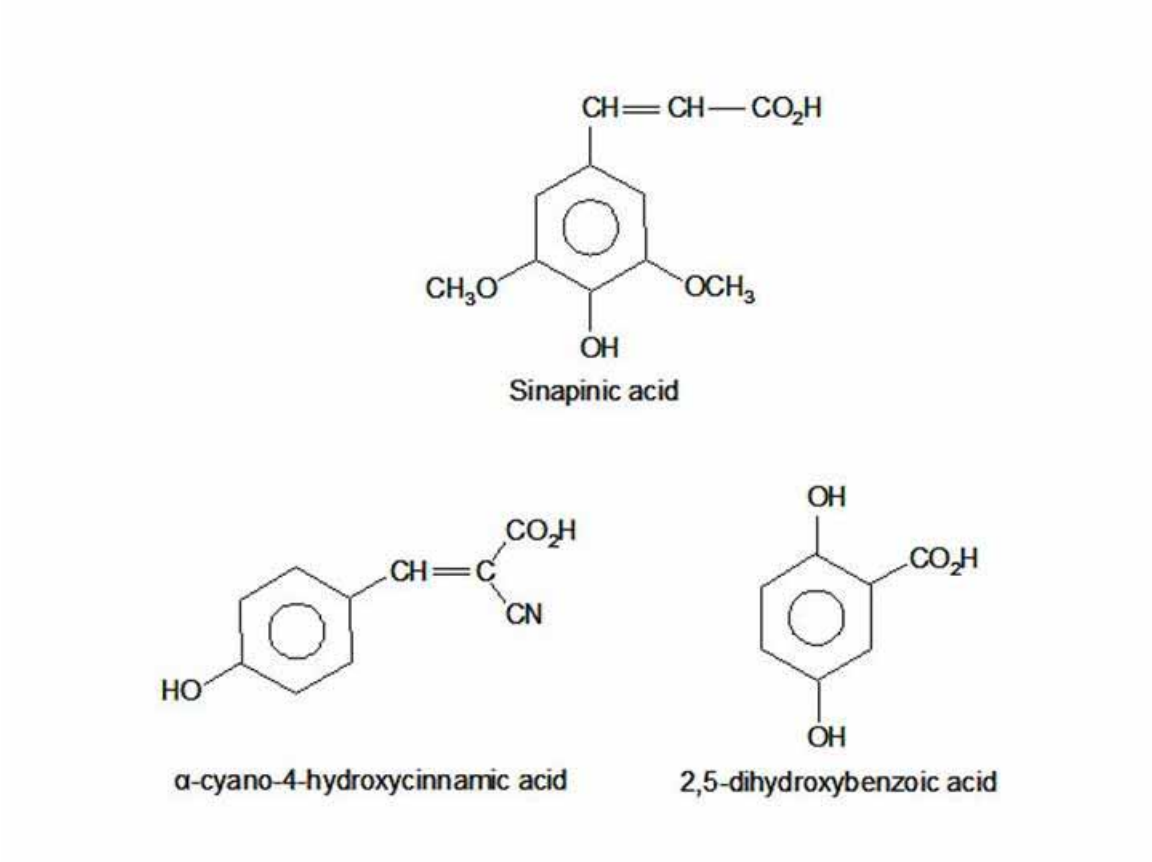

Figure 1-2. Commonly used MALDI matrices for peptide and protein analysis

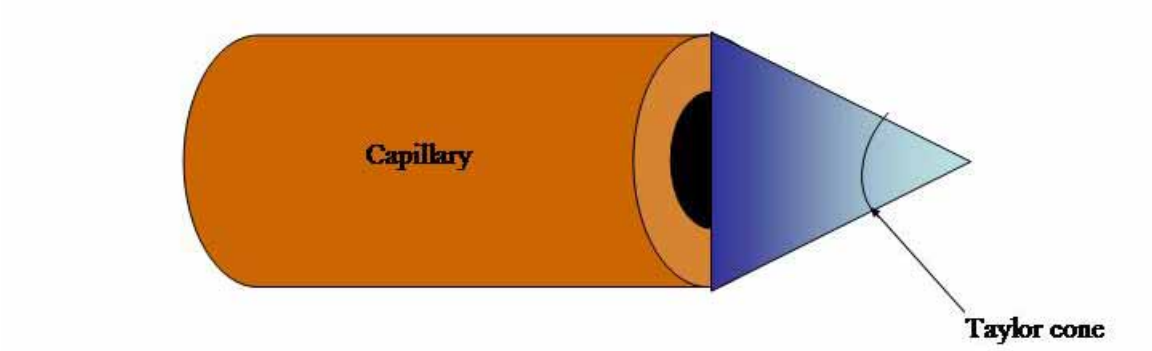

Figure 1-3. Taylor cone formation

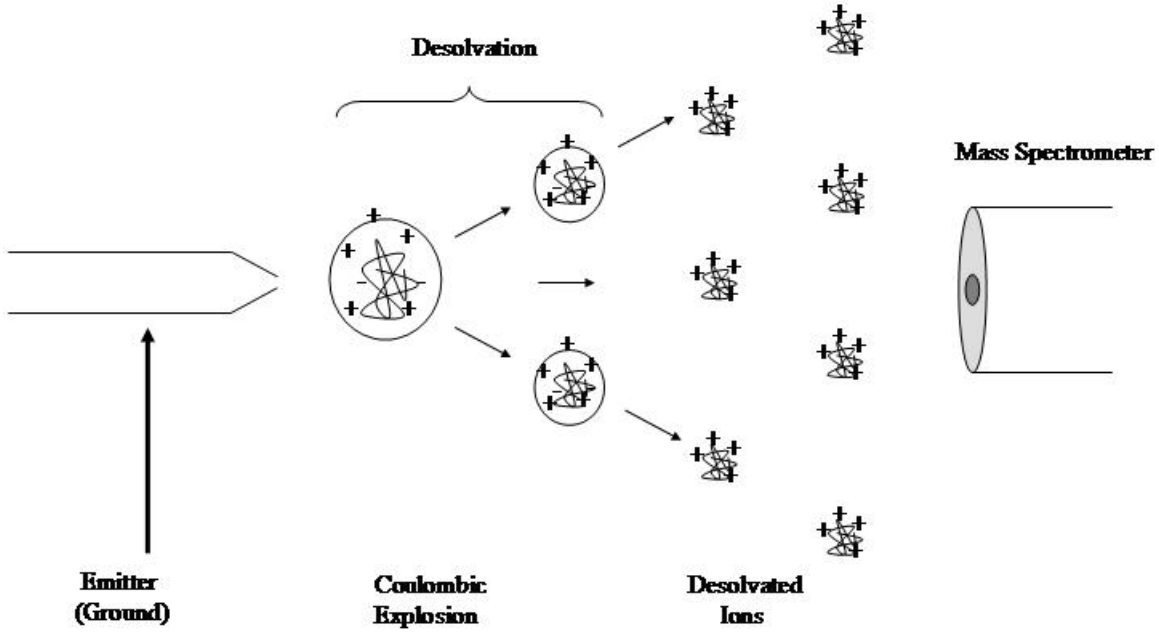

Figure 1-4. The electrospray ionization process

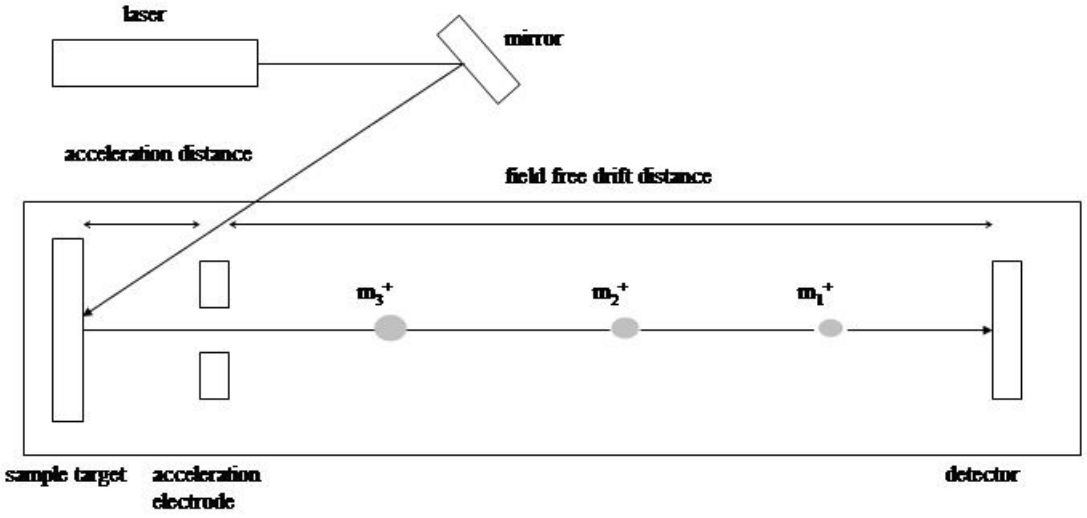

Figure 1-5. A linear time-of-flight mass spectrometer

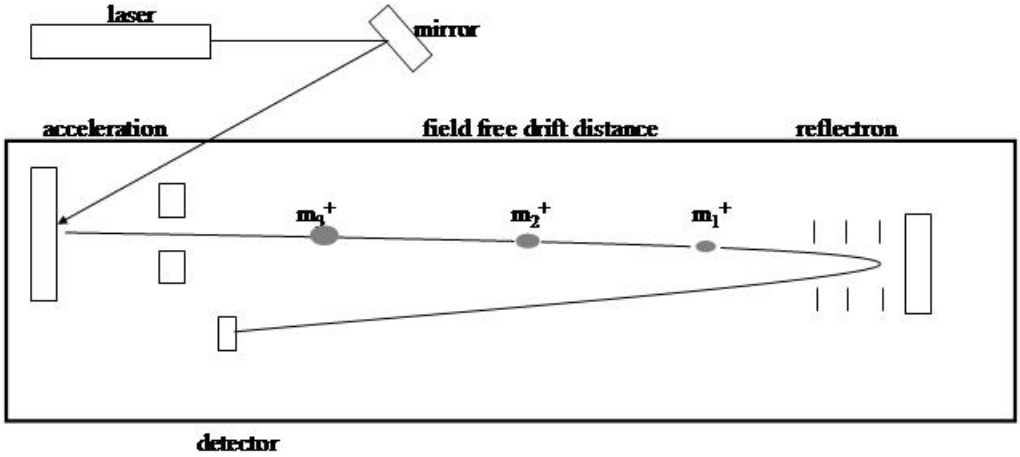

Figure 1-6. A reflectron time-of-flight mass spectrometer

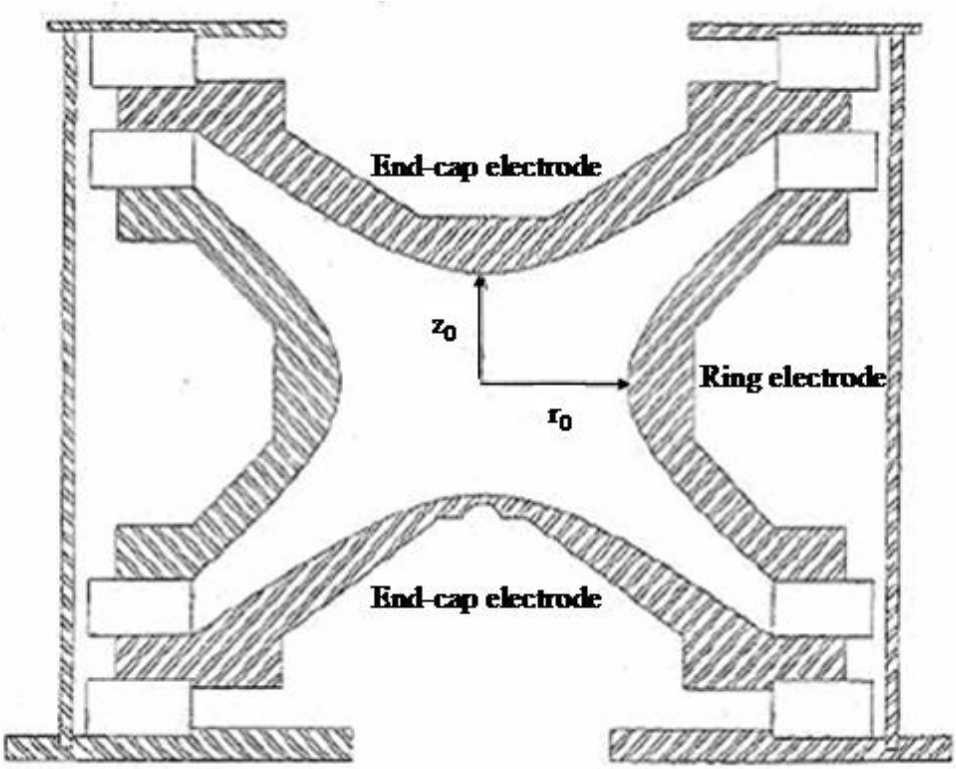

Figure 1-7. Three-dimensional ideal ion trap showing the asymptotes and the dimensions r_0 and z_0. Modified from March, R.E. *International Journal of Mass Spectrometry*. **2000**, *200*, 285-312. Figure 1, page 287.

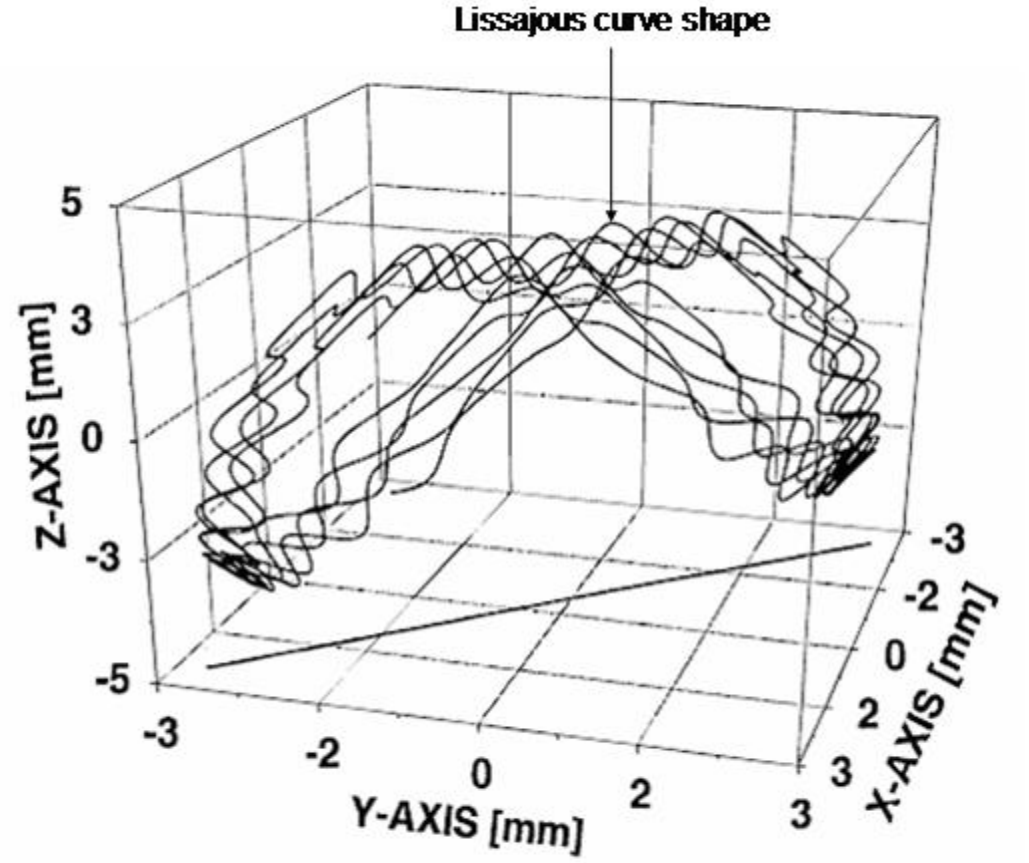

Figure 1-8. Trajectory of a trapped ion. Modified from March, R.E. *Journal of Mass Spectrometry*. **1997**, *32*, 351-369. Figure 8, page 356.

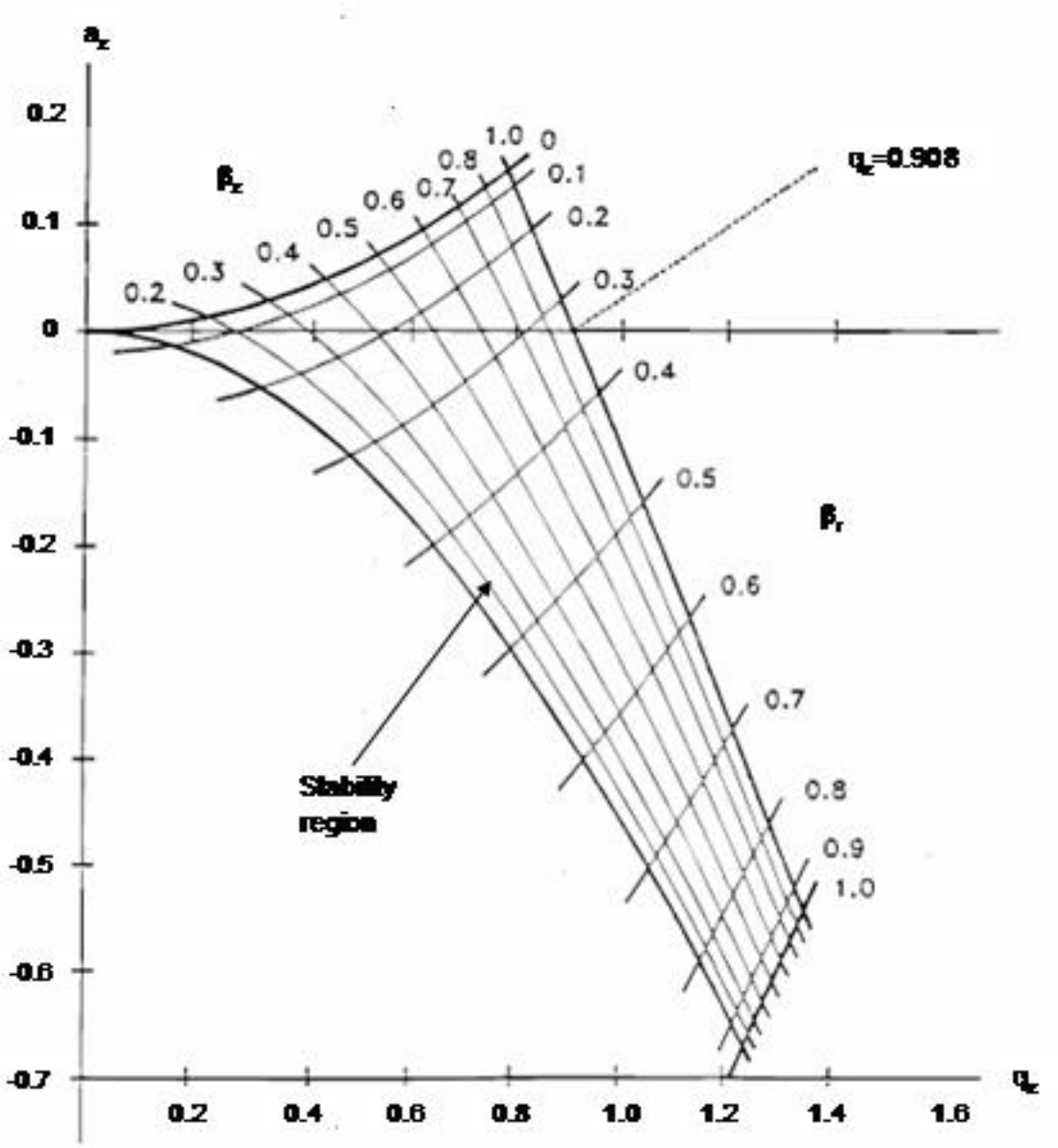

Figure 1-9. Mathieu stability diagram in (a_z, q_z) space for the region of simultaneous stability in both the *r*- and *z*-directions near the origin for the three-dimensional quadrupole ion trap. The q_z-axis intersects the β_z=1 boundary at q_z=0.908, which corresponds to q_{max} in the mass-selective instability mode. Modified from March, R.E. *Journal of Mass Spectrometry*. **1997**, *32*, 351-369. Figure 7, page 356.

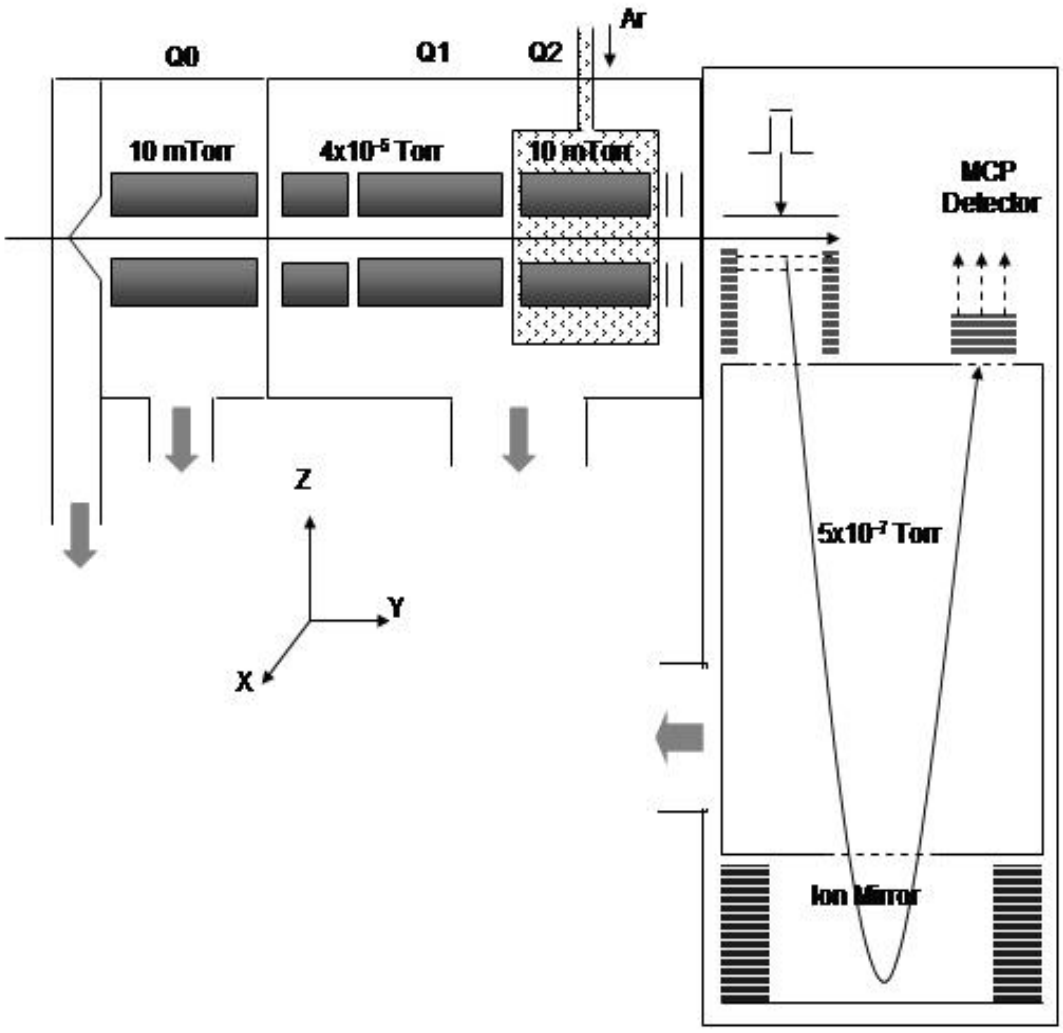

Figure 1-10. Tandem QqTOF mass spectrometer. Adapted from Chernushevich, I.V.; Loboda, A.V.; Thomson, B.A. *Journal of Mass Spectrometry*. **2001**, 36, 849-865. Figure 1, page 860.

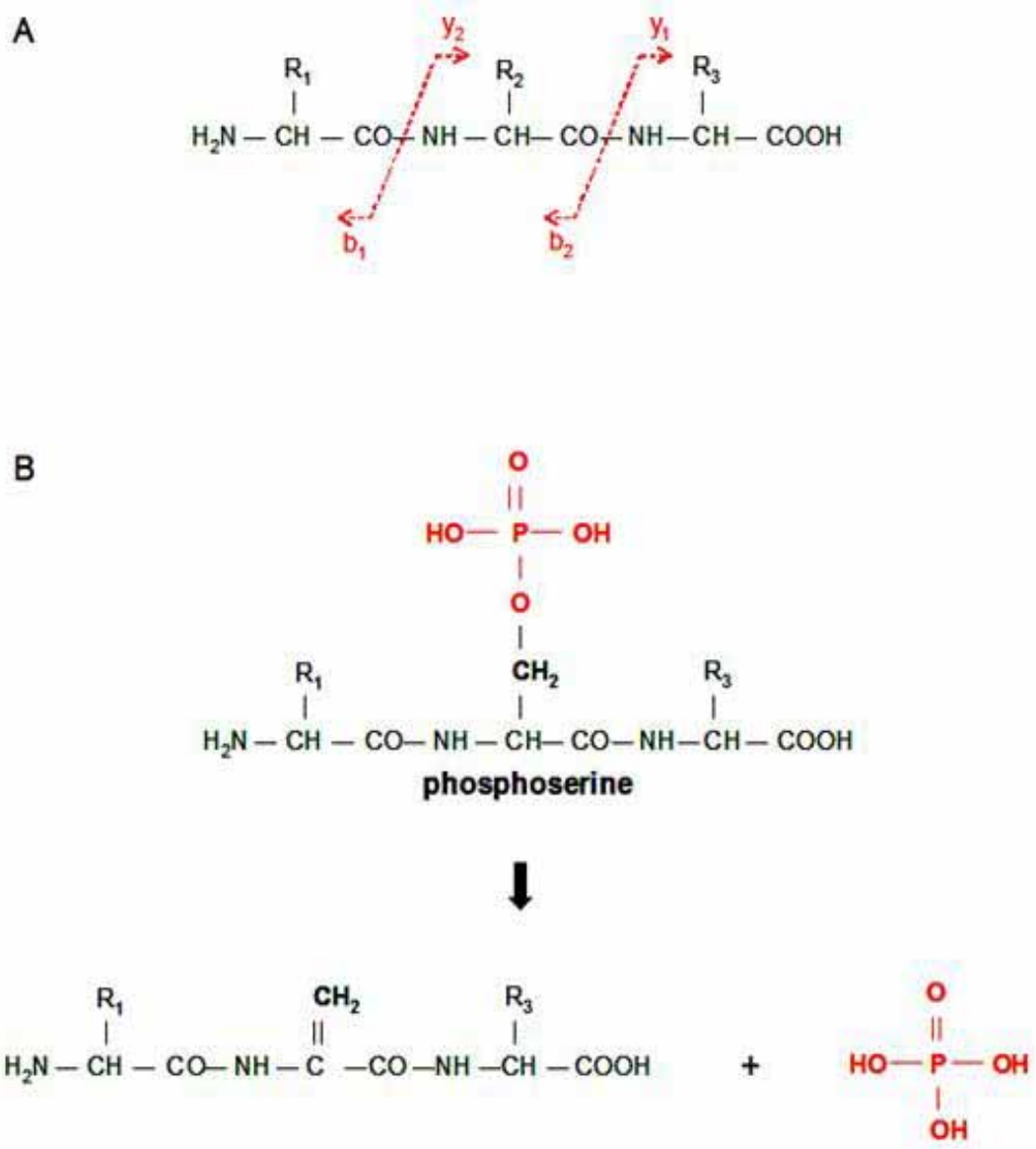

Figure 1-11. Peptide fragmentation. A) Typical low-energy CID fragmentation of a peptide forming mainly *b* and *y*-ions. B) Fragmentation of a phosphopeptide resulting in mainly β-elimination of the phosphate moiety as phosphoric acid.

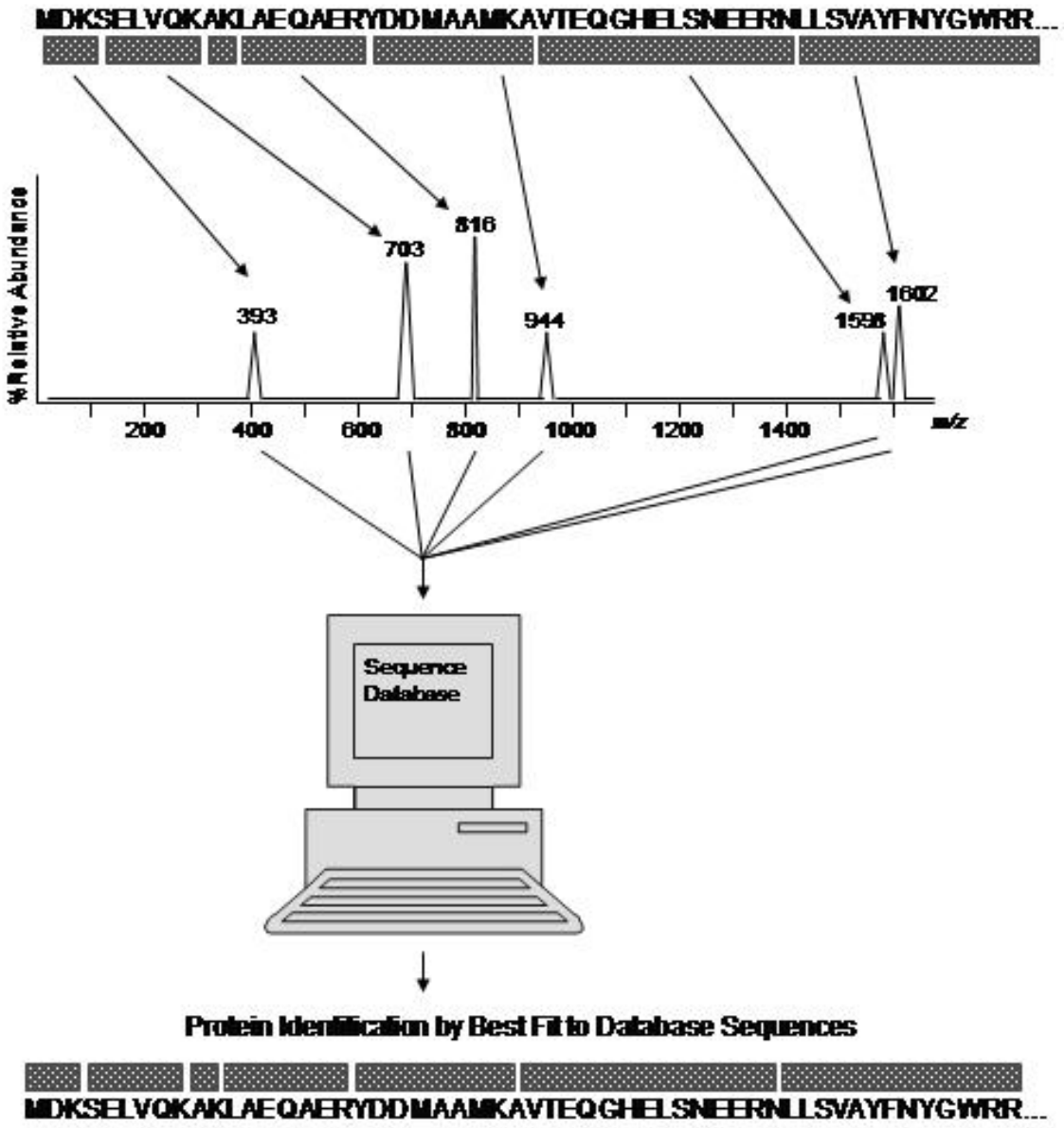

Figure 1-12. Peptide mass mapping. A protein sequence can be verified by site-specific digestion and measurement of the peptide ions for correlation with those predicted by the sequence. Conversely, if the identity of the protein is not known the peptide mass map can be used to search the protein database to find the sequence that best fits the mass map. Adapted from Yates, J.R. *Journal of Mass Spectrometry*. **1998**, 33, 1-19. Figure 3, page 8.

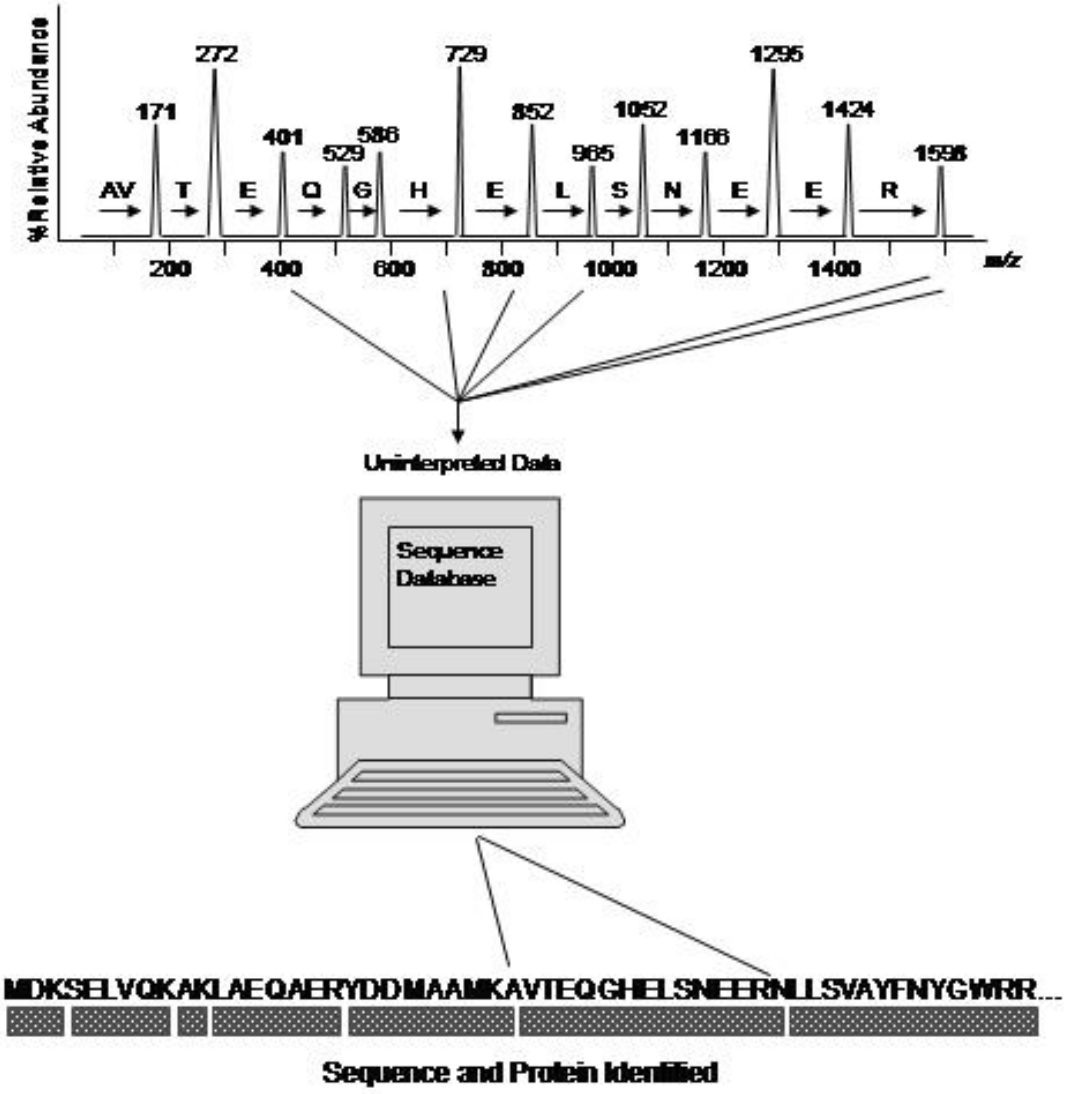

Figure 1-13. Tandem mass spectrometric sequencing. The ladder of fragment ions represents the amino acid sequence of the peptide. By subtracting the *m/z* values for adjacent ions of the same type the sequence can be elucidated. Conversely, the fragmentation pattern can be used to search the protein or nucleotide database to find the amino acid sequence that best fits the tandem mass spectrum. Adapted from Yates, J.R. *Journal of Mass Spectrometry*. **1998**, 33, 1-19. Figure 4, page 10.

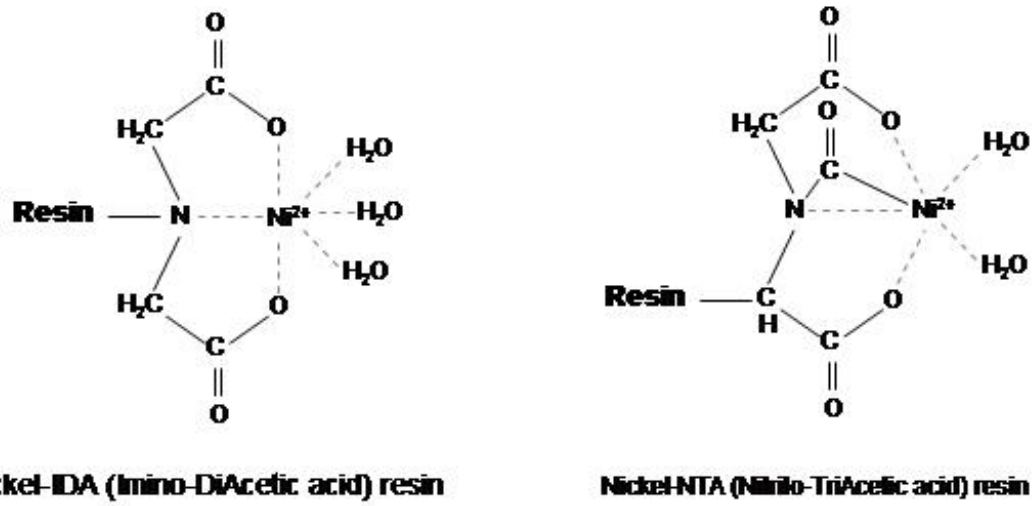

Figure 1-14. Two of the most frequently used chelating ligands for IMAC: imino-diacetic acid and nitrilo-triacetic acid.

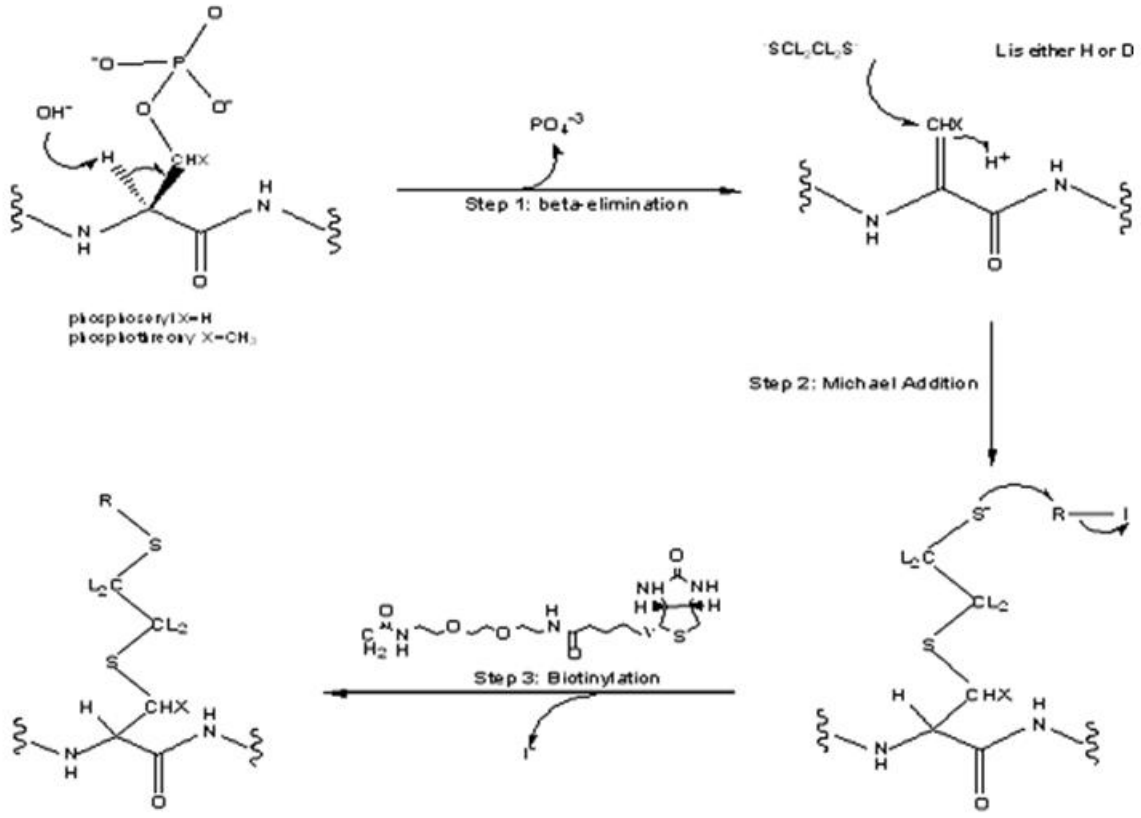

Figure 1-15. Chemical derivatization method for the affinity purification of a phosphorylated peptide. Adapted from Goshe, M.; Conrads, T.; Panisko, E.; Angell, N.; Veenstra, T.; Smith, R. *Analytical Chemistry*. **2001**, 73, 2578-2586. Figure 1, page 2581.

CHAPTER 2
PHOSPHORYLATION DETECTION AND ENRICHMENT

Reversible phosphorylation on serine, threonine and tyrosine residues on a protein is one of the most common and important post-translational modifications involved in virtually all cellular processes. Key to the molecular understanding of these signals, is the identification of kinases, their substrates, and the specific site of phosphorylation. However, the process of studying phosphorylation can be a very difficult and tedious task as mentioned in Chapter 1. A major contributor to the difficulty of this process is suppression by unphosphorylated peptides in a complex mixture, making phosphopeptide isolation and enrichment a necessary step.

For the development of a method for phosphoprotein analysis of a complex mixture (for example *Arabidopsis thaliana* protein extract), one of the first requirements is the determination of a means of detecting phosphoproteins along with the limits of detection of the chosen method. Second, since isolation and enrichment of the phosphorylated protein is necessary if using a MALDI-TOF-MS or QIT-MS, examination and comparisons of available enrichment methods would be advantageous for determining the most appropriate method.

MALDI-TOF-MS can provide a means for rapid sample preparation and analysis for method development, however, optimal conditions for MALDI has to be determined prior to any other method development strategies. As mentioned earlier, selection of matrix and solvent conditions is critical. This is even more important for the analysis of phosphorylated peptides because of their low response to mass spectrometry in positive

ion mode due to the presence of the negatively charged phosphate group. This negative charge interference in detection is even more pronounced when multiple phosphate groups are present in the peptide. Asara and Allison determined that one way to alleviate this problem is to reduce the negative charge interference by the addition of a positively charged species to the matrix spot.[78] Typically, this may be done by adding a solution of ammonium citrate to the already co-crystallized analyte and matrix spot. Hence, optimization of matrix conditions prior to this step is necessary.

Once matrix conditions for phosphopeptide analysis have been optimized, comparison of enrichment techniques may be performed. As mentioned in Chapter 1, two published methods for phosphopeptide isolation and enrichment are IMAC and chemical derivatization methods. During the initial stages of this project, there were two commercially available IMAC columns on the market, the Pierce Phosphopeptide Isolation Kit and Millipore's ZipTipMC, each having a potential advantage over the other. The isolation kit from Pierce is a small minispin column with Ga(III) chelated to an IDA resin, thus reducing the workload for the user since the metal ions are already chelated and the minispin column format makes it possible to enrich several samples at once. On the other hand, Millipore's ZipTipMC comes as a ziptip with the IDA resin only, that is, the user binds his or her metal of choice during the preparation, thus giving the user the opportunity to test various metal ligands for optimization. At the time, several methods had been reported for isolation and enrichment of phosphopeptides by chemical derivatization methods, with Goshe's[77] publication being the most recent.

Presented are the method development steps performed for determining the best isolation and enrichment techniques for the purposes of this project. Beginning the

method development was the optimization of matrix conditions for MALDI, followed by optimization of conditions for both available IMAC products as well as the chemical derivatization method published by Goshe.[77] Once optimized, both enrichment techniques (IMAC and chemical modification) were compared for determination of the best method for the goal of this project. It should be noted that at the later stages of this project, a proprietary Phosphoprotein Purification Kit was developed by Qiagen Inc. which will not be discussed until Chapter 4.

Experimental Methods

Materials and Instruments

The Pro-Q® Diamond Phosphoprotein Gel Stain (Pro-Q) and SYPRO Ruby Gel Stain (SYPRO) were obtained from Molecular Probes, Inc (Eugene, OR). 10% Bis-Tris Novex NuPAGE polyacrylamide gels were from Invitrogen (Carlsbad, CA). Protein standards (β-casein, ovalbumin, bovine serum albumin, and carbonic anhydrase) as well as the evaluated MALDI matrices (α-cyano-4-hydroxycinnamic acid and dihydroxybenzoic acid) were purchased from Sigma Aldrich (St. Louis, MO). Sequencing grade trifluoroacetic acid was obtained from Applied Biosystems (Foster City, CA). The Phosphopeptide Purification Kit (IMAC mini-spin columns), (+)-biotinyl-iodoacetamidyl-3,6-dioxaoctanediamine (iodoacetyl-PEObiotin) and tris(2-carboxyethyl) phosphine hydrochloride (TCEP-HCl) were from Pierce (Rockford, IL). Reversed-phase C18 ZipTips and ZipTipMC were obtained from Millipore (Billerica, MA). Sequencing-grade modified trypsin was purchased from Promega (Madison, WI). Acetonitirile, ethanol, methanol, formic acid, glacial acetic acid, barium hydroxide, sodium hydroxide, ammonium bicarbonate, and Coomassie Brilliant Blue R250 (Coomassie) were obtained

from Fisher Scientific (Fairlawn, NJ). HPLC grade water was purchased from Burdick & Jackson (Pleasant Prairie, WI).

Fluorescent imaging of gels was acquired with a Typhoon Scanner (GE Healthcare, Piscataway, NJ). Mass spectrometric measurements were made using a Voyager-DE Pro Biospectrometry Workstation (Applied Biosystems, Foster City, CA).

Phosphoprotein Detection

1.86 μg and 0.93 μg β-casein, ovalbumin, bovine serum albumin, and carbonic anhydrase were separated by 1-dimensional gel electrophoresis using a 10% Bis-Tris Novex NuPAGE polyacrylamide gel with MOPS running buffer at 200V for 1 hr. The gel was stained with the Pro-Q Diamond and imaged according to the manufacturer's protocol. The gel was then counter-stained with SYPRO Ruby, imaged with the Typhoon, and counter-stained again with Coomassie. A serial dilution (250, 100, 50, 20, 10 and 1 ng) of a mixture of equal amounts of β-casein, ovalbumin, and bovine serum albumin was then separated on a gel and stained as described above for detection limit determination.

Matrix Optimization

Saturated solutions of the MALDI matrices were prepared in the selected solvent systems in Table 2-1 and mixed with a synthetic peptide (PLARTLpSVAGLPGK). Unless otherwise noted, all MALDI analyses were performed using the dried-droplet method by mixing a 1:1 ratio of analyte to matrix and allowing the spot to dry on the MALDI plate. Two spots of each matrix-analyte solution were spotted and once dried, 25 mM ammonium citrate was spotted on top of one of the two spots and allowed to dry.

Table 2-1. Matrix solvent systems for evaluation.

Matrix	**Solvent combinations**	**Additive after spotting**
α-cyano-4-hydroxycinnamic acid	50% ACN, 0.3% TFA	
	50% ACN, 0.3% TFA	25 mM ammonium citrate
	50% ACN, 3% FA	
	50% ACN, 3% FA	25 mM ammonium citrate
2,5-dihydroxybenzoic acid	50% ACN, 0.3% TFA	
	50% ACN, 0.3% TFA	25 mM ammonium citrate
	50% ACN, 3% FA	
	50% ACN, 3% FA	25 mM ammonium citrate

ACN – acetonitrile, TFA – trifluoroacetic acid, FA – formic acid

Mass spectra were obtained with a MALDI-TOF MS instrument equipped with a 337 nm nitrogen laser and reflectron optics. All spectra were acquired in positive ionization mode under delayed extraction conditions in reflectron mode. Spectra were obtained with an acceleration voltage of 20 kV and 100 laser shots at a laser repetition rate of 3.0 Hz. Laser intensity, extraction delay time, grid voltage and guide wire were all adjusted to obtain the best spectrum for each sample. An external calibration was performed before each spectrum was obtained with a calibration mixture consisting of 1.0 pmol Des-Arg^1-Bradykinin, 1.3 pmol Angiotensin I, 1.3 pmol Glu^1-fibrinopeptide B, 2.0 pmol ACTH (1-17 clip), and 1.5 pmol ACTH (18-39 clip).

Phosphopeptide Enrichment

A stock solution of β-casein digest was prepared by incubating 1 mg of protein with 10 μg trypsin in 100 mM ammonium bicarbonate at 37 ^{0}C overnight.

Immobilized metal-ion affinity chromatography

Both the Phosphopeptide Purification Kit and ZipTipMC were used according to the manufacturer's protocol. Briefly, this involves binding the sample at low pH (0.1% acetic acid), followed by wash steps with low organic (0.1% acetic acid with 10% acetonitrile, and water) for the removal of unphosphorylated peptides, and then elution of

the phosphopeptides with higher pH (0.3N ammonium hydroxide solution). For the ZipTipMC, the metal ion of choice is bound to the resin. Fe^{3+}, Ga^{3+}, Cu^{2+}, and Ni^{3+} were used as the metal ions of choice. Isolated peptides were analyzed by MALDI analysis as described above.

Chemical derivatization

Chemical derivatization of the synthetic phosphopeptide was performed according to the protocol published by Goshe.[77] Reaction conditions such as time and temperature were investigated for optimal yield.

Results and Discussion

Phosphoprotein Detection

Analysis of the four protein standards (including the phosphorylated proteins, β-casein and ovalbumin) with the Pro-Q stain showed selective staining of the phosphorylated proteins (Figure 2-1). Counter-staining with both SYPRO and Coomassie proved that all four proteins were indeed on the gel and the Pro-Q stain was selective for only the phosphorylated proteins. Counter-staining also indicated that the Pro-Q stain could be as sensitive as the SYPRO stain (as low as 1 ng)[53] compared to the Coomassie stain (as low as 100 ng),[53] according to the intensity of the gel bands. Subsequent detection limit experiments determined that the Pro-Q stain could detect phosphoproteins at levels as low as 20 ng (Figure 2-2). However, it was noticed that large amounts (several μg) of unphosphorylated proteins were detected by the stain but appeared more as smears rather than very distinct bands. This problem could be alleviated by always including an unphosphorylated protein as a standard among the samples and subtracting the corresponding signal for removal of background staining. However, if there is a

greater amount of an unphosphorylated protein present in the sample of interest than the standard, background subtraction may not completely correct for this problem.

Matrix Optimization

Initial studies utilized several matrix compounds, with α-cyano giving the best results for peptide analysis. This matrix (α-cyano matrix containing 50% ACN/0.3%TFA) was then optimized for phosphopeptide enhancement by the addition of ammonium citrate. Figure 2-3 compares spectra of a phosphopeptide of 1459 Da obtained with and without the addition of ammonium citrate. The increase in intensity of the $[M+H]^+$ upon addition of ammonium citrate shows that ammonium citrate does enhance the phosphopeptide signal and can be used with the matrix for optimal conditions for phosphopeptides.

Phosphopeptide Enrichment

Figure 2-4 shows a MALDI spectrum of a β-casein digest. Expected unphosphorylated peptides can be seen, however, the phosphopeptides at 2062 (1 phosphorylation site) and 3121 (4 phosphorylation sites) m/z are not visible, proving that isolation and enrichment of phosphopeptides is necessary. Isolation and enrichment of the β-casein phosphopeptides was then performed with the Millipore ZipTipMC. A variety of metal ions were tested to identify which would give the best isolation and enrichment of the phosphopeptides, as well as the least nonspecific binding. Figure 2-5 shows spectra of the isolated phosphopeptides obtained using Fe^{3+}, Ga^{3+}, Cu^{2+}, and Ni^{3+} proving that the ZipTipMC does isolate and enrich for phosphopeptides. Of the four metal ions used, Ga^{3+} and Fe^{3+} seem to give the best spectra with respect to good signal being obtained for both the monophosphorylated and multiply phosphorylated peptides. The monophosphorylated peptide could not be found in the spectrum obtained using Ni^{3+}, however, the signal

intensity of the multiply phosphorylated peptide was higher than those obtained for the other spectra.

The Pierce Phosphopeptide Isolation Kit was then used to isolate and enrich the β-casein phosphopeptides, giving similar results as the ZipTipMC with Ga^{3+} (results not shown). These results were expected since the Pierce Kit also uses Ga^{3+}. The sensitivity of the column was then tested by isolating and enriching for phosphopeptides from 1 picomole of β-casein digest (Figure 2-6). At first glance, it seemed as if only the monophosphorylated peptide was isolated (Figure 2-6A), however, after the addition of ammonium citrate to the spot, significant enhancement of the multiply phosphorylated peptide (3121 m/z) was seen, however, the monophosphorylated peptide could not be seen possibly due to ionization suppression by the multiply phosphorylated peptide. This shows that this IMAC column does isolate and enrich phosphopeptides at very low amounts (1 pmol) and also that ammonium citrate can be added to the MALDI spot for the enhancement of the multiply phosphorylated peptides.

Both IMAC methods were compared to decide which would be best for this project. In regards to levels of sensitivity for phosphopeptides, both gave similar results (femtomole range), leaving the decision up to ease of use and cost. Of the two IMAC products mentioned, the Pierce product was the least labor intensive, however, it was more costly. Due to the amount of samples that would possibly have to be enriched, the Pierce Phosphopeptide Isolation Kit was chosen as the better method of the two.

For a comparison of the chemical derivatization and IMAC methods, the protocol published by Goshe[77] was tested for its utility as an enrichment technique. A synthesized phosphopeptide PLARTLpSVAGLPGK was used for optimization of each reaction step

(β-elimination, Michael addition, and biotinylation). For β-elimination, optimal reaction conditions of temperature and time were examined. Once established, parameters for the addition of the linker were examined, that is, type of linker used (ethaneditiol (EDT) or 2-mercaptoethylamine), addition of a reducing agent (TCEP-HCl) to prevent dimerization of the linker molecule, length of time of the reaction, and also whether the reaction could occur simultaneously with β-elimination. At this point, removal of the EDT prior to the addition of biotin was critical for the prevention of competition for the biotin between the EDT or 2-mercaptoethylamine-linked peptides and the free EDT or 2-mercaptoethylamine molecules. If complete removal of the linkers is not achieved, the reaction yield will be minimal, which is not acceptable in real samples where phosphorylation can be at very low levels. Several methods such as extraction with diethyl ether or size-exclusion chromatography were utilized for the efficient removal of the linker molecules. Following the removal step, the addition of biotin was performed and optimized.

Figure 2-7 shows the chemical derivatization of a synthetic phosphopeptide. A spectrum of the phosphopeptide was first obtained (Figure 2-7A), followed by β-elimination at determined optimal conditions (55 ^{0}C for 1 hour in 0.5 M sodium hydroxide (Figure 2-7B). Figure 2-7C shows the addition of the EDT linker at the determined optimal conditions, that is, the β-elimination and EDT addition steps are performed concurrently (55 ^{0}C for 1 hour in 0.5 M sodium hydroxide, 3.6 mM TCEP-HCl and a 5 molar excess of EDT), followed by removal of unreacted EDT with a Sephadex G-10 spin column. It should be mentioned that the addition of 2-mercaptoethylamine as a linker was also attempted but was unsuccessful. Biotinylation

was then performed for 90 minutes in the dark with constant stirring and the reaction mixture desalted with a Sephadex G-10 spin column (Figure 2-7D). This experiment proves that the chemical derivatization steps are possible, however, the isolation of the biotinylated peptide with an avidin column was not done due to the low product yield which would not be acceptable for the purposes of this project because of low amounts of starting material that would be available.

This procedure proved to be very labor-intensive and time-consuming so an alternative method was attempted. If a label could be synthesized and stored, sample handling and derivatization time would be reduced. With this in mind, the above label (EDT and biotin) was first synthesized, purified and then added to the phosphopeptide, unfortunately, no reaction seemed to have occurred when monitored by MALDI-TOF-MS.

The synthesis of a novel phosphorylation label was also attempted by the reaction of H-Cys(Trt)-NH_2 and biotin-LC-OSu. The expected product (702 m/z) was not found in the MALDI spectrum of the reaction mixture, however, the sodiated and potassiated forms could be seen at 724 and 740 m/z. The reaction mixture was then purified on an HPLC system and the fraction of interest collected and a MALDI spectrum obtained as shown in Figure 2-8A. In order to validate that this was indeed the product of interest, several other matrices were investigated to see if the expected product could be seen without the adducts. Figure 2-8B shows a spectrum of the expected product with 3-hydroxypicolinic acid matrix, proving that the expected product was formed. Removal of the tert-butyl protecting group from the product was then performed and the expected loss of 57 m/z is seen in Figure 2-8C, resulting in the deprotected label at 645 m/z, as

well as the product with a sodium adduct as well as a potassium adduct. Even though the experiments were successful in synthesizing a possible phosphorylation label, the reaction yield was very low and reaction with the phosphopeptide did not seem to occur after several attempts.

Of the above methods attempted for isolation and enrichment of phosphopeptides, the IMAC columns seem to be more favorable when time and labor are considered. Samples can be prepared for analysis by IMAC in less than an hour, compared to 1 day by the chemical derivatization method, that is, providing that the isolation with the avidin column works effectively. Also, the chemical derivatization method requires a larger quantity of sample and there is a much higher potential of sample loss due to the many steps involved. With all of this in mind, the IMAC columns seemed to be the best method for isolation and enrichment of phosphopeptides.

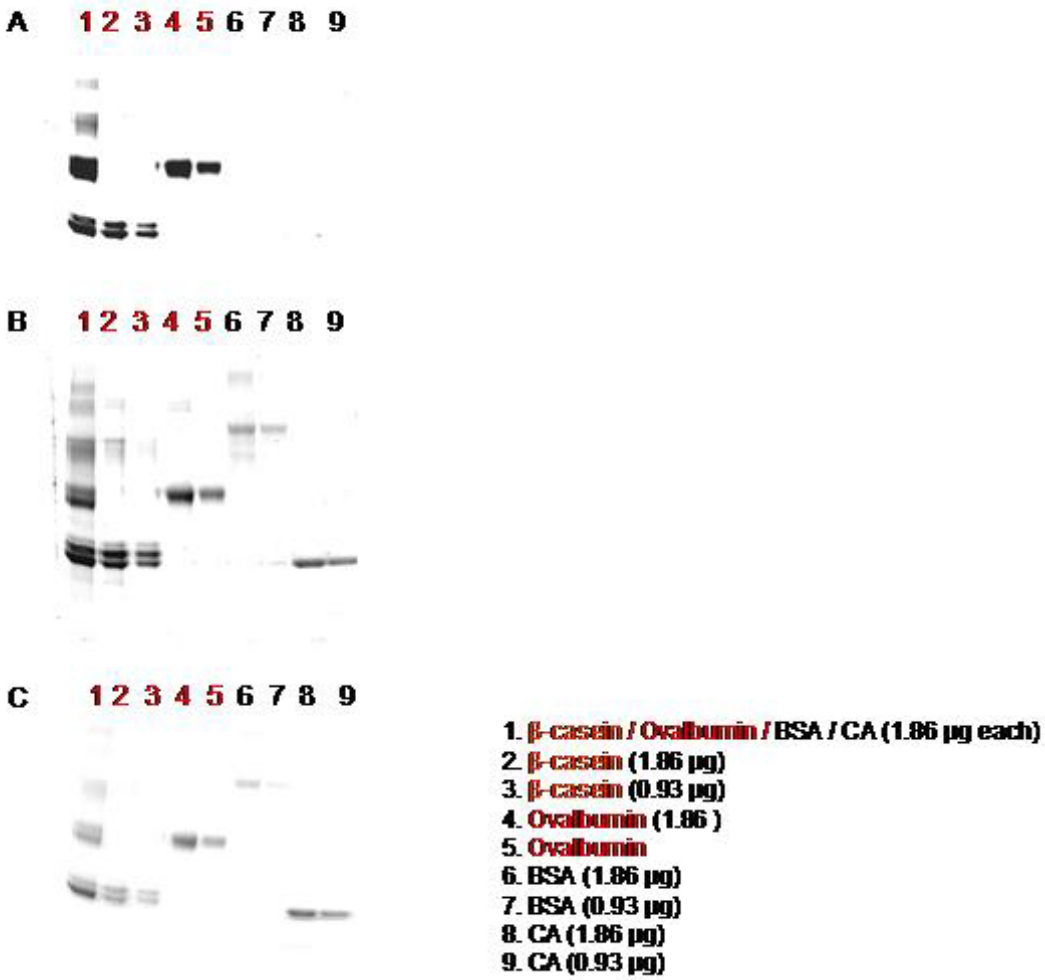

Figure 2-1. One-Dimensional gel electrophoresis of four protein standards. A) Selective phosphoprotein detection with the Pro-Q Diamond Phosphoprotein Gel Stain. B) Total protein staining with SYPRO Ruby protein stain. C) Total protein staining with Coomassie R-250 protein stain.

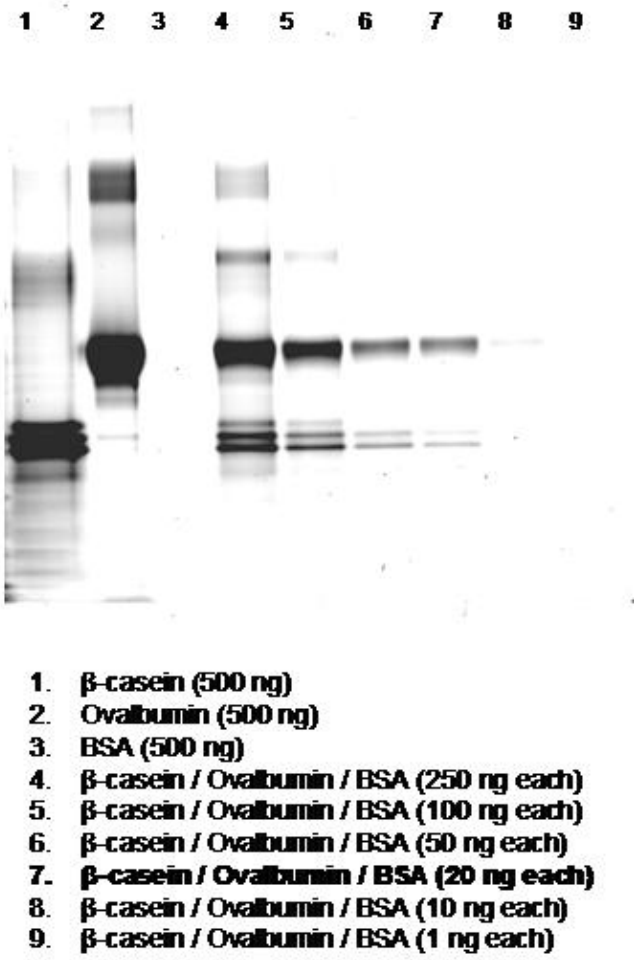

Figure 2-2. Pro-Q stained gel of varying concentrations of protein standards. The lowest detection level of the two phosphoprotein standards appears to be 20 ng.

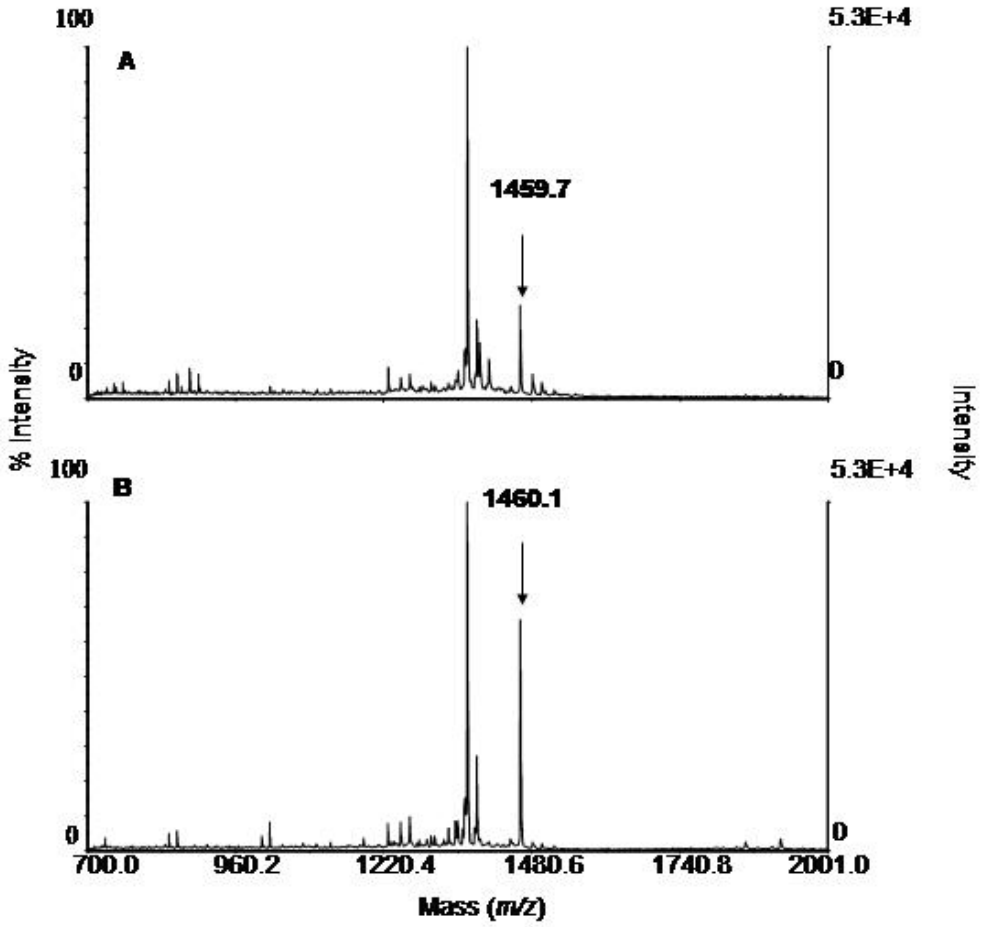

Figure 2-3. Comparison of matrix conditions for phosphopeptide spectra. A) Spectrum obtained with α-cyano matrix. B) Spectrum obtained with α-cyano matrix and the addition of ammonium citrate. Enhancement of the phosphopeptide at m/z 1460 is observed.

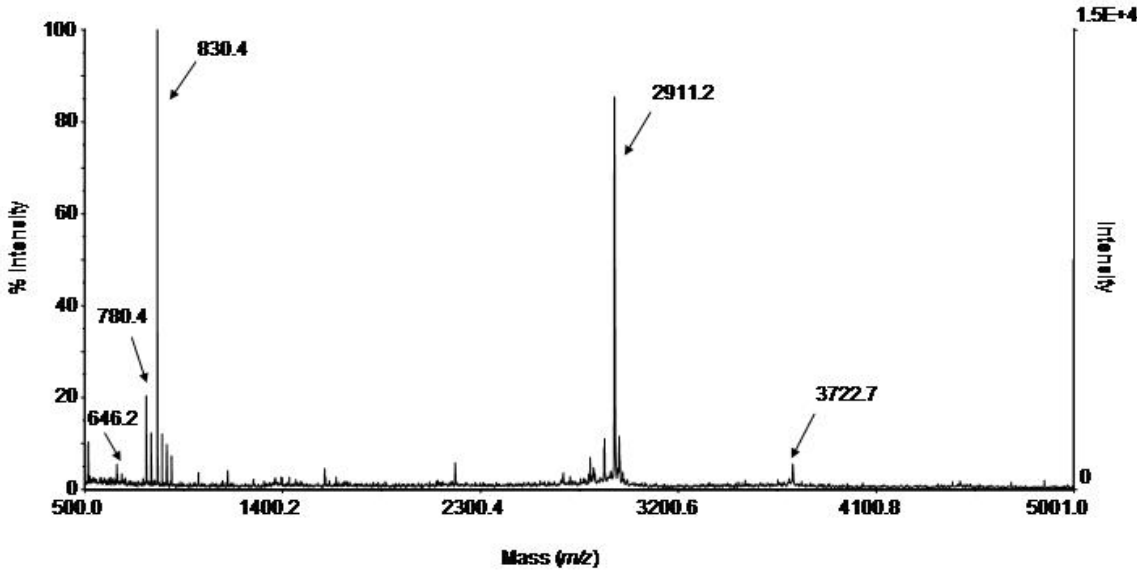

Figure 2-4. β-casein digest spectrum. Unphosphorylated peptides are observed, however, the two known phosphopeptides at *m/z* 2062 and 3123 are not evident.

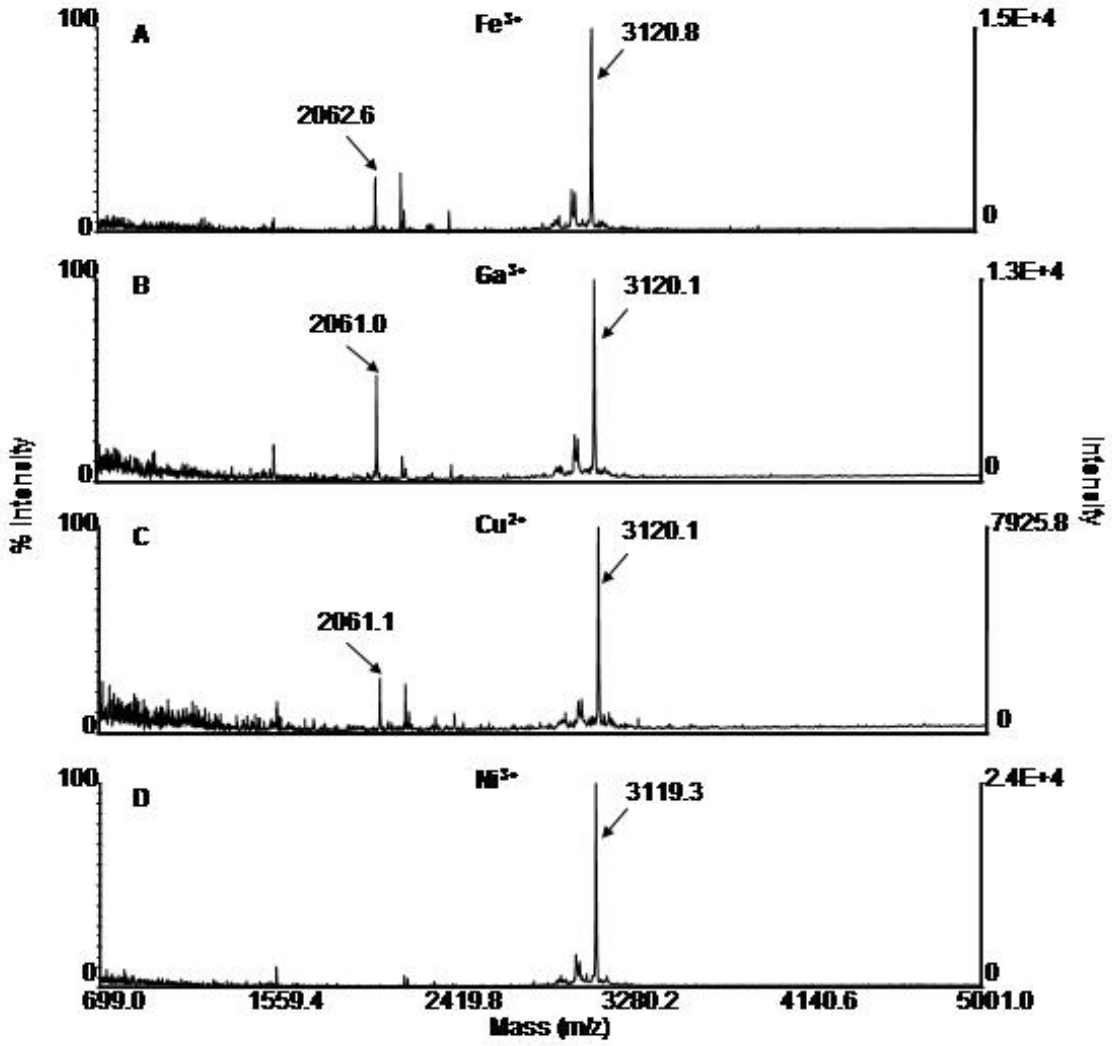

Figure 2-5. Comparison of different metal ions on the ZipTipMC for isolation of β-casein phosphopeptides. A) Fe^{3+}. B) Ga^{3+}. C) Cu^{2+}. D) Ni^{3+}.

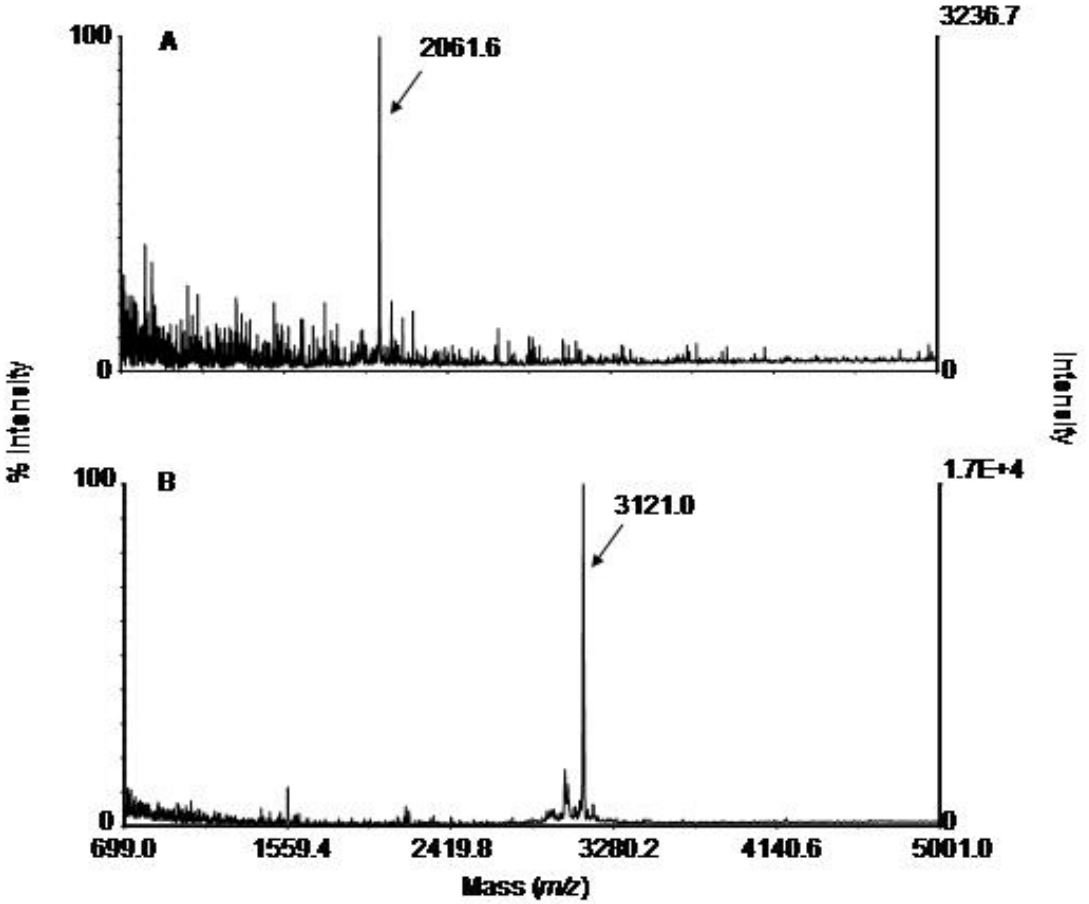

Figure 2-6. Spectra of β-casein phosphopeptides isolated by the Pierce Phosphopeptide Isolation Kit. A) Isolated phosphopeptides spotted with α-cyano matrix. B) Isolated phosphopeptides spotted with α-cyano matrix and the addition of ammonium citrate.

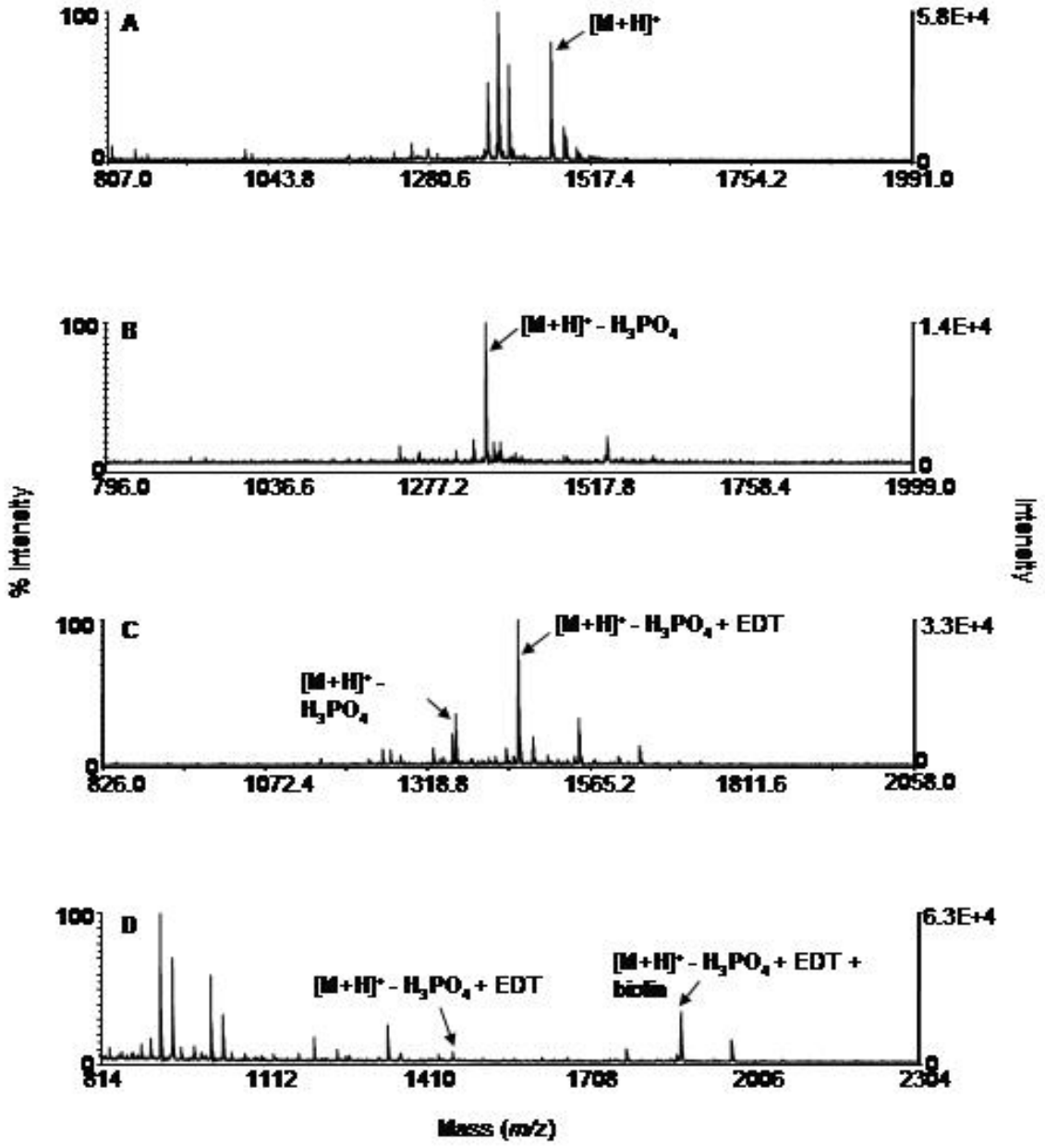

Figure 2-7. Chemical derivatization of phosphopeptide. A) Underivatized phosphorylated peptide. B) β-elimination of peptide. C) Michael addition of EDT to the β-eliminated peptide. D) Biotinylation of peptide.

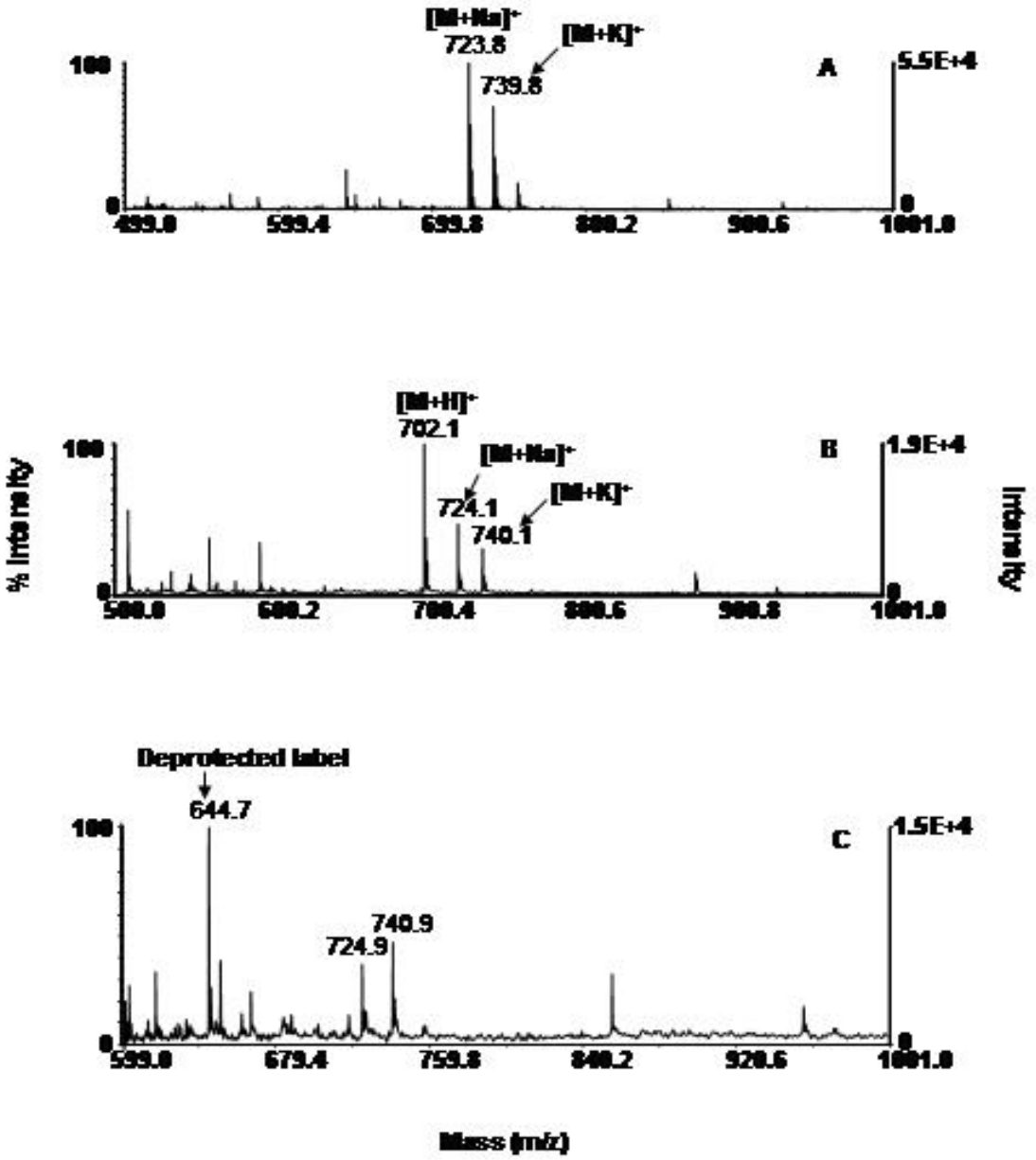

Figure 2-8. Spectrum of chemical derivatization label. A) Spectrum of HPLC purified label using α-cyano. B) Spectrum of HPLC purified label using 3-hydroxypicolinic acid matrix. C) Spectrum of the deprotected label using α-cyano.

CHAPTER 3
INVESTIGATING THE AUTOPHOSPHORYLATION SITES OF A CALCIUM-DEPENDENT PROTEIN KINASE

Since the regulatory properties of autophosphorylation of CDPKs still remain unclear, one of the goals of this project was to develop methods for the identification of phosphorylation sites of phosphorylated proteins and subsequently apply these methods for autophosphorylation site identification of CDPKs. With the availability of three different types of mass spectrometers each possessing a unique feature, the utilization of all three instruments was considered beneficial for complete coverage of autophosphorylation site mapping.

As mentioned in Chapter 1, both the QIT-MS and QqTOF-MS come equipped with similar features for phosphorylation detection, that is, the reconstructed neutral fragment chromatogram and precursor ion scanning, respectively, both taking advantage of the lability of the phosphate group. Also, in-line reversed phase chromatography may be used with these mass spectrometers for separation of peptides prior to MS and MS/MS analysis. Additionally, phosphorylation site mapping is possible with the MS/MS feature.

The MALDI-TOF-MS on the other hand, does not have the capability of MS/MS analysis; however, it has the advantage of rapid sample preparation and analysis as well as affordability for smaller laboratories interested in proteomics. With this in mind, the development of a novel, simple, cost effective method for preliminary phosphorylation identification was proposed. The idea behind this was to take advantage of several difficulties associated with phosphopeptide analysis, including hydrophilicity, ionization

efficiency and suppression effects by unphosphorylated peptides. Often, protein digests are desalted with a ziptip which can result in the loss of phosphopeptides due to their hydrophilicity. Phosphopeptides may not be seen in a MALDI-TOF spectrum due to either loss of the phosphopeptide in the desalting step or to suppression effects. However, if these peptides were first dephosphorylated by β-elimination prior to desalting and spotting, the peptides should be visible in the spectrum. The appearance of these peptides in the spectrum when compared to the spectrum of the untreated digest would then be an indication of a previously phosphorylated peptide that has been dephosphorylated. Targeted MS/MS analysis can then be performed on these peptides for phosphorylation verification.

Presented is the development of a novel, cost effective method for preliminary phosphorylation site identification on a MALDI-TOF-MS. Also presented is the comparison of various available methods for autophosphorylation site mapping of a calcium-dependent protein kinase.

Experimental Methods

Materials and Instruments

NuPAGE 10% Bis-Tris SDS-PAGE gels were obtained from Invitrogen (Carlsbad, CA). Sequencing-grade modified trypsin was purchased from Promega (Madison, WI). The Phosphopeptide Purification Kit (IMAC mini-spin columns) was from Pierce (Rockford, IL). Reversed-phase C18 ZipTips were obtained from Millipore (Billerica, MA). The MALDI calibration mixture Sequazyme Peptide Mass Standards Kit was purchased from PE Biosystems (Foster City, CA). Mass spectrometric measurements were made using either an LCQ Deca ion trap (ThermoFinnigan, San Jose, CA) equipped with a PicoView electrospray ionization source (New Objective, Ringoes, NJ) and an

ABI 140D Solvent Delivery System (Perkin Elmer, Wellesley, MA) or a Voyager-DE Pro Biospectrometry Workstation (Applied Biosystems, Foster City, CA) or a QSTAR (Applied Biosystems, Foster City, CA) equipped with an LC Packings Ultimate nanoHPLC system (LC Packings, Sunnyvale, CA).

Kinase Assay and Protein Preparation

Recombinant calcium-dependent protein kinases 4 and 5 (CPK4 and CPK5) from *Arabidopsis thaliana* were a gift from Estelle Hrabak, University of New Hampshire. Both kinases were autophosphorylated by incubation for one hour at room temperature in 50 mM Tris, pH 7.5, 10 mM $MgCl_2$, 1 mM EGTA, and 1.2 mM $CaCl_2$, and either 1 mM [γ-P^{32}] ATP or 1 mM unlabeled ATP. P^{32}-labelled kinases as well as the untreated kinases were resolved on an SDS gel for autoradiography. In preparation for mass spectral analysis GST-CPK5 autophosphorylated with cold ATP was resolved on an SDS gel along with several standards (α-casein, β-casein, BSA and ovalbumin). Protein bands visualized by staining with Coomassie Brilliant Blue R250 were excised from the gel for in-gel tryptic digestion [54]. An in-solution tryptic digestion of β-casein was also performed, however, reduction and alkylation of the protein was omitted. Samples were then dried in a centrifugal vacuum system (SpeedVac) to near dryness.

Phosphopeptide Enrichment

Phosphopeptides from the autophosphorylated CPK5 digest were isolated and enriched with the Pierce Phosphopeptide Purification Kit (gallium IMAC mini-spin columns) according to protocol. The eluted phosphopeptides were dried and reconstituted in 0.1% TFA. The IMAC-enriched sample as well as a sample of the digest (unenriched sample) were desalted with a reversed-phase C18 ZipTip according to protocol and the

eluted peptides were dried and reconstituted in 5% ACN/0.5% acetic acid for mass spectrometric analysis on the LCQ Deca and QSTAR.

β-Elimination

β-elimination of the in-solution and in-gel tryptic digests of CPK5 was performed according to Knight's procedure[68] in which the peptides were dissolved in a 4:3:1 solution of H_2O/DMSO/ethanol (50 μL) followed by 23 μL saturated $Ba(OH)_2$ and 1 μL 5 M NaOH. Samples were incubated at 37 ^{0}C for 2 hours then neutralized with HCl and dried with a SpeedVac. The samples were reconstituted in 0.1% TFA and desalted with a reversed-phase C18 ZipTip for analysis with the MALDI-TOF mass spectrometer. It should be noted that the in-solution digest of β-casein was first used to test this method followed by the in-gel digests of all the standards (β-casein, α-casein, ovalbumin, and BSA).

Data-dependent LC/MS/MS on the QIT-MS

Samples were introduced to the ion trap mass spectrometer via an on-line reversed-phase capillary HPLC (50 μm i.d. x 5 cm C18 produced in-house) with an isocratic solvent delivery at 200 nL/min with 0% Solvent A (5% ACN/95% water/0.5% acetic acid) for 5 min, and a linear gradient was performed for 20-30 min to 60% Solvent B (95% ACN/5% water/0.5% acetic acid). The tryptic peptides were detected using data-dependent acquisition whereby a full scan between *m/z* 300.0-2000.0 was first obtained followed by a CID spectrum of the top 4 precursor ions (collision energy = 35%). ESI conditions were as follows: capillary temperature, 200 ^{0}C; sheath gas flow, 0 L/min; auxiliary gas flow, 0 L/min; ESI voltage, 1.20 kV; capillary voltage, 7.00 V; tube lens offset, -5.00 V. The CID mass isolation window was set to 2.00 *m/z* units. The subsequent information was input into the protein database searching programs Sequest

(ThermoQuest, San Jose, CA, USA) or MASCOT (Matrix Science Inc, Boston, MA, USA). Additional phosphopeptides were identified by investigating the neutral fragment chromatograms at *m/z* 98, 49 or 32.6 and manually interpreting the MS/MS data.

MALDI-TOF-MS Analysis of the β-eliminated Digests

The peptide samples of the untreated and β-eliminated digests of β-casein, α-casein, ovalbumin, BSA and CPK5 were prepared using a matrix solution consisting of 53 mM HCCA in 50% acetonitrile/0.1% TFA in a 1:1 ratio, that is, 1 µL sample to 1 µL matrix. The samples were then air dried at room temperature on a stainless steel plate. Mass spectra were obtained with a MALDI-TOF MS instrument equipped with a 337 nm nitrogen laser and reflectron optics. All spectra were acquired in positive ionization mode. The instrument was operated under delayed extraction conditions in reflectron mode, mirror voltage ratio of 1.12, a delay time of 150 ns and grid voltage 70% of full acceleration voltage (20 kV). Spectra were obtained with 100 laser shots at a laser intensity of 2625 and laser repetition rate of 3.0 Hz. An external calibration was performed before each spectrum was obtained with a calibration mixture consisting of 1.0 pmol Des-Arg^1-Bradykinin, 1.3 pmol Angiotensin I, 1.3 pmol Glu^1-fibrinopeptide B, 2.0 pmol ACTH (1-17 clip), and 1.5 pmol ACTH (18-39 clip). Once the spectra of the untreated and β-eliminated digests were obtained, they were superimposed and peaks present in the β-eliminated digest that were not visible in the untreated sample were investigated further as possible phosphopeptides that had been dephosphorylated.

Precursor Scanning

Tryptic digests were loaded into an EconoTip (New Objective, Woburn, MA) emitter and interfaced with the QSTAR instrument operated in precursor ion scanning mode for the PO_3^- fragment ion (-79 m/z). The needle voltage was maintained between -

700 V and -800 V while the declustering and focusing potentials were -70 V and -225 V, respectively. Mass spectra were acquired using a stepsize of 0.25 Da and a dwell time of 40 ms per m/z. The collision energy was adjusted to higher values (70-80 eV) in order to optimize the production of the phosphate-derived low-mass fragment ions using nitrogen as the collision gas.

Data-dependent LC/MS/MS on the QqTOF-MS

Capillary rpHPLC separation of protein digests was performed on a 15 cm x 75 um i.d. PepMap C18 column (LC Packings, San Francisco, CA) in combination with an Ultimate Capillary HPLC System (LC Packings, San Francisco, CA) operated at a flow rate of 200 nL/min. Inline mass spectrometric analysis of the column eluate was accomplished by a hybrid quadrupole time-of-flight instrument (QSTAR, Applied Biosystems, Foster City, CA) equipped with a nanoelectrospray source. A two-point mass calibration was performed in MS/MS mode of operation using the known fragment ion masses of [Glu]-Fibrinopeptide (*m/z* 175.119 and *m/z* 1056.475).

Results and Discussion

Kinase Autophosphorylation

Figure 3-1 shows autoradiography results of the P^{32}-labelled kinases demonstrating that that the kinases were indeed autophosphorylated under autophosphorylation conditions.

Ion Trap: Data-dependent LC/MS/MS of the Autophosphorylated CPK5 Digest

Analysis by Sequest software of data-dependent LC/MS/MS spectra of the unenriched sample identified the phosphopeptide NSLNI**pS**MR. An additional phosphopeptide DIY**pT**LSRK was revealed by manual analysis of the reconstructed neutral fragment chromatogram of 98 *m/z* and the corresponding MS/MS spectrum of

947.39 *m/z* (Figure 3-2). Analysis of the IMAC-enriched digest by manual inspection of the neutral fragment chromatograms at 98 m/z and 32.6 m/z revealed a triply charged phosphopeptide G**pS**FKDKLDEGDNNKPEDYSK at 789.73 *m/z* as well as the previously identified peptide DIY**pT**LSRK (Figure 3-3). This shows that without the use of the IMAC columns for enrichment, the former peptide would have gone unnoticed possibly due to suppression effects from the other unphosphorylated peptides present in the digest. Overall, three autophosphorylation sites were identified by this data-dependent LC/MS/MS method and IMAC enrichment.

Analysis of data from a similar experiment by MASCOT software resulted in the identification of two phosphorylation sites (DIY**pT**LSR and TMRNSLNI**pS**MR) previously identified by using both Sequest and neutral fragment scans. Neutral fragment scans from this run revealed MS/MS data at *m/z* values matching *m/z* values of the possible phosphopeptides TPNIRDIY**pT**L**pS**R and EMFQAMD<u>T</u>DN<u>S</u>GAI<u>T</u>FDELK (doubly phosphorylated but ambiguous as to which two sites). However, due to insufficient fragmentation, confirmation of these phosphorylation sites was not possible. It should also be noted that the phosphopeptide G**pS**FKDKLDEGDNNKPEDYSK identified previously by IMAC enrichment was not identified in this analysis possibly due to suppression effects. Overall, utilizing these various methods with the ion trap mass spectrometer resulted in the identification of six possible phosphorylation sites of CPK5.

MALDI-TOFMS

Several controls (β-casein, α-casein, ovalbumin, and BSA) were first used to test the MALDI-TOFMS β-elimination method as a first pass method for predicting peptide phosphorylation. Figure 3-4 shows the results of the superimposed spectra of the untreated and β-eliminated β-casein digests. Peaks that were present in the β-eliminated

spectrum and absent in the untreated digest were matched with *m/z* values of theoretically dephosphorylated phosphopeptides. That is, peaks representing potentially dephosphorylated peptides were compared to a list of all possible phosphorylated peptides of β-casein that was generated by the MS-Digest function of the database-searching program Protein Prospector. Both known phosphopeptides FQ**pS**EEQQQTEDELQDK and ELEELNVPGEIVE**pS**L**pSpSpS**EESITR were detected by this method, however, the monophosphorylated peptide was also indicated to be doubly phosphorylated and the tetraphosphorylated peptide also indicated to have six phosphorylation sites.

Although these possibly false sites were detected, both phosphopeptides were the only peptides identified as being phosphorylated. Similar results (not shown) were also attained for the other standards (phosphorylated and unphosphorylated proteins) whereby the known phosphopeptides were identified, however, several false positives were also identified. Although there were false positives, actual phosphopeptides were identified so it was decided that this approach could be used as a first pass method to find possible phosphopeptide targets for MS/MS verification.

This method was applied to the autophosphorylated CPK5 digest and three possible phosphopeptides were found, RTMRNSLNI**pS**MR also found with the LC/MS/MS Sequest and MASCOT analysis shown previously, L**pT**AHEVLRHPWICENGVAPDR and IIQRGHYSERKAAELTK (one phosphorylation site but exact site is uncertain) (Figure 3-5).

QqTOF-MS analysis

Precursor ion scanning of the autophosphorylated digest indicated the presence of at least 12 phosphorylation sites. This was done by manually performing peptide mass

mapping whereby all *m/z* values with the loss of PO_3^- obtained from the precursor ion scan were compared to a list of theoretically possible phosphopeptides generated by Protein Prospector MS Digest, however, sequence verifications have proven to be difficult. Figure 3-6 shows the precursor ion scan from 400-500 *m/z*. Among these were the three phosphopeptides previously identified by the ion trap experiments.

Data-dependent LC/MS/MS analysis followed by MASCOT database searching resulted in the identification of six phosphorylation sites, including previously identified sites DIYTL**pS**R, NSLNI**pS**MRDA and G**pS**FKDKLDEGDNNKPEDYSK, as well as additional sites N**pS**LNI**pS**MR, ML**pS**SKPAER, and TSTTNLSSNSDHSPNAADIIAQEFSK (1 site) (Figure 3-7). It should be noted that the phosphopeptide fragmentation on this instrument was significantly better than that of the ion trap as many of the b and y ions were generated on the QSTAR as compared with the ion trap which showed mainly the dephosphorylated peptide after collision-induced dissociation.

The development of a simple, cost effective method for phosphorylation identification using β-elimination and a MALDI-TOF MS has been shown to provide targets for sequence verification. This will be useful for laboratories with only a MALDI-TOF instrument, however, sequences will have to be verified once the targets are found due to possible false positives.

Comparison of results obtained by these three instruments gave significant overlap showing that these methods are indeed complimentary (Figure 3-8). Overall, at least 17 possible phosphorylation sites were observed (Figure 3-9). Of these sites, 5 sites are seen in at least two of these methods and 7 have been verified by MS/MS analysis. Combining

the three methods gave an increased identification of autophosphorylation sites that would not have been possible by any one method. Having these complimentary results gave an increase in the confidence that these sites are indeed valid, however, future experiments need to be performed to verify the sites, after which we will attempt to interpret their significance.

Because the autophosphorylation properties of CDPKs are not fully understood, localization of the autophosphorylation sites of an *Arabidopsis thaliana* member of this family by various complimentary methods are presented as a contribution towards their understanding. Compilation of these autophosphorylation sites with those of previously published sites such as that of the tomato (*Le*CPK1),[79] ice plant (McCPK1)[9] or *arabidopsis thaliana* (CPK1)[80] will aid in their understanding by possibly identifying a common sequence motif.

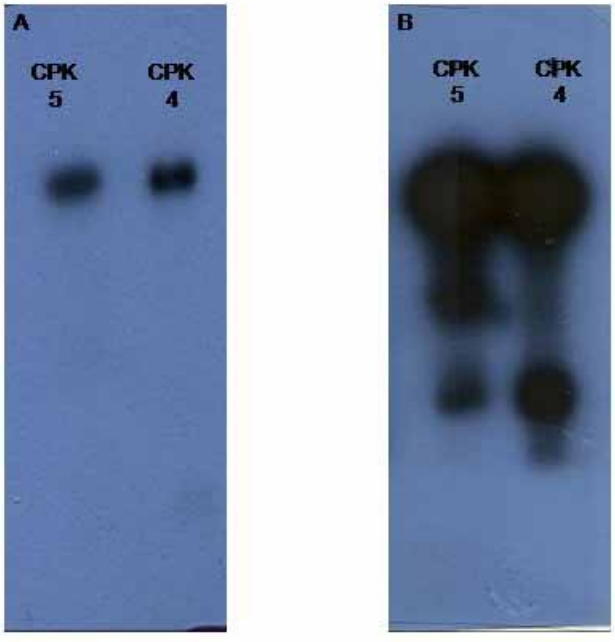

Figure 3-1. Autoradiography of SDS-PAGE separated autophosphorylated CPDKs. A) 2 hour exposure. B) 2 day exposure.

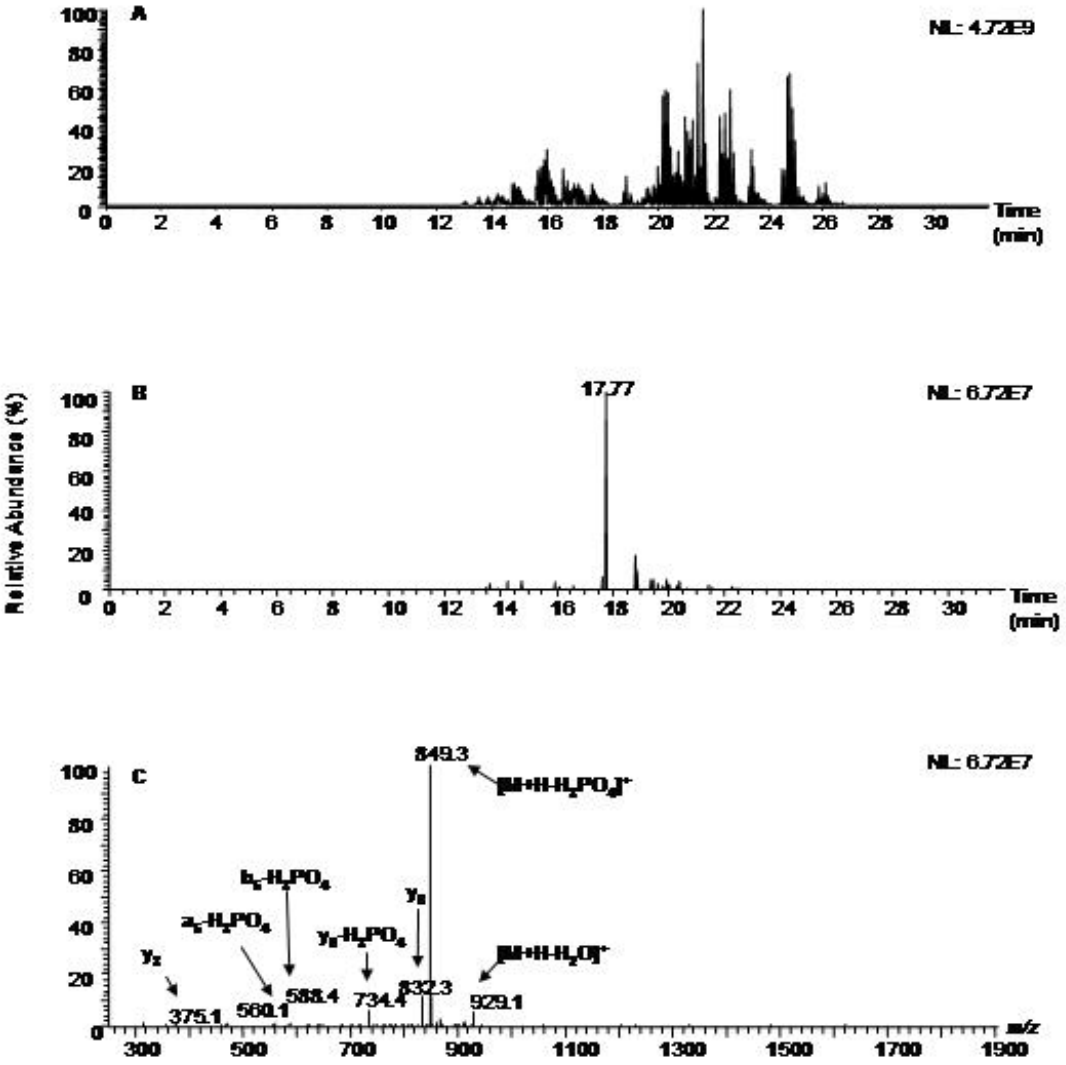

Figure 3-2. Autophosphorylated CPK5 digest. A) Base peak chromatogram. B) Reconstructed neutral fragment 98 chromatogram. C) MS/MS of the phosphopeptide DIY**pT**LSRK (947.39 m/z) located by the neutral fragment chromatogram at the retention time 17.77 min.

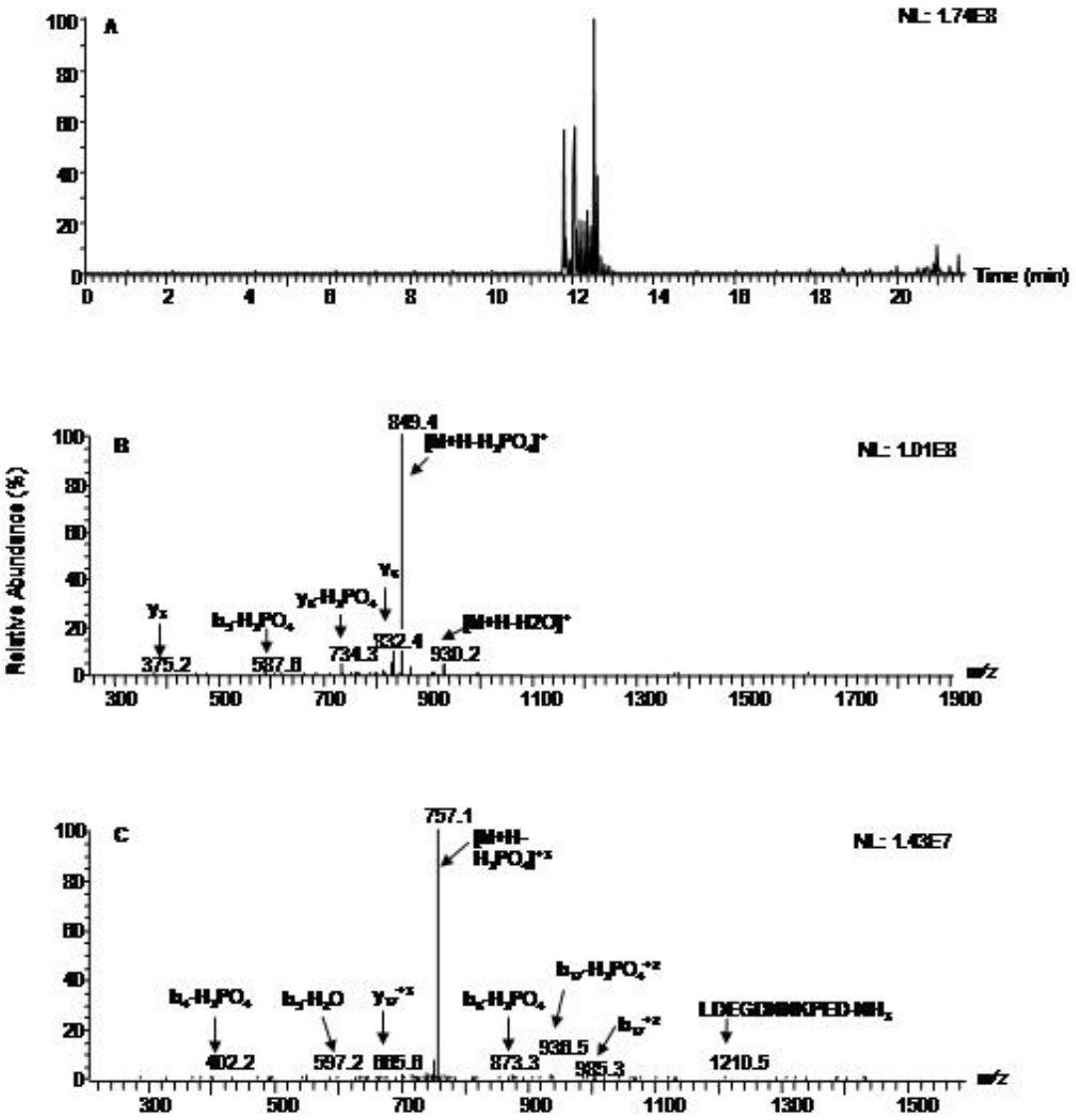

Figure 3-3. IMAC enriched phosphopeptides from autophosphorylated CPK5 digest. A) Base peak chromatogram. B) MS/MS of the phosphopeptide DIY**pT**LSRK located by the neutral fragment 98 chromatogram at 12.06 min. C) MS/MS of the phosphopeptide G**pS**FKDKLDEGDNNKPEDYSK located by the neutral fragment 32.6 chromatogram at 11.80 min.

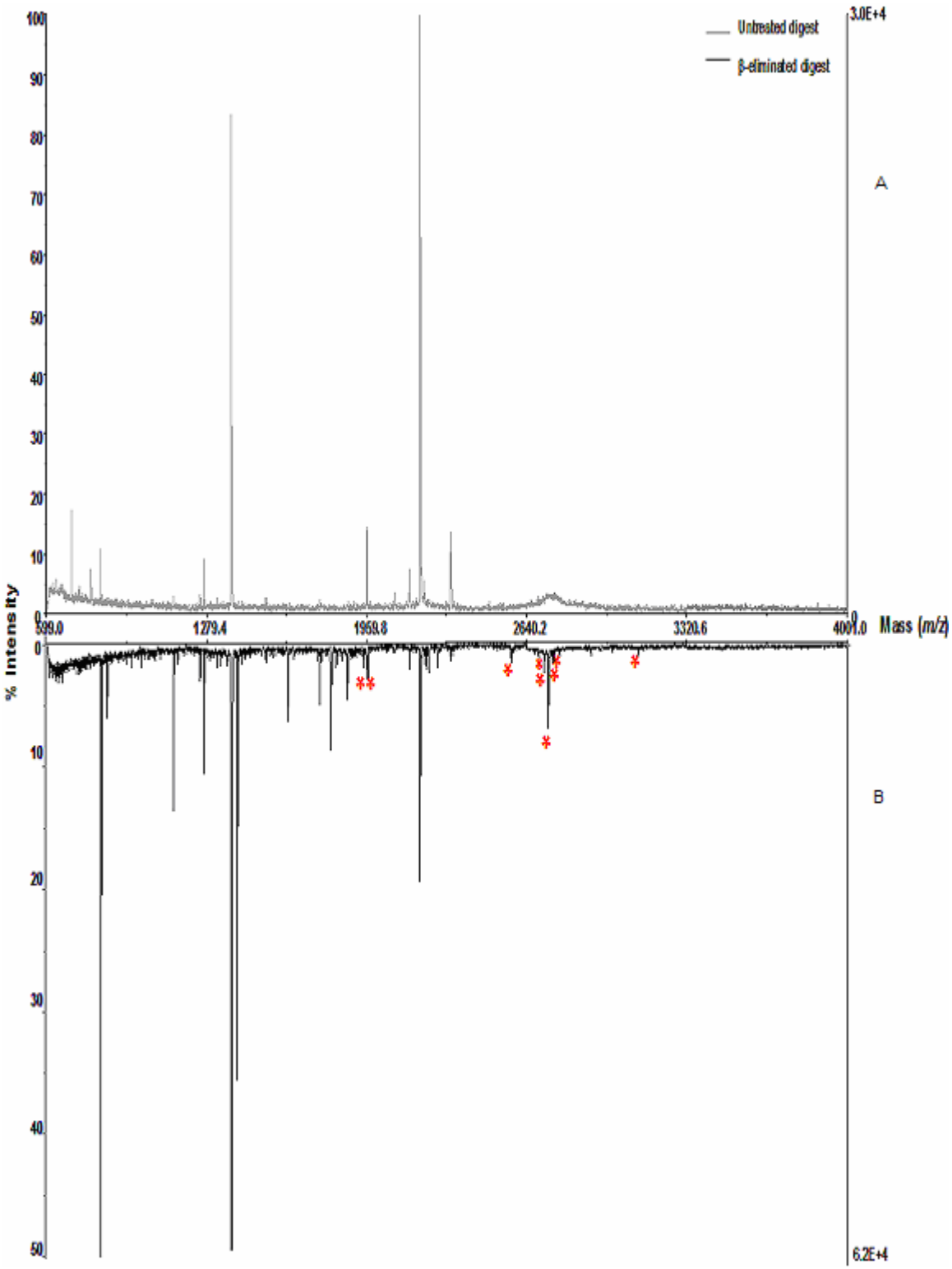

Figure 3-4. Spectra of β-casein digest. A) Untreated digest. B) β-eliminated digest. Dephosphorylated phosphopeptides in the β-eliminated digest are indicated with *.

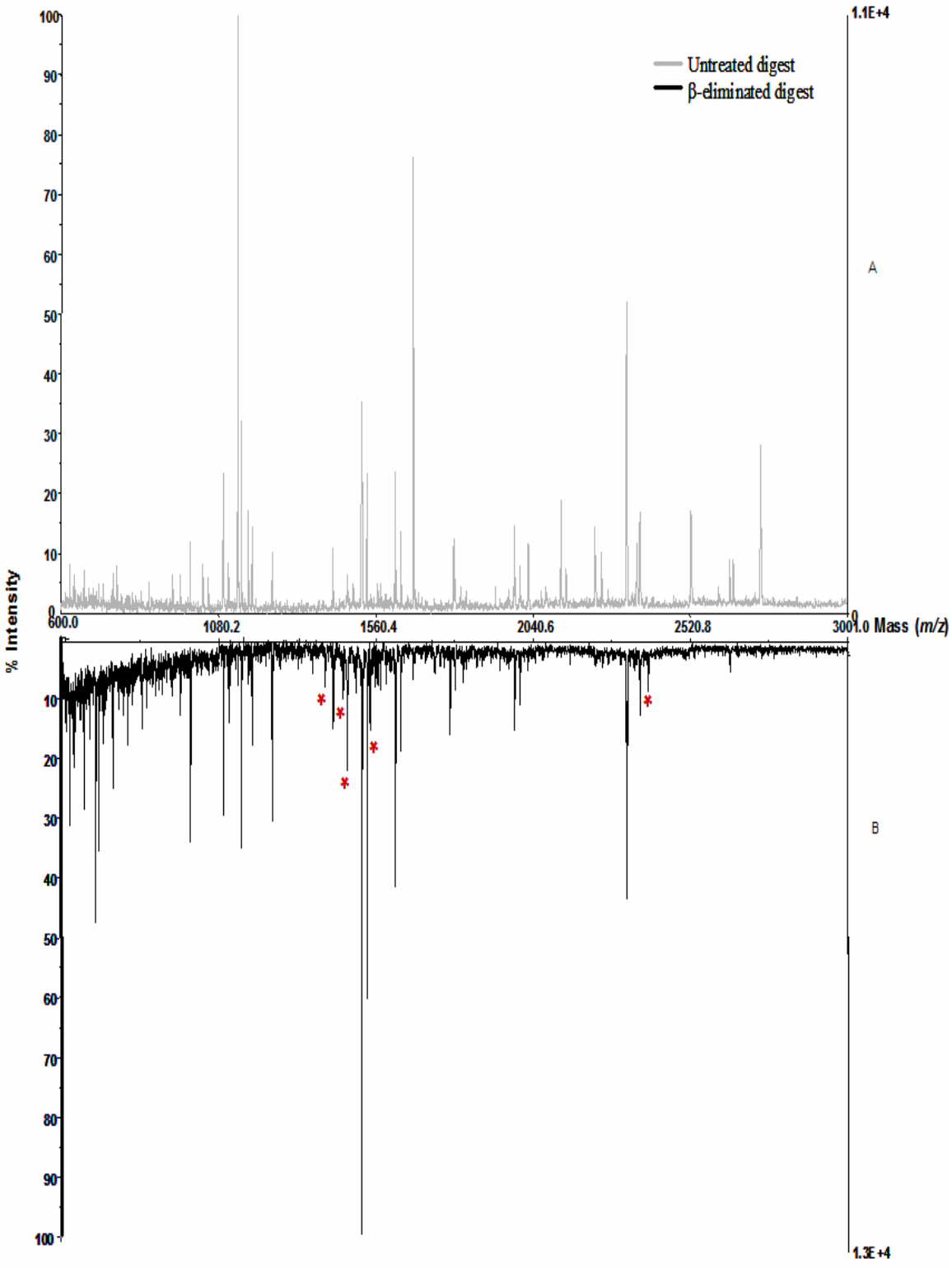

Figure 3-5. Spectra of CPK5 digest. A) Untreated digest. B) β-eliminated digest. Dephosphorylated phosphopeptides in the β-eliminated digest are indicated with *.

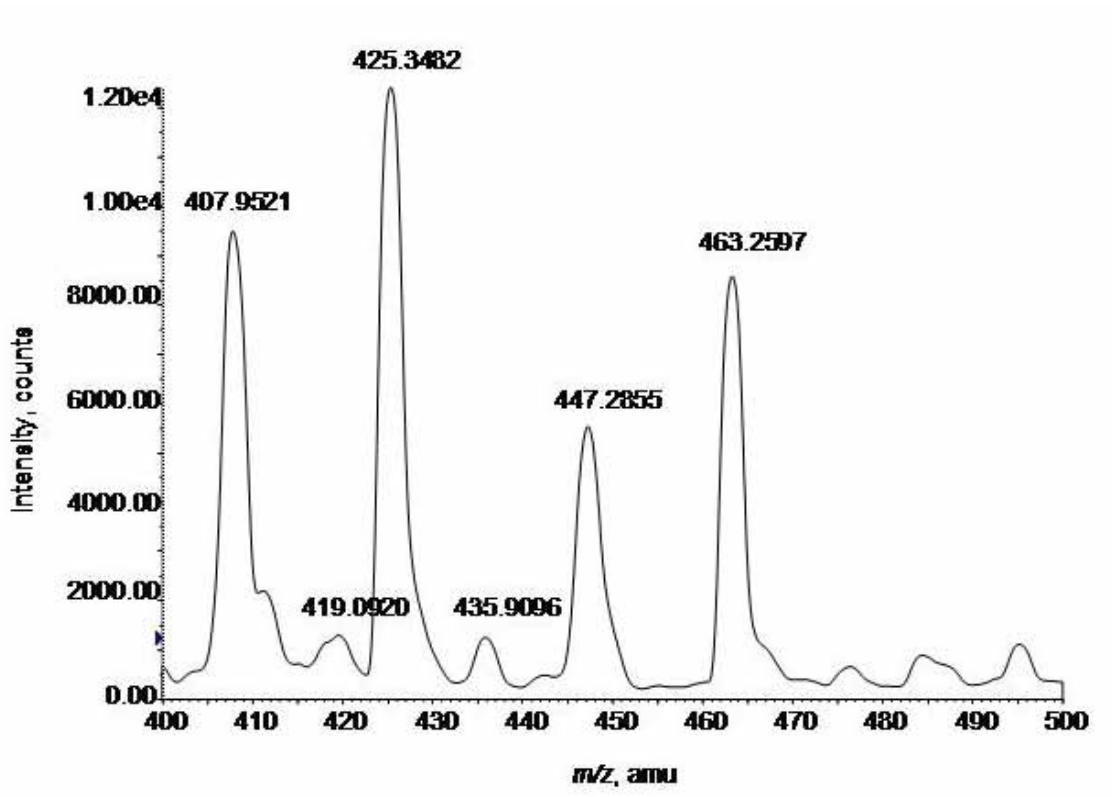

Figure 3-6. Precursor ion scan (400 - 500 *m/z*) of autophosphorylated CPK5 digest. *m/z* values were matched with predicted phosphorylated protein digests.

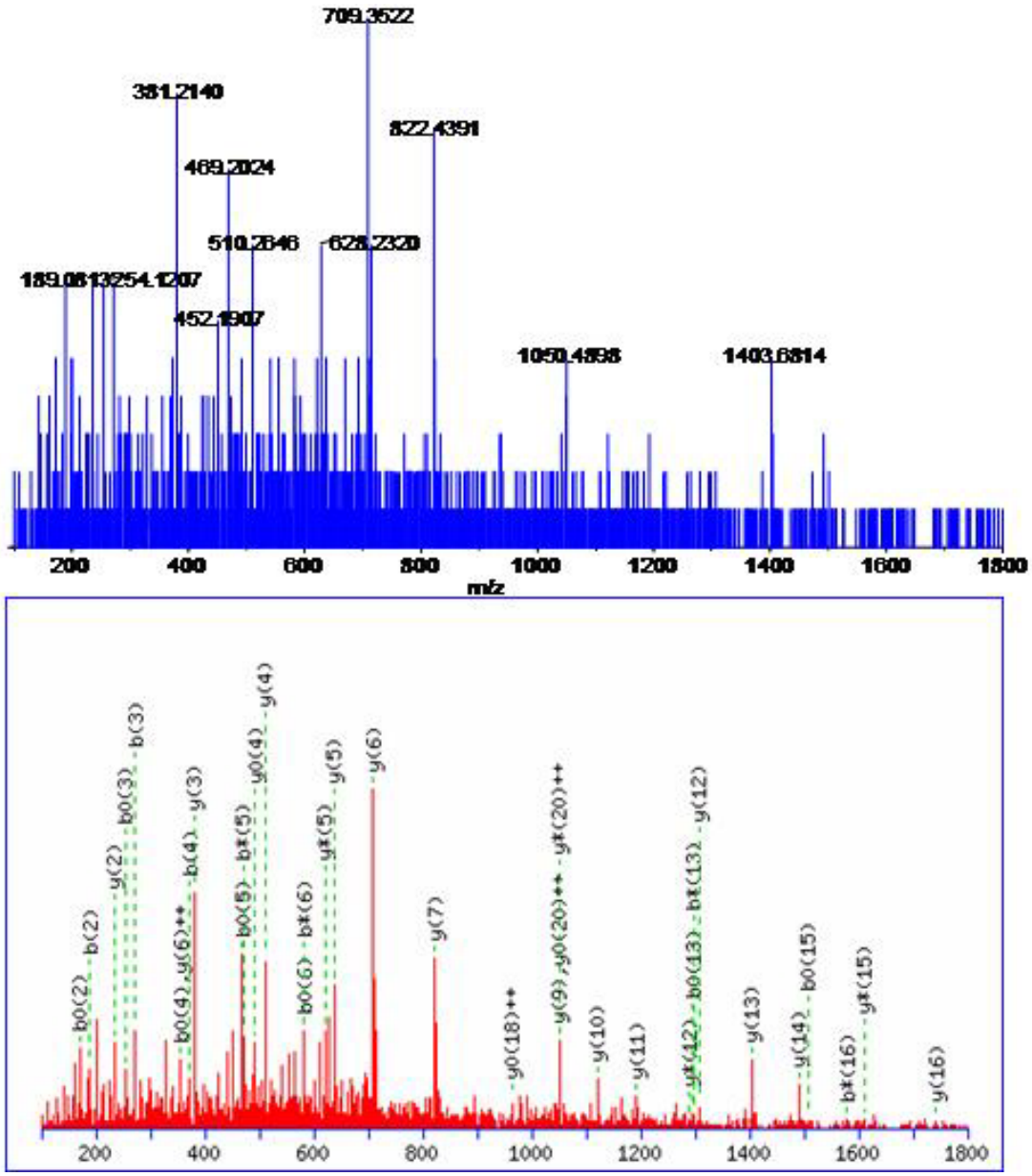

Figure 3-7. MS/MS spectrum and corresponding MASCOT search result of the phosphopeptide TSTTNLSSNSDHSPNAADIIAQEFSK (*m/z* 939).

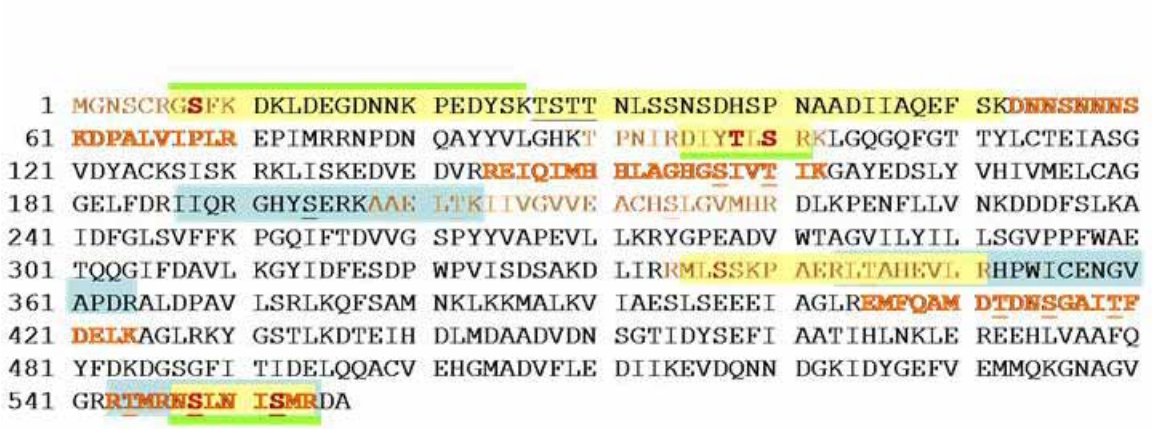

Figure 3-8. Overlap of phosphopeptides identified by all methods. Blue squares indicate peptides identified on the MALDI-TOF-MS, green squares indicate those identified on the QIT-MS, and yellow indicate those identified on the QSTAR. Sequences in orange are sequences identified by precursor ion scanning.

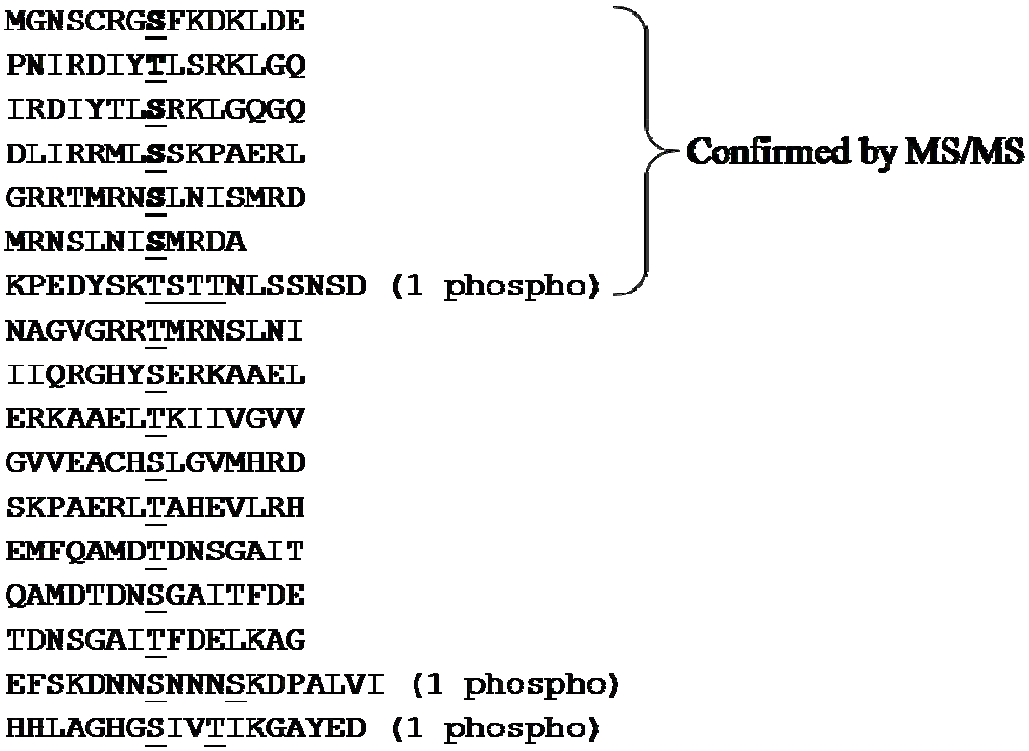

Figure 3-9. Phosphorylated peptides identified by all methods.

CHAPTER 4
IDENTIFICATION OF IN VITRO SUBSTRATES OF A CALCIUM-DEPENDENT PROTEIN KINASE

In order to gain an understanding of the specificity and function of calcium-dependent protein kinases, substrates of these kinases need to be determined. Several approaches may be applied to achieve this goal: 1) mass spectrometric identification of substrates phosphorylated by CDPKs *in vitro*, 2) tandem affinity purification and identification of *in vivo* substrates, and 3) CDPK substrate traps with yeast two-hybrid systems. For the purposes of this project, the mass spectrometric identification of *in vitro* substrates approach was determined as the method of choice due to the availability of mass spectrometry facilities.

Several problems are associated with this approach. First, if cell lysates are used as the source of substrate proteins, protein phosphatases and kinases already present in the cell lysate need to be inhibited. Second, all phosphates already present in the substrate proteins need to be removed prior to phosphorylation by the kinase in order to be able to determine the source of protein phosphorylation. Finally, since many important substrates may be present at low abundance, isolation and enrichment of these proteins is necessary.

Several commercial products were available to solve these problems. First, denaturation of proteins results in the inhibition of phosphatases and kinases. Several denaturing extraction methods have been compared by Saravanan and Rose[49] with the phenol-based method giving the highest protein yield and best resolution and spot intensity with gel electrophoresis. Invitrogen's TRIzol Reagent, a ready-to-use reagent

consisting of a mono-phasic solution of phenol and guanidine isothiocyanate, can be used for easy extraction of proteins from plant tissue. This occurs by the disruption of cells and dissolving of cell components by the reagent, resulting in an overall purified protein extract after the removal of RNA and DNA with chloroform and ethanol, respectively.

For dephosphorylation of proteins in the extracts, several phosphatases were available: calf intestinal alkaline phosphatase (CIP) (in-solution or immobilized on agarose), Antarctic phosphatase, and biotinylated phosphatase. Each of these phosphatases has its advantage. CIP in solution is the cheapest of the four and since derivatizations such as immobilization are not performed on the protein, active sites should be easily accessible for dephosphorylation of proteins in the extract. However, inhibition of the phosphatase after treatment and prior to kinase treatment is necessary. Agarose-immobilized calf intestinal phosphatase and biotinylated phosphatase have the advantage that they can be removed easily from the protein extract; however, as mentioned above, immobilization can possibly inhibit the activity of a percentage of the phosphatase due to blocking of the active sites. Antarctic phosphatase, which can be completely deactivated by a short heat treatment, also has the advantage of facile inhibition prior to kinase treatment. Since several types of phosphatases were available, determination of the most appropriate phosphatase for the purposes of this project was deemed necessary.

Finally, the development of a proprietary product from Qiagen (The PhosphoProtein Purification Kit) designed for the specific purification of phosphorylated proteins from complex cell lysates has made enrichment of low abundance proteins possible. The principle of this method is that proteins that carry a phosphate group on any

amino acid are bound with high specificity to a PhosphoProtein Purification Column, while proteins without phosphate groups do not bind to the column and can therefore be found in the column flow-through fraction. Binding of phosphorylated proteins occurs by flowing the lysate (~0.1 mg/mL) through the column at a flow rate of about 0.5 mL/min. Low lysate concentration and flow rate are used to ensure that all phosphate groups are easily accessible and are not hidden within protein complexes, and that complete binding of phosphorylated proteins occurs. Since 7-15% of proteins from cells are expected to carry one or more phosphate groups, the expected yield from one of these columns for 2.5 mg of protein from a cell lysate is 175–500 μg of phosphorylated protein. The maximum binding capacity of one of these columns is 500 μg of phosphorylated protein.

This chapter demonstrates the development of a method for the identification of substrates phosphorylated by kinases *in vitro*. A comparison of the performance of available phosphatases for dephosphorylating plant extract is shown as well as the comparison of various phosphatase inhibition methods. Application of the optimized dephosphorylation and inhibition steps followed by enrichment of the resulting *in vitro* phosphorylated substrates is shown.

Experimental Methods

Materials and Instruments

TRIzol® Reagent and NuPAGE 10% Bis-Tris SDS-PAGE gels were obtained from Invitrogen (Carlsbad, CA). Calf Intestinal Alkaline Phosphatase, Antarctic Phosphatase, Biotinylated Phosphatase and Streptavidin Magnetic Beads were purchased from New England Biolabs, Inc. (Beverly, MA). Immobilized Calf Intestinal Alkaline Phosphatase was purchased from Sigma (St. Louis, MO). The PhosphoProtein Purification Kit was from Qiagen Inc. (Valencia, CA). Pro-Q® Diamond Phosphoprotein Gel Stain was from

Molecular Probes (Eugene, OR). Sequencing-grade modified trypsin was purchased from Promega (Madison, WI). CPK4 was a gift from Estelle Hrabak.[81]

Mass spectrometric measurements were made using an LCQ Deca ion trap (ThermoFinnigan, San Jose, CA) equipped with a PicoView electrospray ionization source (New Objective, Ringoes, NJ) and an ABI 140D Solvent Delivery System (Perkin Elmer, Wellesley, MA).

Method Development

Protein extract preparation

Proteins were extracted from mature *Arabidopsis thaliana* leaves with TRIZOL® Reagent by grinding the leaves with a pre-chilled mortar and pestle and liquid nitrogen. Ground leaves were then transferred to a chilled Corex centrifuge tube and 5 mL of TRIZOL® Reagent was added for every 500 mg plant leaves. Ground leaves were allowed to sit in the TRIZOL® Reagent for 5 minutes at room temperature after which the sample was centrifuged at 10,900 x*g* for 5 minutes at 4^0C. The supernatant was then transferred to a fresh Corex tube, centrifuged for another 5 minutes, and the resulting supernatant transferred to an Oakridge tube. One mL of chloroform was then added and the tube shaken vigorously by hand for 15 seconds then allowed to stand at room temperature for 2 minutes. The sample was centrifuged at 10,900 x*g* for 15 minutes and the upper (aqueous) layer completely removed. 1.5 mL of absolute ethanol was then added to the lower phase (phenol-chloroform phase) and the sample mixed by inversion followed by incubation for 3 minutes at room temperature. Centrifugation of the sample at 483 x*g* was followed by the transfer of the supernatant to a fresh Corex tube. Three volumes of ice cold acetone was added to the supernatant followed by centrifugation for 2 minutes at 2,860 x*g*. The supernatant was then decanted, the acetone wash was

repeated, and the pellet was air dried. The protein pellet was solubilized with 1% SDS in 50 mM Tris-HCl, pH 7.5 for 30 minutes at 50 ^{0}C. The sample was diluted to obtain a final SDS concentration of 0.1%, and it was dialyzed against three changes of 1L of 50 mM Tris-HCl, pH 7.5. Dialyzed extract was concentrated by ultrafiltration in an Amicon concentrator containing a 10,000 molecular weight cut-off membrane. Protein quantification was performed by Bradford Protein Assay.

Dephosphorylation of the protein extract

Several phosphatases were used for dephosphorylation optimization of the protein extract: Calf Intestinal Alkaline Phosphatase, Antarctic Phosphatase, Biotinylated Phosphatase, and Immobilized Calf Intestinal Alkaline Phosphatase. All phosphatases were used according to the manufacturers' protocols and the resulting samples were separated on an SDS gel and stained with the Pro-Q® Diamond Phosphoprotein Gel Stain according to the manufacturer's protocol. CIP was determined as the most suitable phosphatase for these experiments.

Phosphatase inhibition

Since the CIP was in solution, several methods of phosphatase inhibition were tested: the addition of sodium orthovanadate, EDTA or phosphoserine. The model protein used for these studies was a fusion protein in which maltose binding protein was linked to the N-terminus of soybean serine acetyl transferase (MBP-SAT), which had previously been phosphorylated by CDPK4. Varying concentrations and times for each inhibitor were used to optimize the inhibition step. Also, kinase action after addition of these inhibitors was observed to ensure that the kinase was not being inhibited. This was done by adding CDPK4 to the inhibited sample with the needed phosphorylation buffer, 1mM

ATP, 10 mM $MgCl_2$, 1 mM EGTA, and 1.2 mM $CaCl_2$. The Pro-Q® Diamond Phosphoprotein Gel Stain was used to monitor all of these reactions.

***In vitro* phosphorylation of *Arabidopsis thaliana* extract with CDPK4**

Proteins were extracted from 2 g of mature *Arabidopsis thaliana* according to the method described previously. Removal of phosphate groups added to the proteins by *in vivo* phosphorylation was then performed by adding 6,750 units of CIP to the extract in a buffer containing 100 mM NaCl, 50 mM Tris-HCl, 10 mM $MgCl_2$, and 1 mM dithiothreitol, pH 7.9 at 37 ^{0}C for 1 hour. Sodium orthovanadate was then added to a final concentration of 10 mM and the sample incubated at room temperature for 5 minutes. Buffers and excess vanadate was removed from the extract by dialysis in 50 mM Tris buffer, pH 7.5. The dephosphorylated protein extract was then split into two equal aliquots for a control and an *in vitro* phosphorylated sample. Phosphorylation of the dephosphorylated extract was performed by incubation with CDPK4 in a buffer containing 1 mM ATP in 50 mM Tris, pH 7.5, 10 mM $MgCl_2$, 1 mM EGTA, and 1.2 mM $CaCl_2$ in the cold overnight, followed by incubation at room temperature for one hour. To the control sample, the same conditions were applied except for the addition of the kinase.

Phosphoprotein enrichment

Excess ATP was removed from the extract by dialysis in 50 mM Tris buffer, pH 7.5 followed by phosphoprotein enrichment with Qiagen's PhosphoProtein Purification columns according to manufacturer's protocol. Eluates and flow-throughs from both samples were then resolved on an SDS gel and visualized by staining with Pro-Q® Diamond Phosphoprotein Gel Stain according to protocol and Coomassie Brilliant Blue

R250. Gel bands were excised for in-gel tryptic digestion[54] and subsequently dried in a centrifugal vacuum system (SpeedVac) to near dryness.

Data-dependent LC/MS/MS on the ion trap

Samples were introduced to the ion trap mass spectrometer via an on-line reversed-phase capillary HPLC (75 μm x 5 cm C18 New Objective) with an isocratic solvent delivery at 200 nL/min with 3% Solvent A (5% ACN/95% water/0.5% acetic acid) for 5 min, and a linear gradient was performed for 85 min to 60% Solvent B (95% ACN/5% water/0.5% acetic acid). The tryptic peptides were detected using data-dependent acquisition whereby a full scan between *m/z* 300.0-2000.0 was first obtained followed by a CID spectrum of the top 4 precursor ions (collision energy = 38.0%). ESI conditions were as follows: capillary temperature, 190 ^{0}C; sheath gas flow, 0 L/min; auxiliary gas flow, 0 L/min; ESI voltage, 1.30 kV; capillary voltage, 36.0 V; tube lens offset, 10.00 V. The CID mass isolation window was set to 2.50 *m/z* units. The subsequent information was input into the protein database searching programs Turbo Sequest (ThermoQuest, San Jose, CA, USA) or MASCOT (Matrix Science Inc, Boston, MA, USA) and comparison of results was performed with the DTA Select program (ThermoQuest, San Jose, CA, USA).

Results and Discussion

Four phosphatases were investigated for dephosphorylation of the protein extract: Calf Intestinal Alkaline Phosphatase, Antarctic Phosphatase, Biotinylated Phosphatase, and Immobilized Calf Intestinal Alkaline Phosphatase (Figure 4-1). Dephosphorylation of the extracts was monitored by the Pro-Q® stain. According to the results obtained from the stain, CIP gave the best results with regards to dephosphorylation of the proteins (Figure 4-1A). The immobilized CIP was expected to be the best choice because of the

easy removal of the phosphatase. However, the dephosphorylation efficiency was lower than expected (Figure 4-1B) and there was also leakage of the phosphatase from the agarose. The biotinylated and Antarctic phosphatases also did not dephosphorylate the extract very well (Figure 4-1C and D). Although the biotinylated phosphatase was supposed to have the advantage of easy removal with streptavidin magnetic beads, all of the phosphatase was not removed by the beads. Heat inactivation of Antarctic Phosphatase by heating at 65 ^{0}C for 5 minutes worked very well with regards to denaturing the phosphatase; however, most of the proteins in the extract were precipitated by the heat treatment. Overall, the soluble calf intestinal alkaline phosphatase was the best choice since it gave the best dephosphorylation of the extract. The disadvantage of this phosphatase is that it had to be inhibited prior to the phosphorylation step.

Once the CIP was chosen as the phosphatase to be used for dephosphorylating the extract, a method of phosphatase inhibition that would not interfere with subsequent phosphorylation of the proteins by CDPK4 had to be determined. This was necessary to ensure that the phosphatase was not dephosphorylating as the kinase was phosphorylating the proteins. Several known methods of phosphatase inhibition were tested: inhibition of the active site of the phosphatase with vanadate, binding the required phosphatase cations with EDTA, and competitive inhibition with phosphoserine (Figure 4-2). Testing was performed by comparing the intensity of the Pro-Q stained band of MBP-SAT with the phosphatase followed by the inhibitor then the kinase with that of a control (MBP-SAT phosphorylated by CDPK4). Inhibition was monitored by observing whether the MBP-SAT was phosphorylated by the CDPK4 after the phosphatase in the sample was inhibited. Of the three methods tested, inhibition with vanadate gave the desired results of

both inhibiting the phosphatase as well as not interfering with the kinase action (Figure 4-2B), that is, the MBP-SAT was phosphorylated by the kinase after inhibition of the phosphatase by vanadate. EDTA and phosphoserine gave the same results, autophosphorylation of the CIP (formation of phosphoserine in the active site as a result of CIP's catalytic mechanism) and only a small degree of phosphorylation of the MBP-SAT. After comparing these three methods, it was determined that vanadate efficiently inhibited CIP and was also compatible with the CDPK4 phosphorylation step.

Once these individual steps were optimized, they could be combined for the overall schematic for identification of substrates of CDPK4 (Figure 4-3). The resulting 13.8 mg protein extracted from 2.0 g of leaves was used according to the sample preparation method in Figure 4-3. Resulting SDS-PAGE separation of the two samples showed more intense bands in the CDPK4 treated sample when visualized by both Pro-Q® and Coomassie staining (Figure 4-4). Unseparated protein bands were a result of the presence of salts and detergents from the elution buffer of the Qiagen kit. Twenty-seven equivalent gel slices were excised and digested from both samples for data-dependent LC/MS/MS (an example may be seen in Figure 4-5) and the resulting spectra analyzed by Turbo Sequest and MASCOT for protein identification. Identified proteins were then input into the DTA Select program for comparison of results. Proteins that were present in only the CDPK4-treated sample were identified as possible substrates. Table 1 shows a list of 29 possible substrates of CDPK4, each of which was identified by at least two peptides with significant database scores. Identified along with these 29 proteins was CDPK4 which is known to be autophosphorylated on a single serine residue. The identification of CDPK4 in this experiment shows that the method does work since CDPK4 is a positive control in

this sample. An additional 100 proteins were also identified from only the CDPK4 treated sample, however, only one peptide was matched by the database searching program, and the probability of false identification in this set of proteins is very high. Since this set of experiments were executed only once, repetition will be necessary to determine reproducibility.

Ideally, identification of the phosphorylation sites of these proteins would be further proof that these proteins are indeed substrates of the kinase. Unfortunately, phosphorylation site identification was very problematic due to the lability of the phosphate moiety. Reconstructed neutral fragment chromatograms showed many possible phosphopeptides, however, due to insufficient fragmentation of the peptides, sequence information for these peptides was not obtained. Because only the *m/z* for these peptides was known, and there were many possible peptide matches for each *m/z*, identification by database searching was not feasible. Also, manual interpretation of the corresponding MS/MS spectra was difficult, because several proteins were identified for each run, which makes identifying the corresponding protein difficult and mapping the phosphorylation site almost impossible.

We have developed a method for the identification of substrates of a kinase from a complex system. This was accomplished by the incorporation of several methods including the utilization of several newer technologies that have become available for phosphorylation analysis, namely, the Pro-Q® Diamond Phosphoprotein Gel Stain for visualization of the optimization steps, and the Qiagen PhosphoProtein Purification Kit for simplification of the sample, both of which eliminated the need for radio-labeling.

Also shown is the comparison of various phosphatases available as well as a means of phosphatase inhibition.

The obtained results of at least 29 potential substrates of CDPK4 may be used for future studies towards the understanding of the function of CDPK4 as well as the CDPK family. This may be accomplished once substrates of these kinases are identified and compared with the obtained results. Also, phosphorylation site mapping of the substrates may be more efficient with another type of mass spectrometer that will produce better fragmentation for database search matching. These proteins could also be used for future verification experiments whereby the purified proteins may be used for testing their interaction with the CDPK4 or other related kinases.

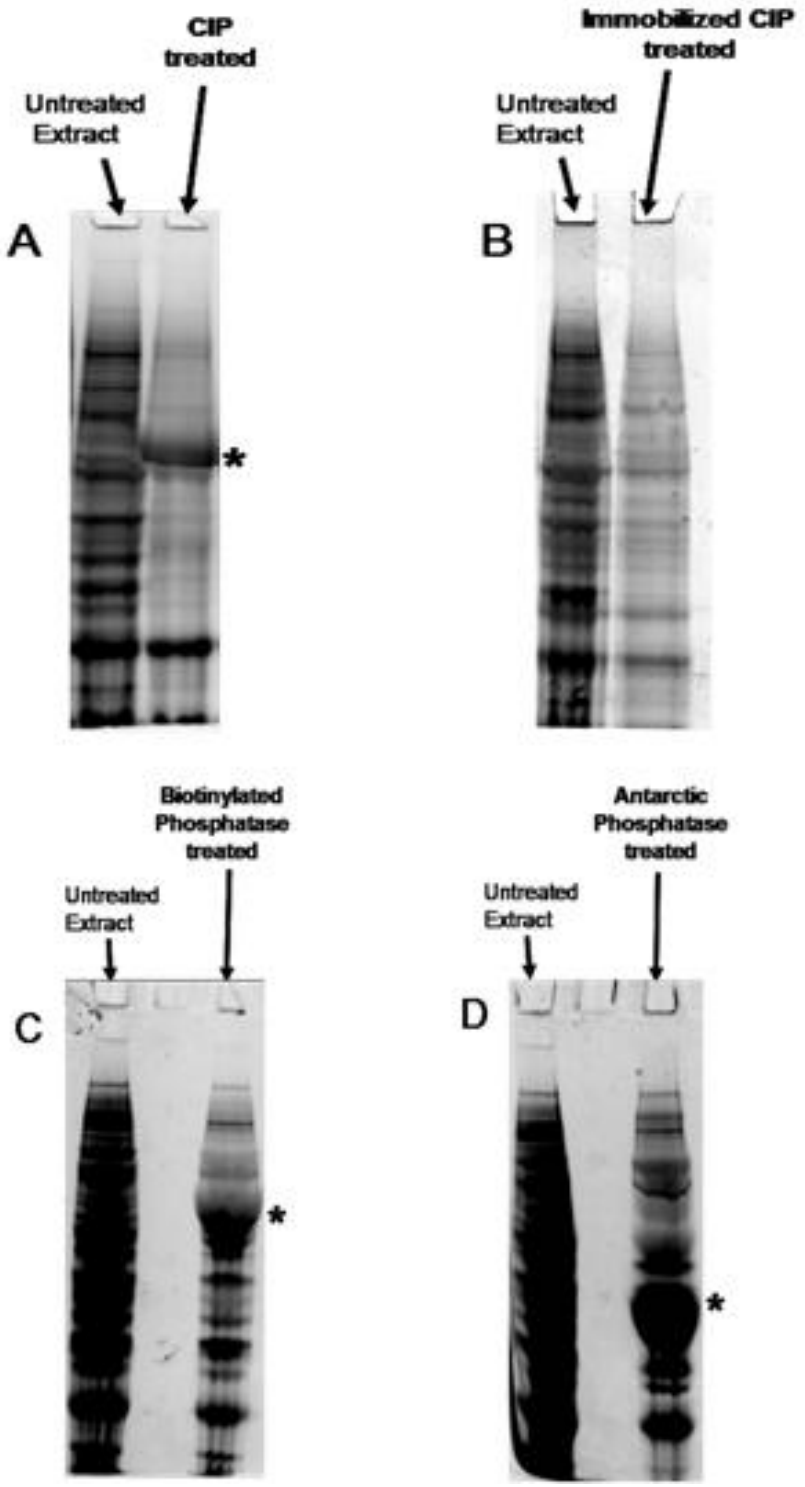

Figure 4-1. Pro-Q® Diamond Phosphoprotein Gel Stain of *Arabidopsis thaliana* extract dephosphorylation with phosphatases. A) Calf intestinal alkaline phosphatase (in solution). B) Immobilized calf intestinal alkaline phosphatase. C) Biotinylated phosphatase. D) Antarctic phosphatase. Phosphatase present in the extract is denoted by *.

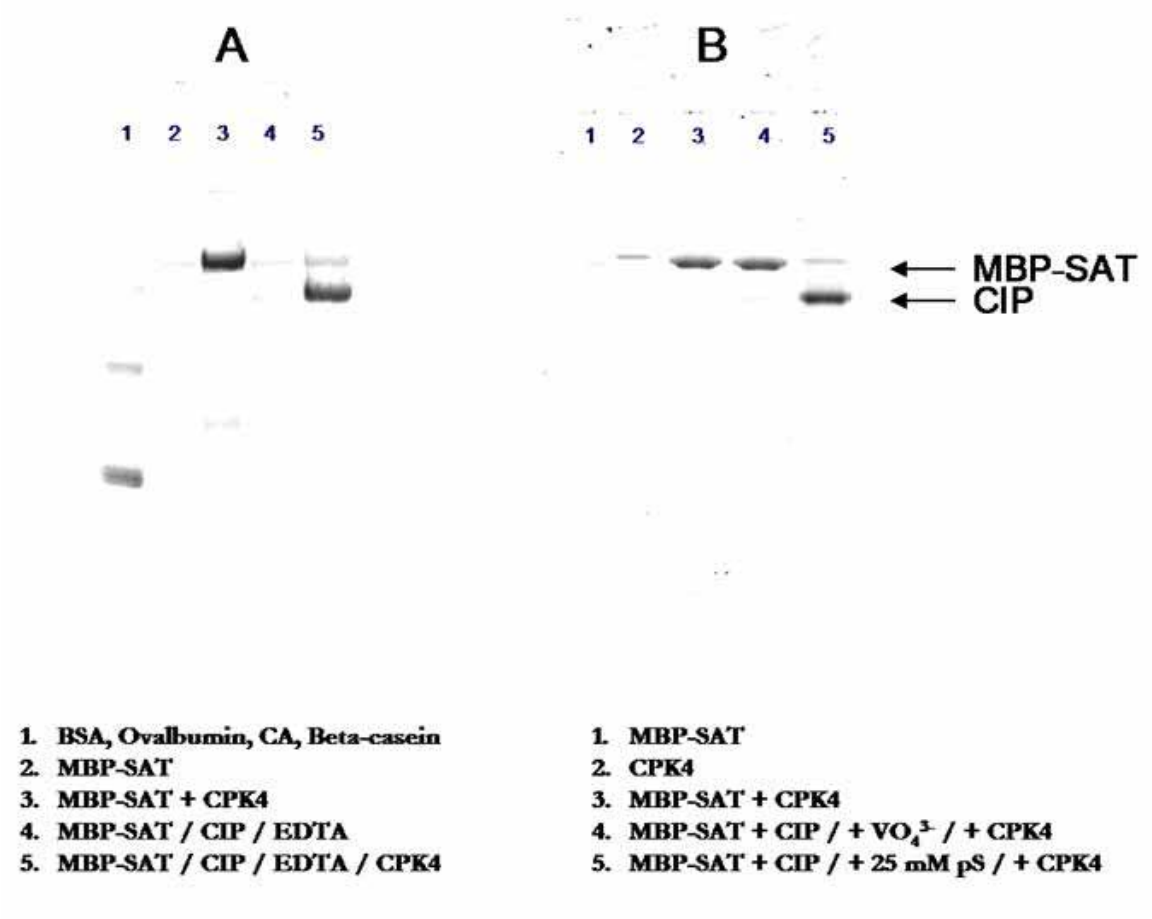

Figure 4-2. Inhibiting Calf Intestinal Alkaline Phosphatase. A) ProQ stained gel showing EDTA inhibition. B) ProQ stained gel showing vanadate and phosphoserine inhibition.

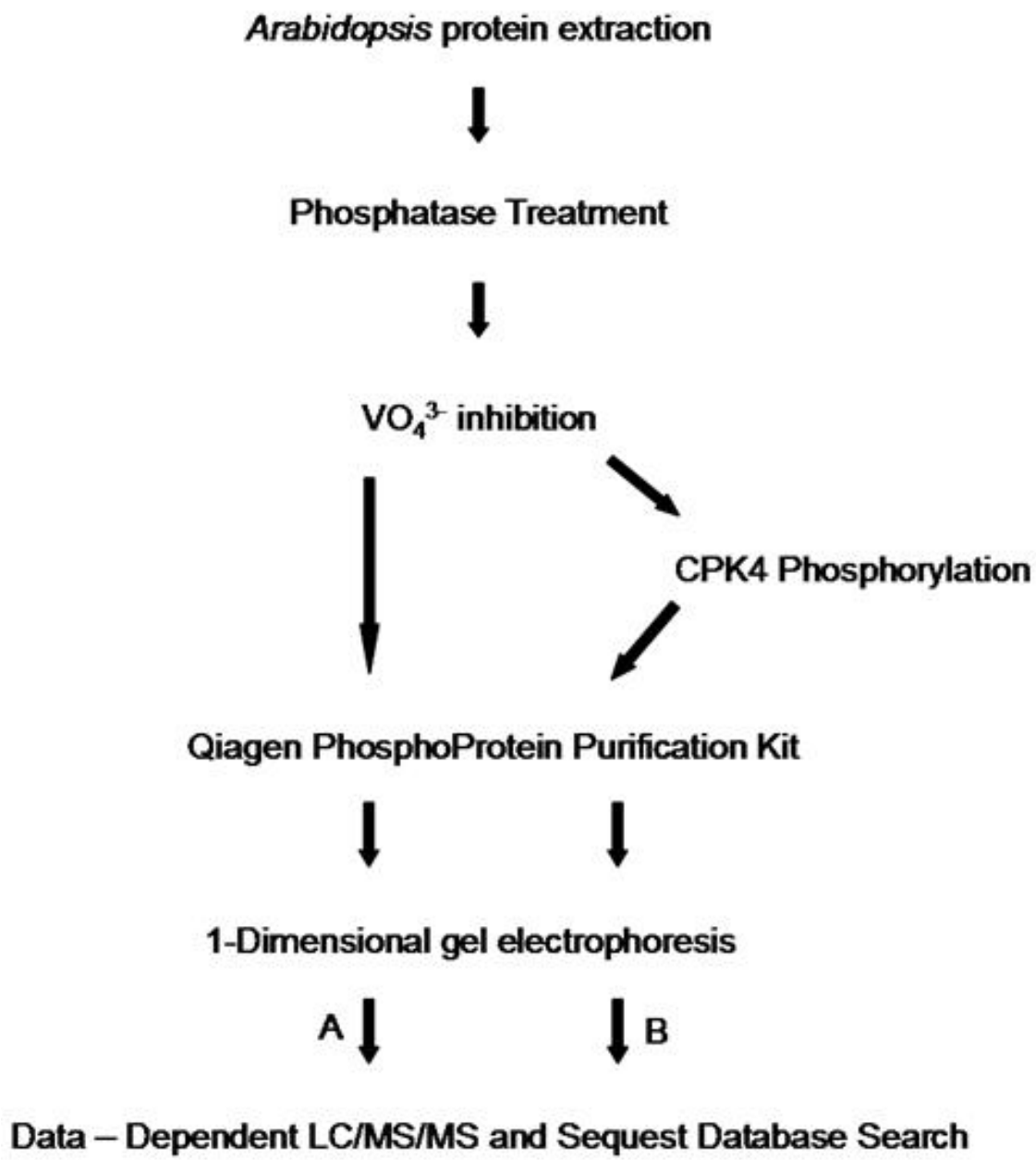

Figure 4-3. Steps for identifying *in vitro* CDPK4 substrates from *Arabidopsis thaliana.* **A** and **B** represent the control and CDPK4 treated samples, respectively.

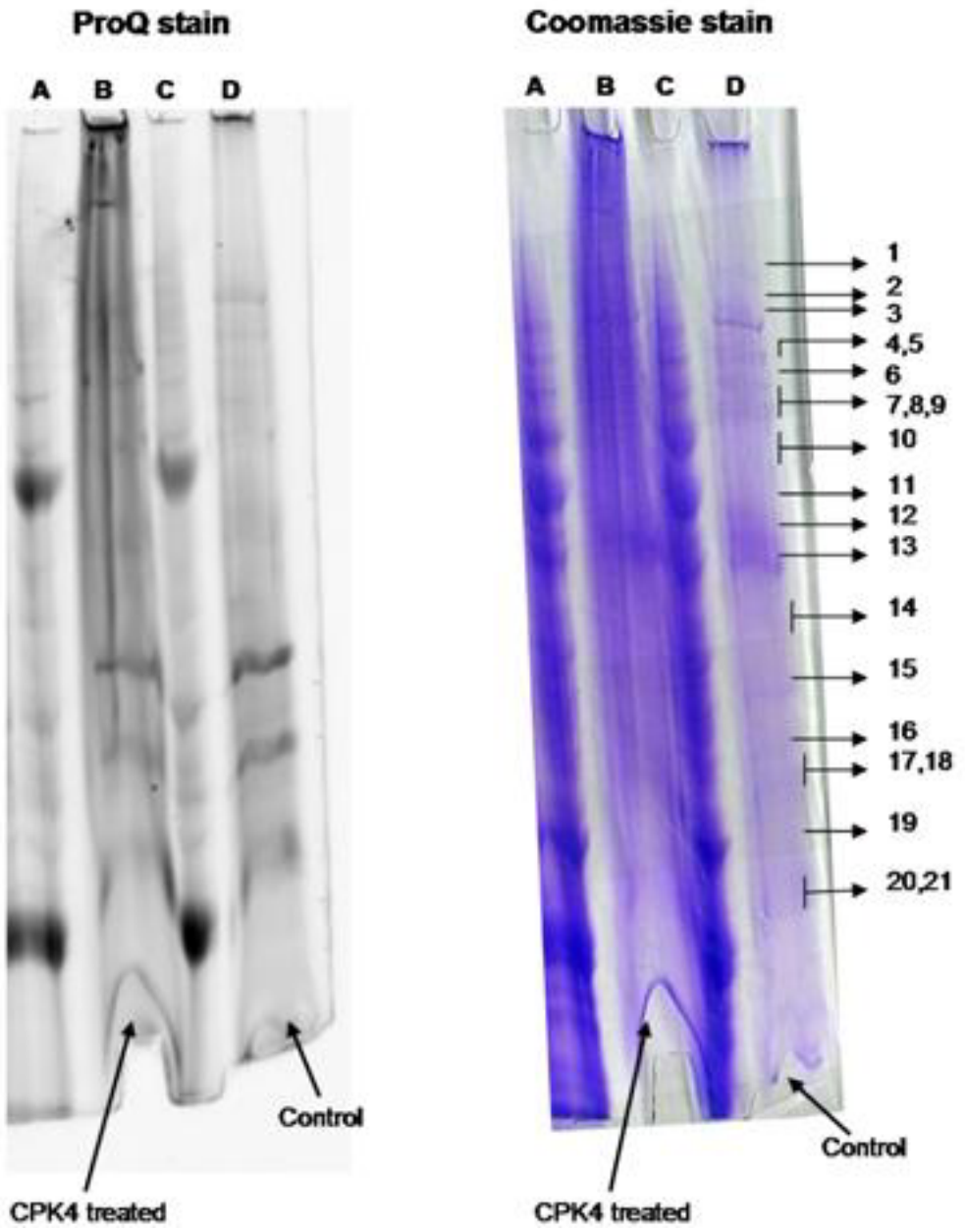

Figure 4-4. SDS gel electrophoresis separation of the control and CDPK4 treated Qiagen samples. A) CPK4 treated flow-through. B) CPK4 treated eluate. C) Control flow-through. D) Control eluate.

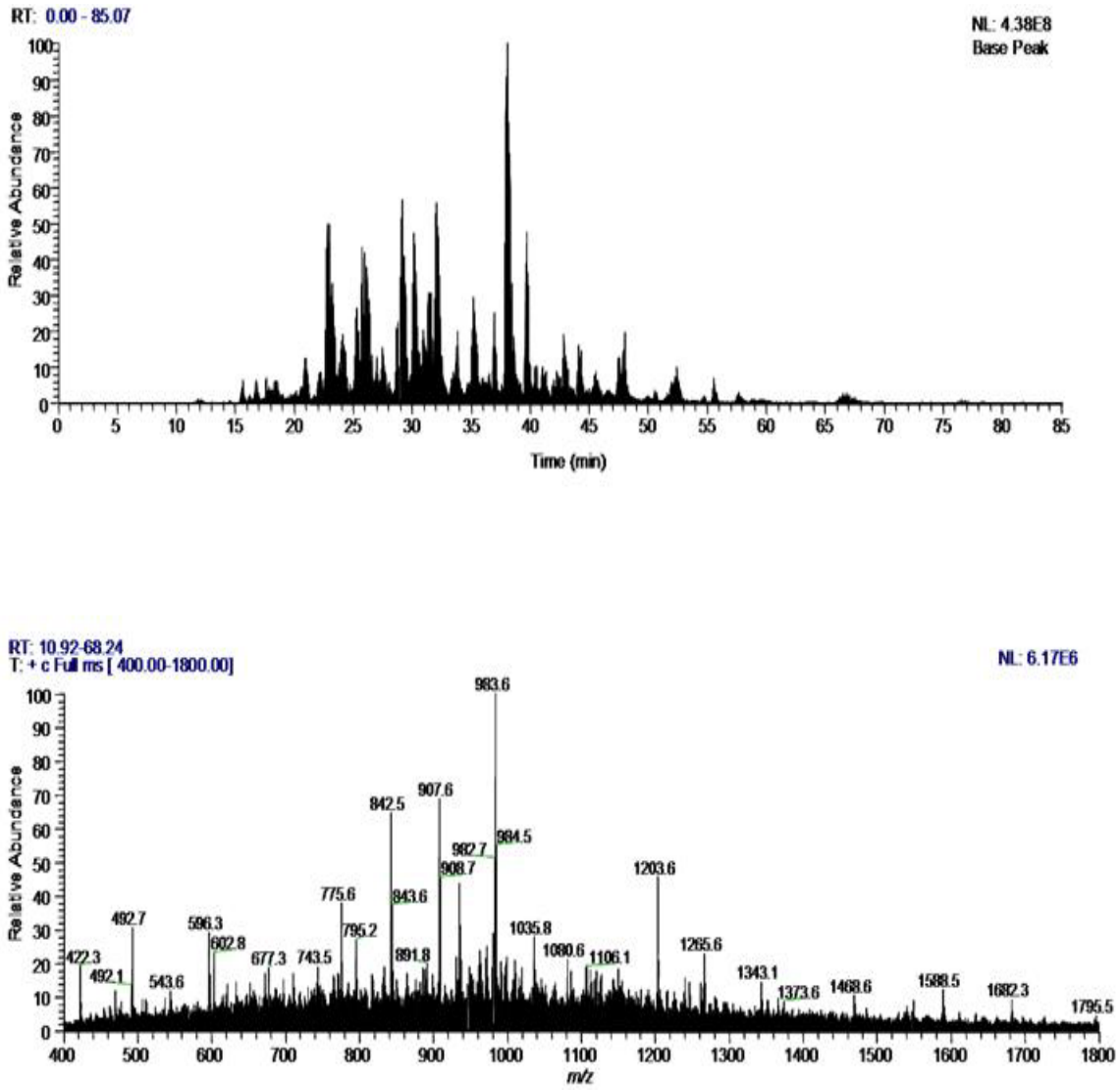

Figure 4-5. Base peak chromatogram and full MS spectrum of tryptic digest of band 20,21.

Table 4-1. List of proteins identified in only the CDPK4 treated sample.

Function or location	Provisional name	Gel band #
Photosynthesis /chloroplast	chlorophyll A-B binding protein	15, b1
	Photosystem I reaction centre subunit N, chloroplast precursor	16, 17,18
	Photosystem II 10 kDa polypeptide, chloroplast precursor	20,21
	Photosystem II 22 kDa protein, chloroplast precursor	20,21, b7
	Oxygen-evolving enhancer protein 2-2, chloroplast precursor	3, 16, 20
	Splice isoform 1 of chlorophyll a-b binding protein CP29.2, chloroplast precursor	7,8,9, 14, 19
	Cytochrome b559 alpha subunit	17,18, b4, b11
	Fructose-bisphosphate aldolase	15, b8
	Fructose-bisphosphate aldolase	15
Ribosomal	50S Ribosomal protein L4, chloroplast precursor	17,18, b10
	60S ribosomal protein L7-3	17,18, 19
	40S ribosomal protein S18	20,21
Nuclear	Nucleosome assembly protein	16, b9
Protein synthesis, processing, trafficking	Nascent polypeptide-associated complex (NAC) domain	17,18
	Similar to nascent polypeptide associated complex alpha chain	17,18
	Peptidyl-prolyl cis-trans isomerase, chloroplast precursor	b1, b2, b8
	Elongation factor 1-beta 1	16
Metabolism	ATP synthase B chain	15, 20,21
	ATP synthase alpha chain	3, 7,8,9, 10
	Putative H+-transporting ATP synthase	10, 17,18
	Glutathione S-transferase 11	20,21
	Glutathione S-transferase like protein	20,21
	Uridylyltransferase-related	17,18
	Serine hydroxymethyltransferase, mitochondrial precursor	11
	Glyceraldehyde-3-phosphate dehydrogenase B, chloroplast precursor	11, 14, 15, b8
	Phosphoribulokinase, chloroplast precursor	b8
	Splice Isoform 2 of Ribulose bisphosphate carboxylase/oxygenase activase, chloroplast precursor	3, 4,5, 7,8,9, 10, 11, b2, b7
Signaling/Stress	Calcium-dependent protein kinase, isoform 4	2, 3, 4,5, b1, b3
	Heat shock cognate 70 kDa protein 2	3, 7,8,9
Other	myrosinase-associated protein	3, 11, 13, 14

CHAPTER 5
14-3-3 INTERACTORS FROM ARABIDIOPSIS THALIANA

As mentioned previously, 14-3-3 proteins regulate many cellular processes by binding to phosphorylated sites in diverse target proteins. In plants, they have emerged as important regulators of phosphorylated enzymes of biosynthetic metabolism, ion channels, and regulators of plant growth.[36] Despite all that is known about these proteins, wide-scale studies of all interactors from a plant have not been performed to our knowledge.

One means of performing this task is to utilize protein-protein affinity chromatography which is a purification method that exploits the unique interaction of one molecule with a second, complementary binding molecule (ligand). The basic procedure for affinity chromatography is shown in Figure 5-1. First, the ligand is covalently coupled to an insoluble matrix such as agarose, and the resulting slurry poured into a column. An impure mixture containing the protein to be isolated is then applied to the column for interaction of the protein of interest with the ligand. That is, the protein of interest will bind specifically to the immobilized ligand while all other proteins will not. The unbound proteins will then wash off the column while the bound interacting protein will remain on the column. Interacting proteins can then be removed from the column by using an elution buffer that will disrupt the protein's interaction with the ligand.

The application of affinity chromatography may be used for the purposes of this project for the identification of interactor proteins from *Arabidopsis thaliana*. Utilization of an entire leaf extract will provide a means for identifying all interactors present, that is,

not only proteins of a certain class. Additionally, the use of denatured proteins will also aid in determining binding motifs for these proteins after the investigation of phosphorylation motifs of identified interactors. Presented is the application of 14-3-3 affinity chromatography for isolating *Arabidopsis thaliana* interactor proteins.

Experimental Methods

Materials and Instruments

TRIzol® Reagent and NuPAGE 10% Bis-Tris SDS-PAGE gels were obtained from Invitrogen (Carlsbad, CA). Microcon YM-10 concentrators were from Millipore (Billerica, MA). Pro-Q® Diamond Phosphoprotein Gel Stain was from Molecular Probes (Eugene, OR). Sequencing-grade modified trypsin was purchased from Promega (Madison, WI).

A FAMOS autosampler from LC Packings (Sunnyvale, CA) was used for automated sample loading. Capillary rpHPLC separation of protein digests was performed on a 15 cm x 75 um i.d. PepMap C18 column from LC Packings (San Francisco, CA) in combination with an Ultimate Capillary HPLC System (LC Packings, San Francisco, CA). Inline mass spectrometric analysis was accomplished by a hybrid quadrupole time-of-flight instrument (QSTAR, Applied Biosystems, Foster City, CA) equipped with a nanoelectrospray source.

Protein Extract Preparation

Proteins were extracted from mature *Arabidopsis thaliana* leaves with TRIZOL® Reagent by grinding the leaves with a pre-chilled mortar and pestle and liquid nitrogen. Ground leaves were then transferred to a chilled Corex centrifuge tube and 5 mL of TRIZOL® Reagent was added for every 500 mg plant leaves. Ground leaves were allowed to sit in the TRIZOL® Reagent for 5 minutes at room temperature after which the

sample was centrifuged at 10,900 x*g* for 5 minutes at 4^0C. The supernatant was then transferred to a fresh Corex tube, centrifuged for another 5 minutes, and the resulting supernatant transferred to an Oakridge tube. One mL of chloroform was then added and the tube shaken vigorously by hand for 15 seconds then allowed to stand at room temperature for 2 minutes. The sample was centrifuged at 10,900 x*g* for 15 minutes and the upper (aqueous) layer completely removed. 1.5 mL of absolute ethanol was then added to the lower phase (phenol-chloroform phase) and the sample mixed by inversion followed by incubation for 3 minutes at room temperature. Centrifugation of the sample at 483 x*g* was followed by the transfer of the supernatant to a fresh Corex tube. Three volumes of ice cold acetone was added to the supernatant followed by centrifugation for 2 minutes at 2,860 x*g*. The supernatant was then decanted, the acetone wash was repeated, and the pellet was air dried. The protein pellet was solubilized with 1% SDS in 50 mM Tris-HCl, pH 7.5 for 30 minutes at 50 ^{0}C. The sample was diluted to obtain a final SDS concentration of 0.1%, and it was dialyzed against three changes of 1L of 50 mM Tris-HCl, pH 7.5. The sample was washed with binding buffer (50 mM Tris-HCl, 250 mM NaCl, and 5 mM $MgCl_2$, pH 7.5) prior to concentration. Protein quantitation was then performed by Bradford Protein Assay.

14-3-3 Affinity Purification

14-3-3 isoforms immobilized on agarose were provided by Dr. Paul Sehnke and Dr. Robert Ferl, Dept. of Horticultural Sciences, University of Florida. The 14-3-3 resin was washed with 10 mL of binding buffer (50 mM Tris-HCl, 250 mM NaCl, and 5 mM $MgCl_2$, pH 7.5). Sample volume was brought up to 10 mL with binding buffer and incubated with the resin for 2 days in a cold room with constant rotation, followed by incubation at room temperature for 1 hour. The protein extract was then flowed off the

column and the column washed with 6 mL of binding buffer. Interacting proteins were then eluted from the resin by washing with 10 aliquots of 500 µL elution buffer (1M acetic acid, 500 mM NaCl, pH 2). Once eluted, samples were neutralized with 1M NaOH and concentrated with microcon concentrators (10,000 molecular weight cut-off) according to the manufacturer's instructions. Eluted proteins were separated on an SDS gel and proteins were visualized by staining and imaging with the Pro-Q Diamond Phosphoprotein stain and Coomassie stain. Gel bands were excised and digested with trypsin overnight. The digested samples were then collected and placed in the FAMOS autosampler for mass spectrometric analysis.

Amino Acid Sequencing by nanoESI QqTOF MS Analysis

MS/MS experiments for peptide sequencing were performed by loading 10 µL of each sample at 10 µL/min for 5 minutes with a FAMOS autosampler onto the C18 precolumn for desalting and concentration of the sample. The switching-valve position was then changed and the trapped peptides were back flushed and separated on the C18 nano column operated at a flow rate of 200 nL/min with a gradient of 5% to 60% acetonitrile over 30 minutes. MS/MS data were acquired by Information Dependent Acquisition (IDA) mode.

Protein Identification

Fragment ion data generated by Information Dependent Acquisition (IDA) via the QSTAR were searched against the NCBI nr sequence database using the Mascot (Matrix Science, Boston, MA) database search engine. Probability-based MOWSE scores above the default significant value were considered for protein identification in addition to validation by manual interpretation of the MS/MS data.

Results and Discussion

Protein extraction from 2.0 g of *Arabidopsis thaliana* leaves resulted in 6.4 mg of protein extract for incubation with the 14-3-3 resin. Visualization of the untreated protein extract, 14-3-3 flow-through and eluate by Pro-Q and Coomassie staining can be seen in Figure 5-2. As can be seen, many protein bands were intensely stained with the Pro-Q Diamond Phosphoprotein stain, indicating that phosphorylated proteins were affinity purified by the 14-3-3 resin. However, once stained by Coomassie, several bands were only faintly stained indicating that these proteins were present at only low amounts. Mass spectrometric analysis of the tryptic digests of these proteins resulted in the identification of 263 proteins (Table 5-1), several of which were identified as being phosphorylated by the MASCOT search engine. Figures 5-3 through 5-25 show the MASCOT search results for the identified phosphorylated proteins. It should be noted that some of these results show insufficient fragmentation resulting in overall low scores for MASCOT, however, according to the peptide mapping it seems as though most of these phosphorylation sites are real. Additionally, low fragmentation for phosphopeptides is typical for MS/MS analysis.

The results presented here show the 14-3-3 affinity purification of 263 putative phosphorylated proteins from *Arabidopsis thaliana*. Interpretation of the biological significance of these results is yet to be determined. However, additional experiments need to be performed to ensure that these proteins interacted with the 14-3-3s because of phosphorylation. One way of performing this would be to dephosphorylate the extract prior to affinity purification with the 14-3-3s. If interaction of the protein does not occur after dephosphorylation of the protein then it may be concluded that the interaction was primarily because of phosphorylation.

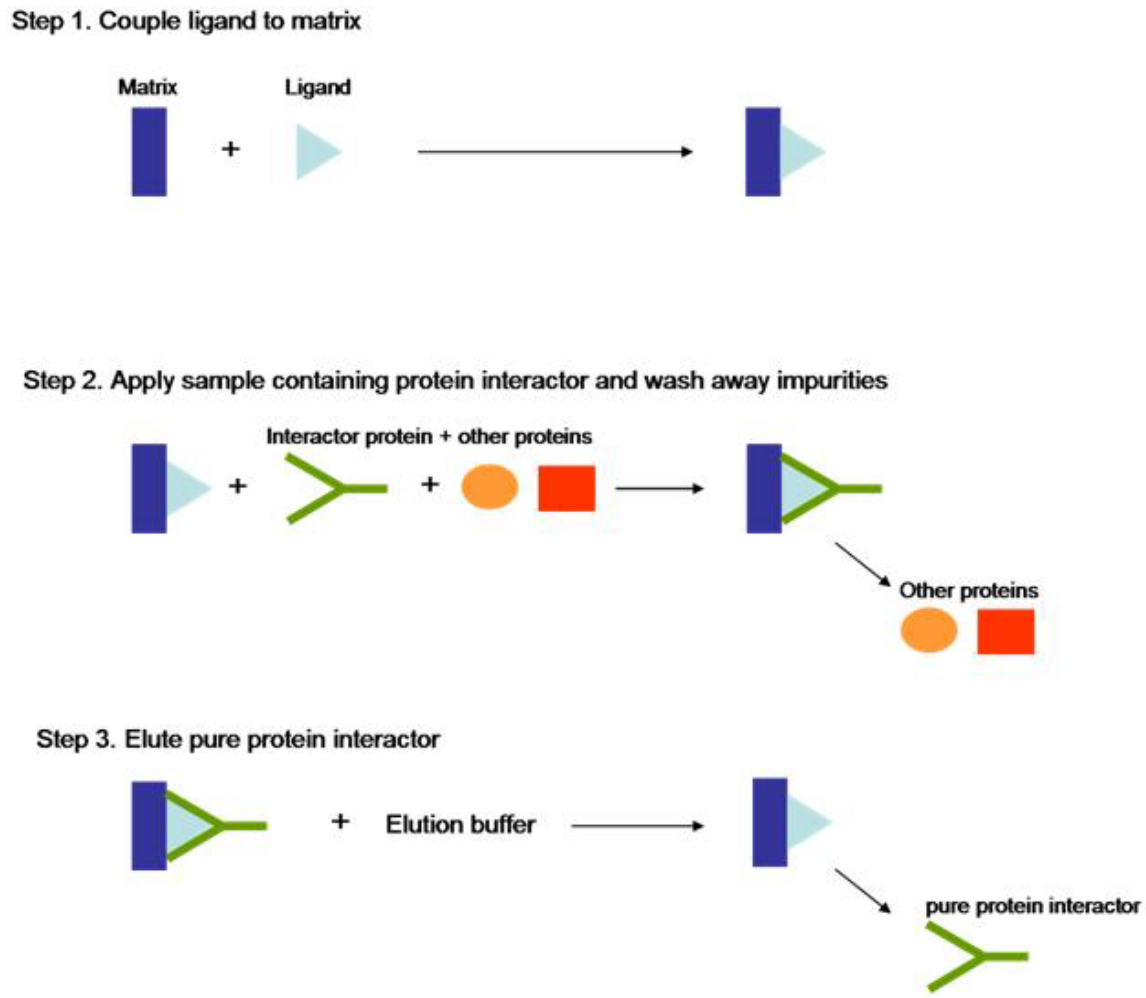

Figure 5-1. General procedure for affinity chromatography.

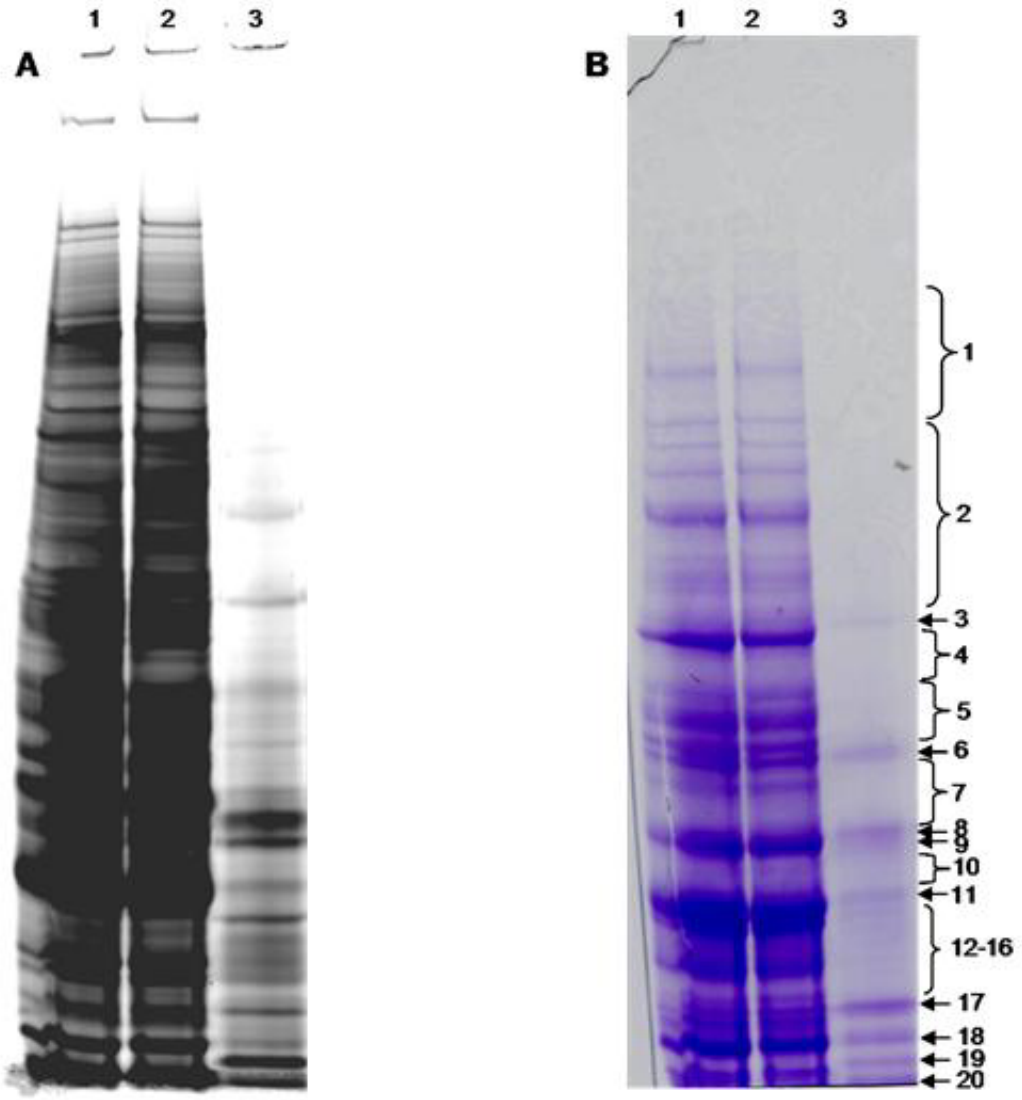

Figure 5-2. Gel images of 14-3-3 affinity purified *Arabidopsis thaliana* proteins. A) Pro-Q Diamond Phosphoprotein Stain. B) Coomassie Stain. Lanes 1, 2, and 3 are TRIzol protein extract, 14-3-3 flow-through, and 14-3-3 eluate, respectively.

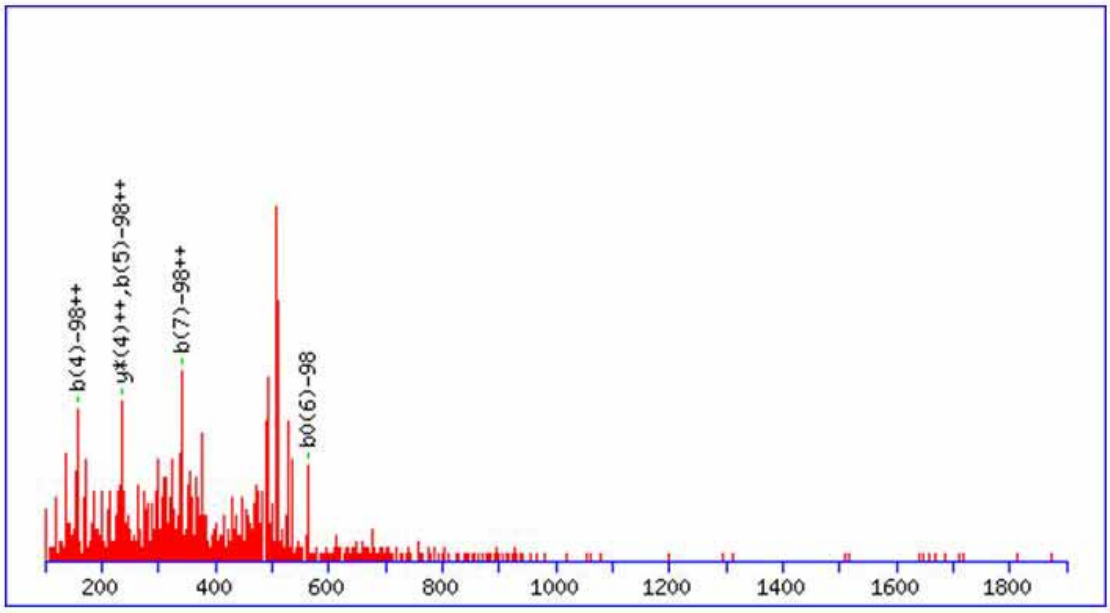

Figure 5-3. MASCOT search results of the phosphopeptide NAGpSRLVVR (m/z 1050.5895) from photosystem I subunit PSI-E-like protein (gi|7269730) identified in band 2.

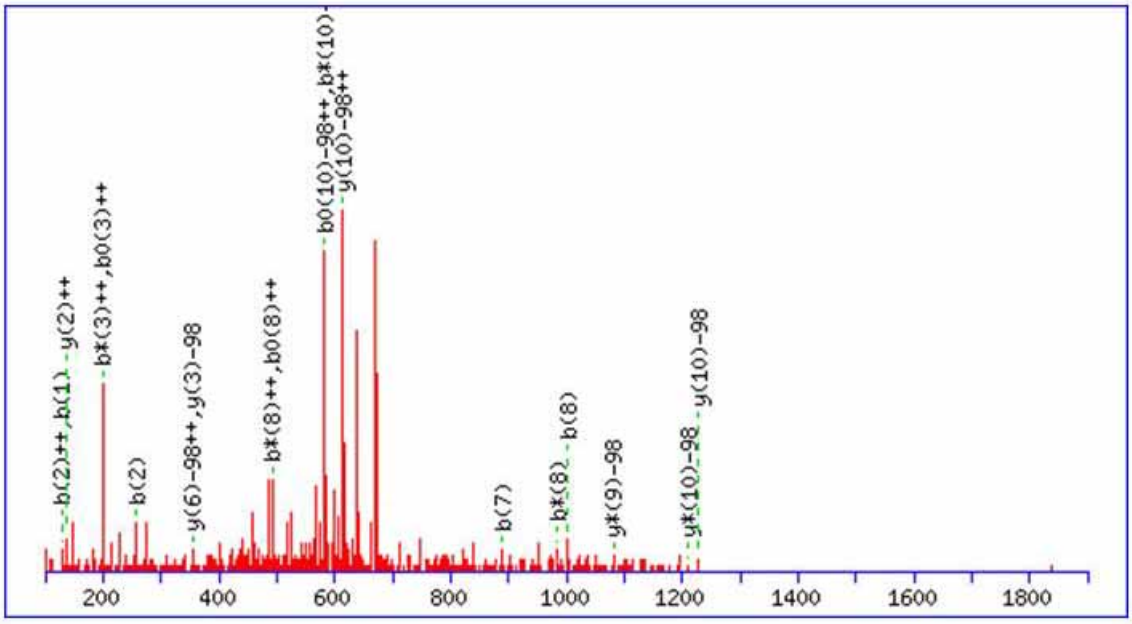

Figure 5-4. MASCOT search results of the phosphopeptide QERFSQILpTPR (m/z 1453.7791) from Nuf2 family protein (gi|15219846) identified in band 3.

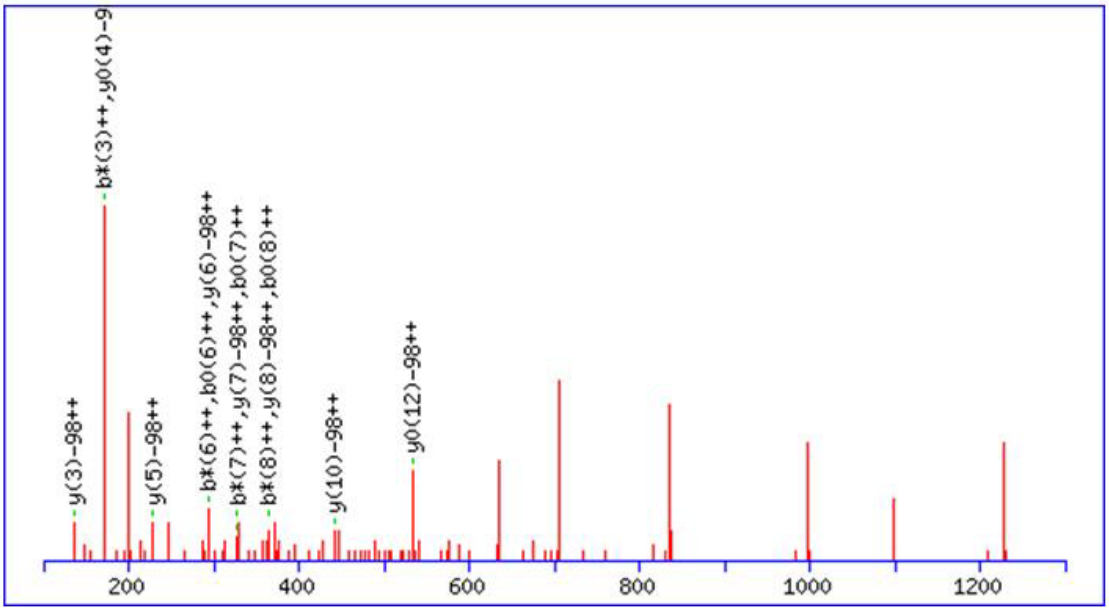

Figure 5-5. MASCOT search results of the phosphopeptide SRLSSAAAKPSVpTA (m/z 1424.7812) from ribosomal protein S6 (gi|2662469) identified in band 4.

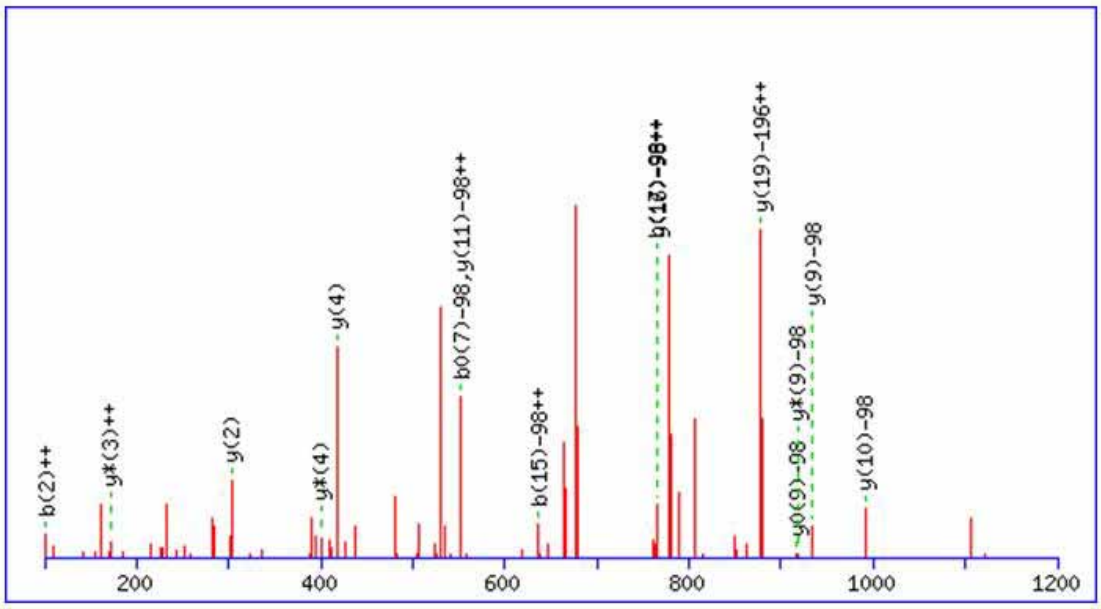

Figure 5-6. MASCOT search results of the phosphopeptide AMAVpSGAVLSGIGSSFLpTGGKR (m/z 2225.2202) from Lhcb6 protein (gi|4741960) identified in band 4.

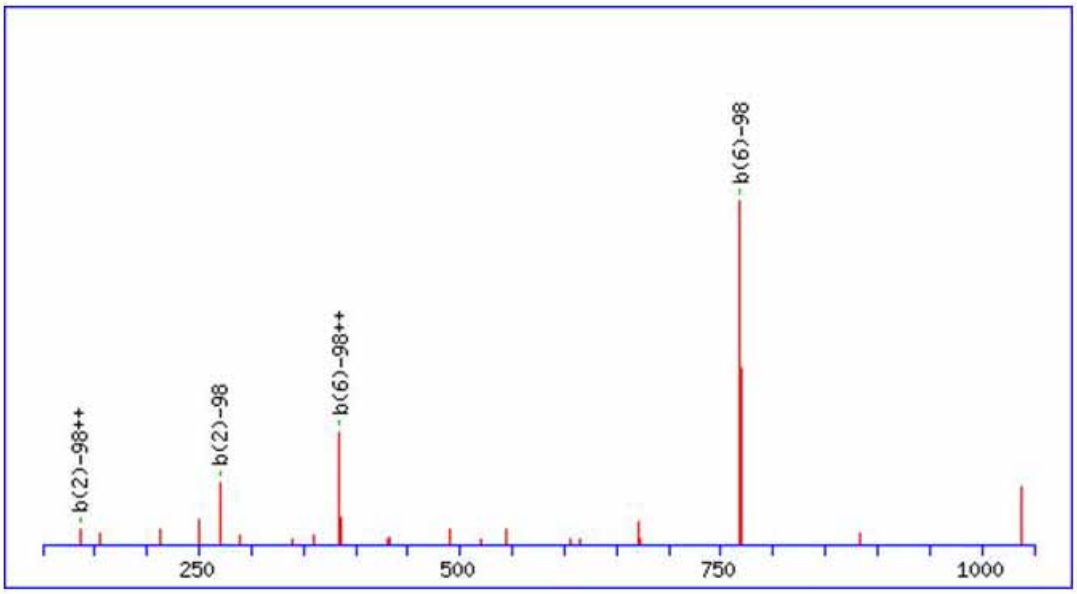

Figure 5-7. MASCOT search results of the phosphopeptide pTWEKLQMAAR (m/z 1312.7833) from laminin receptor homologue (gi|16380) identified in band 6.

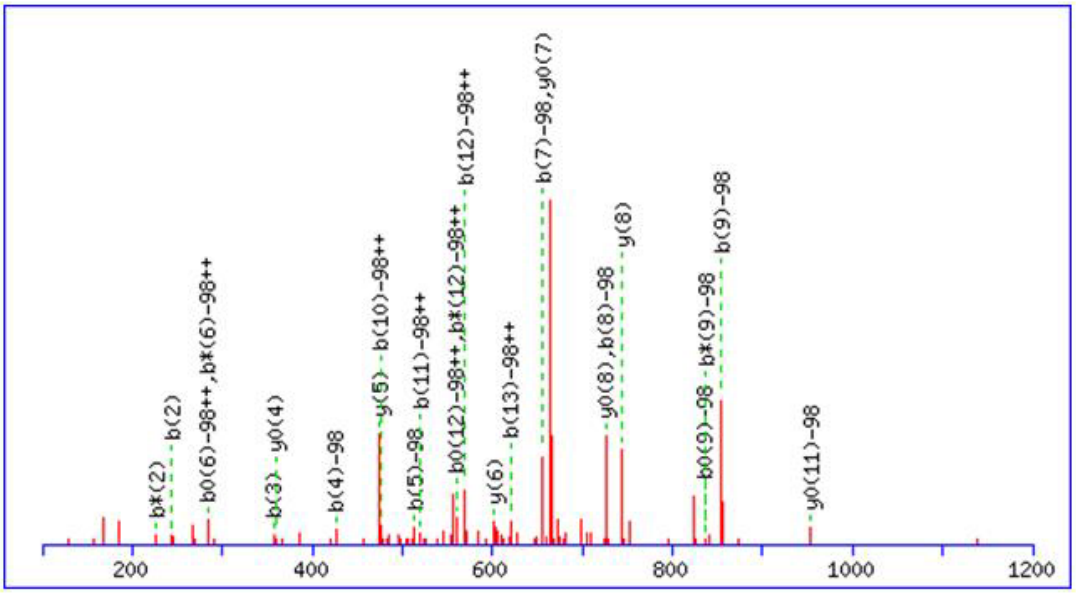

Figure 5-8. MASCOT search results of the phosphopeptide LEAIEpTAK (m/z 953.5758) from cysteine synthase (gi|1488519) identified in band 6.

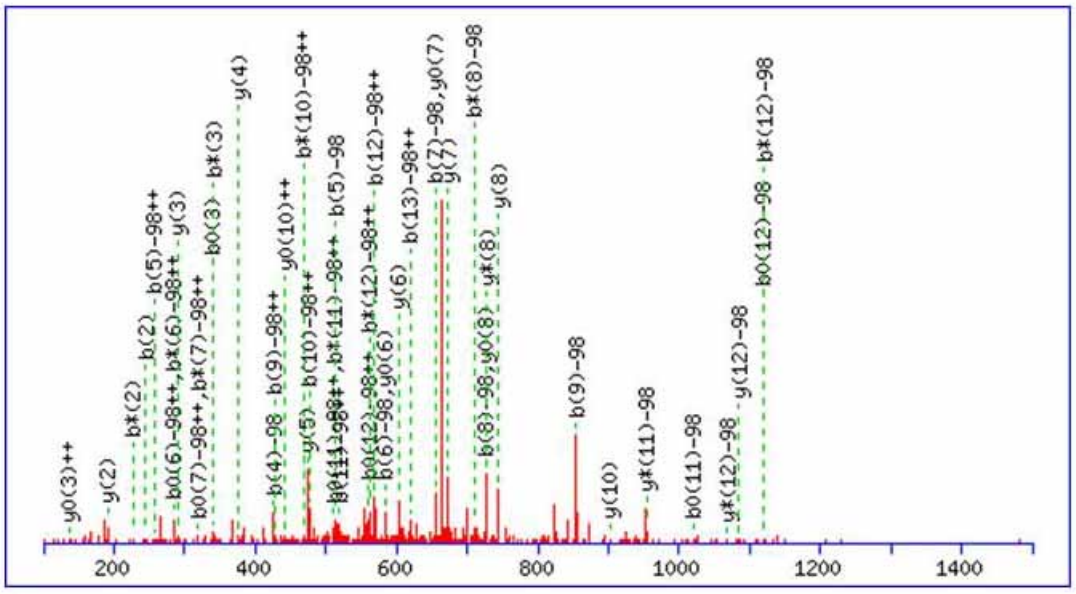

Figure 5-9. MASCOT search results of the phosphopeptide SRLpSSAAAKPSVTA (m/z 1424.7800) from ribosomal protein S6 (gi|2662469) identified in band 7.

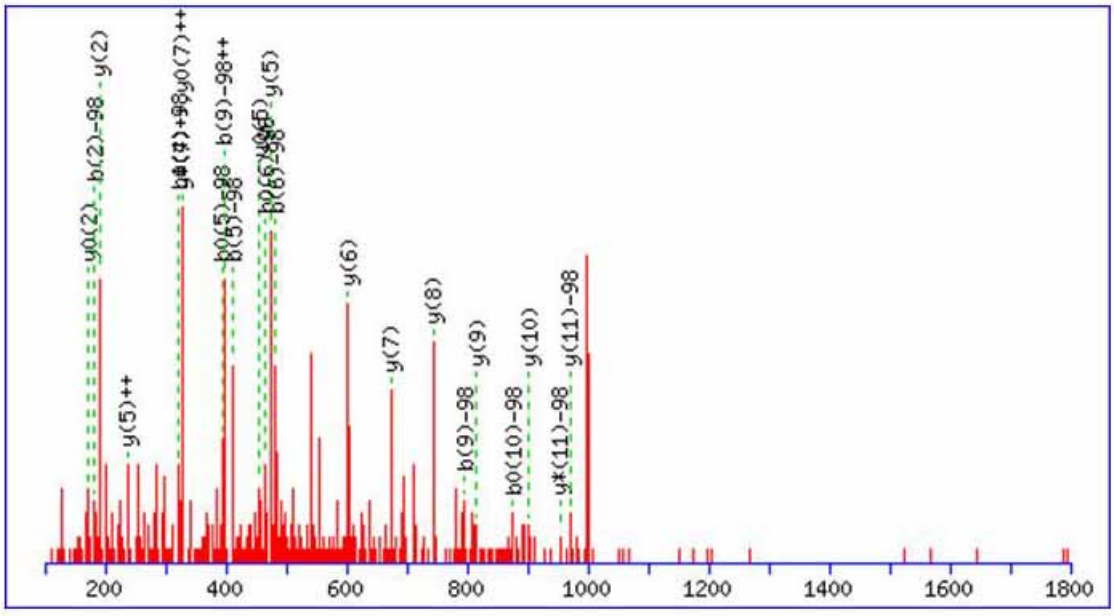

Figure 5-10. MASCOT search results of the phosphopeptide LpSSAAAKPSVTA (m/z 1181.6309) from ribosomal protein S6-like (gi|7270073) identified in band 8.

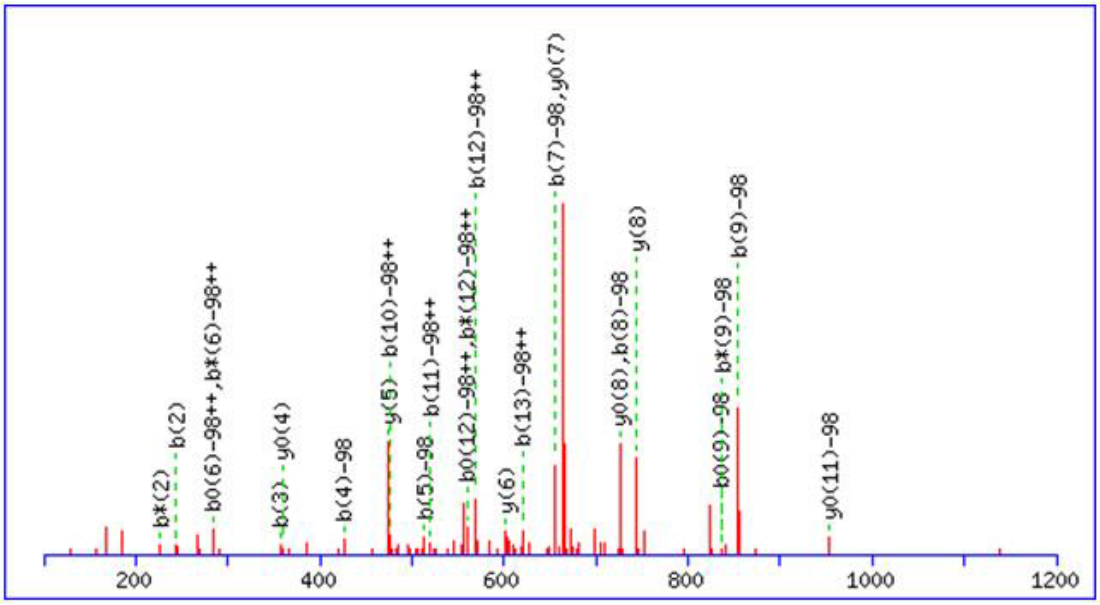

Figure 5-11. MASCOT search results of the phosphopeptide SRLpSSAAAKPSVTA (m/z 1424.7857) from ribosomal protein S6-like (gi|7270073) identified in band 8.

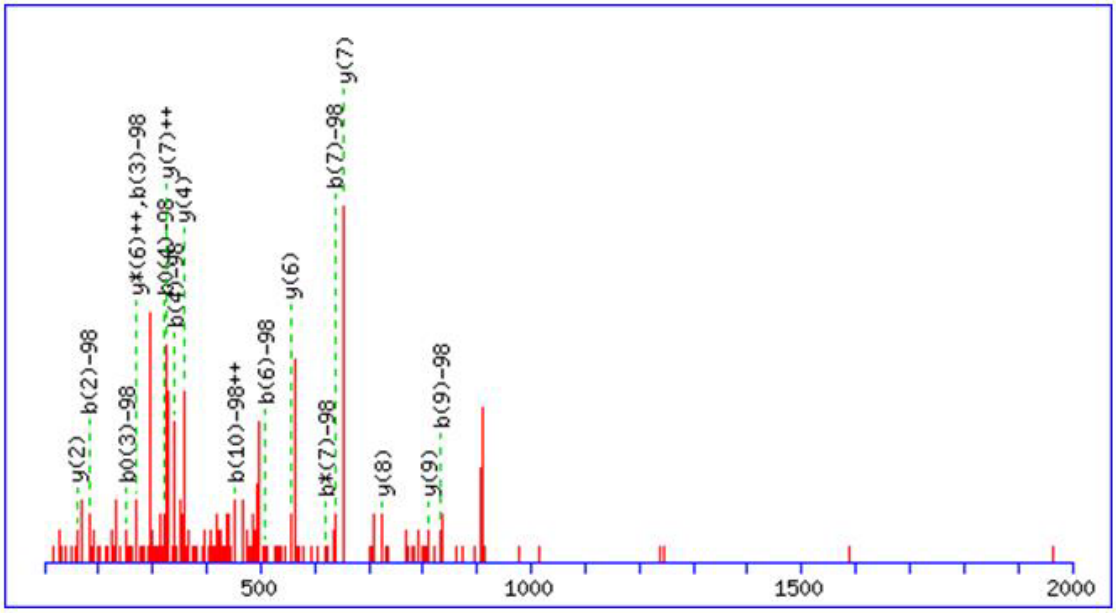

Figure 5-12. MASCOT search results of the phosphopeptide LpSSAPAKPVAA (m/z 1090.6041) from ribosomal protein S6 (gi|2224751) identified in band 8.

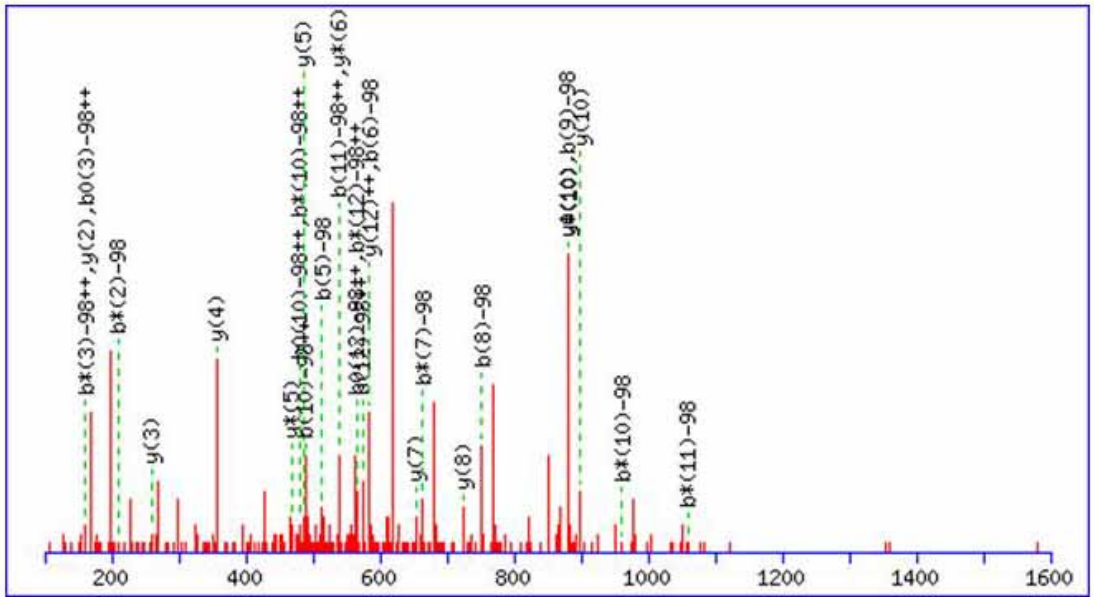

Figure 5-13. MASCOT search results of the phosphopeptide pSRLSSAPAKPVAA (m/z 1333.7495) from ribosomal protein S6 (gi|2224751) identified in band 8.

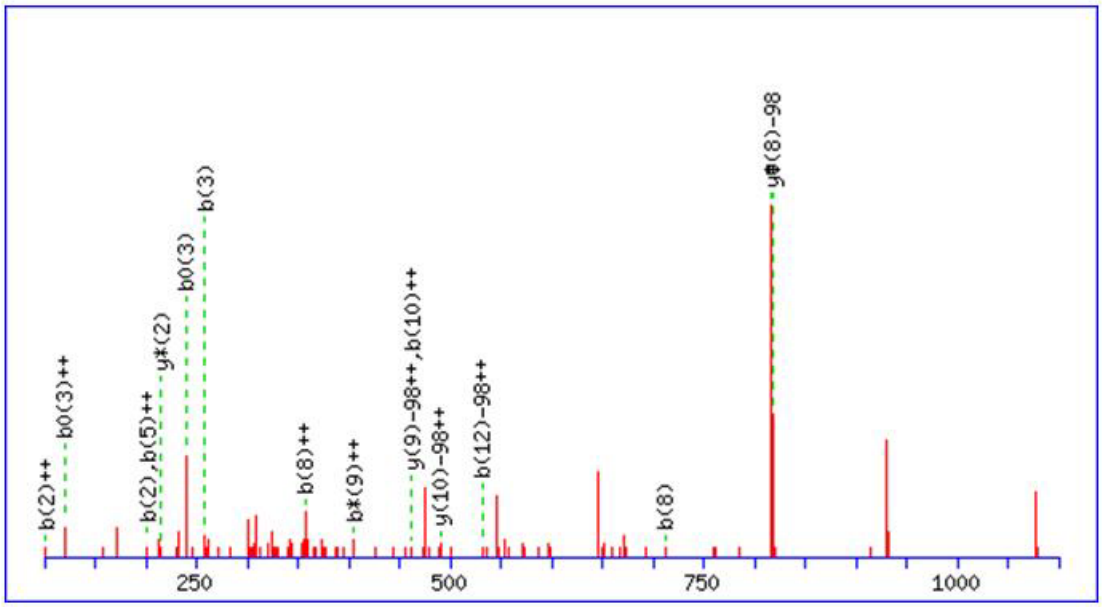

Figure 5-14. MASCOT search results of the phosphopeptide SLGGSRPGLPpTGR (m/z 1333.7944) from unknown (gi|21592536) identified in band 8.

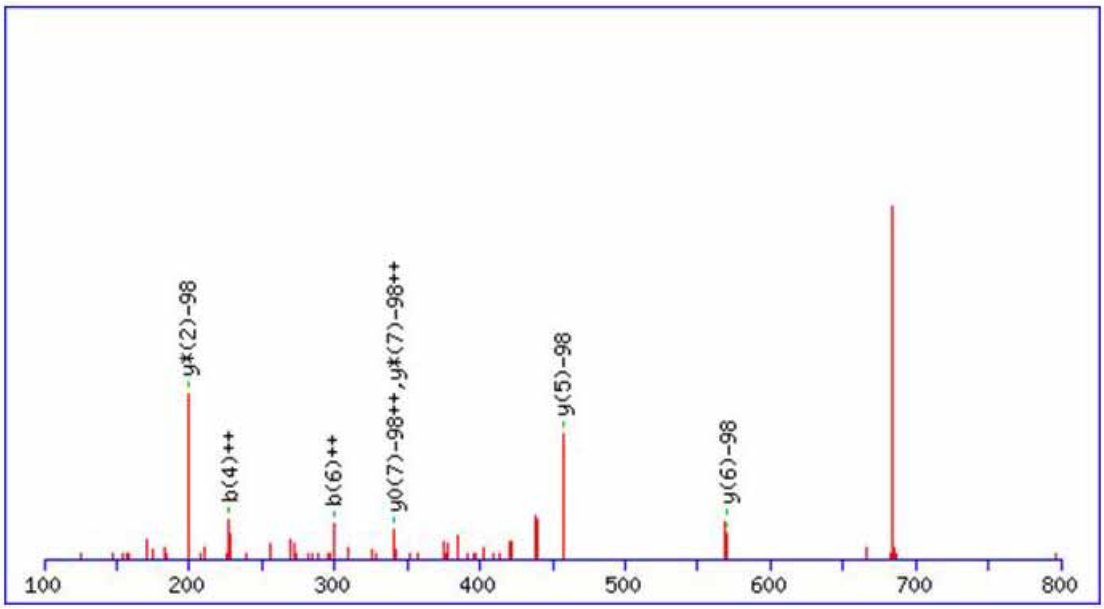

Figure 5-15. MASCOT search results of the phosphopeptide IKLPSGpSK (m/z 908.6211) from 60S ribosomal protein L2 (gi|22135870) identified in band 10.

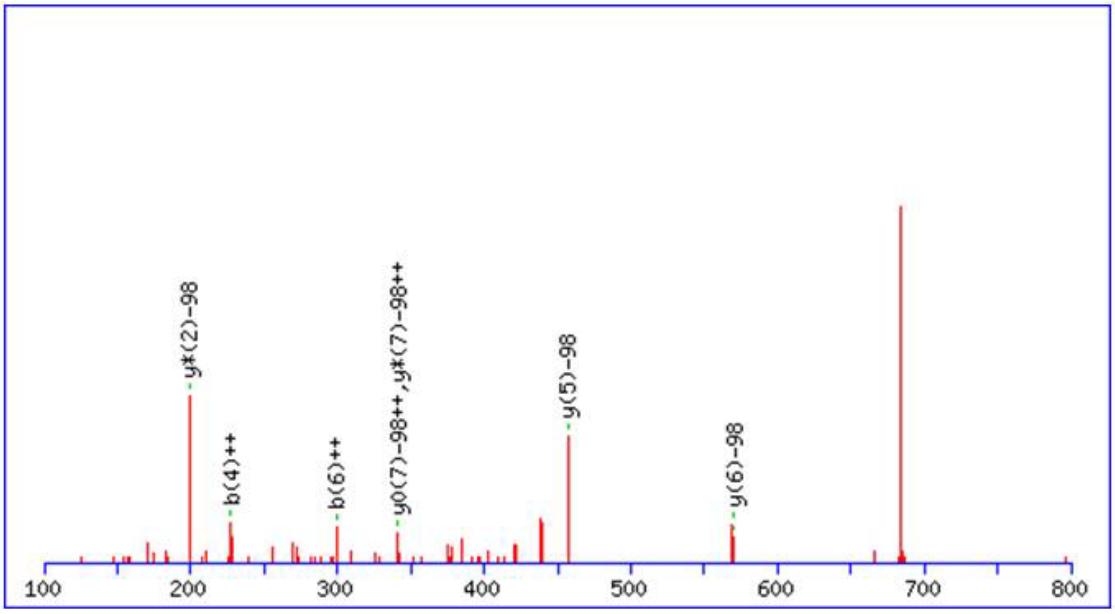

Figure 5-16. MASCOT search results of the phosphopeptide IKLPSGpSK (m/z 908.6211) from putative ribosomal protein L8 (gi|7270565) identified in band 10.

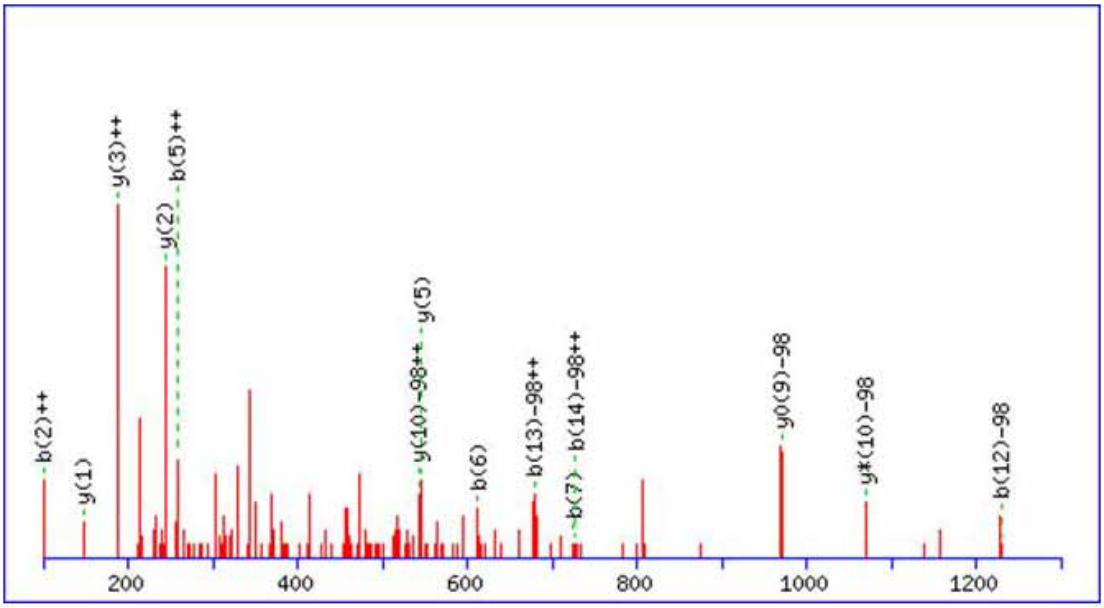

Figure 5-17. MASCOT search results of the phosphopeptide AESLNPLNFpSSSKPK (m/z 1697.9166) from ATP-dependent Clp protease proteolytic subunit ClpR4, putative (gi|21593086) identified in band 11.

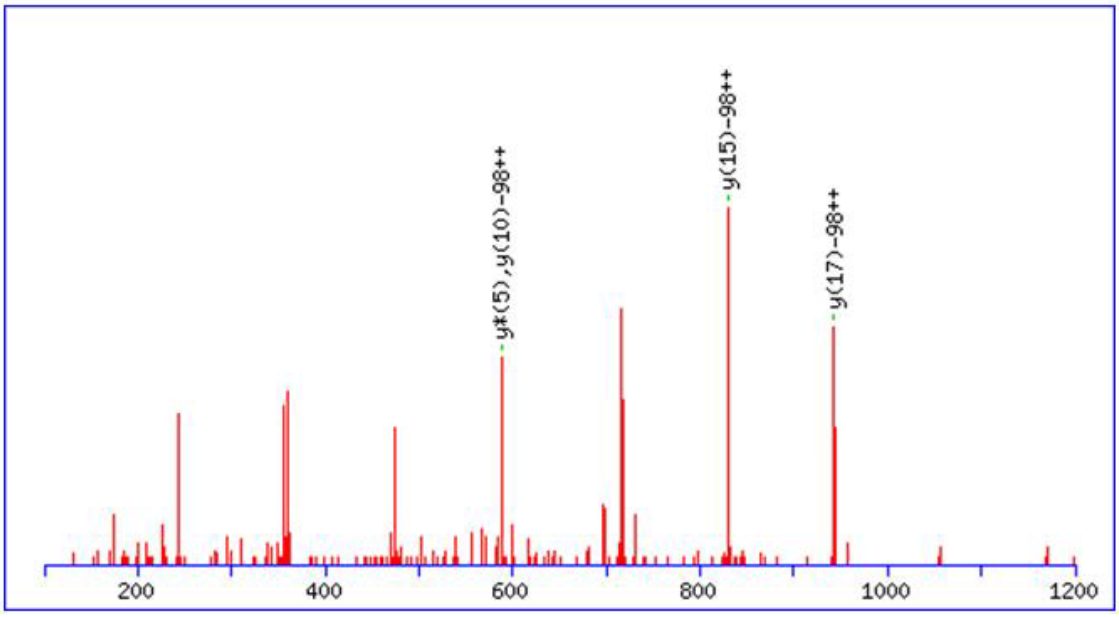

Figure 5-18. MASCOT search results of the phosphopeptide ALVTLIEKGVAFEpTIPVDLMK (m/z 2366.3912) from Glutathione S-transferase (gi|27363352) identified in band 12.

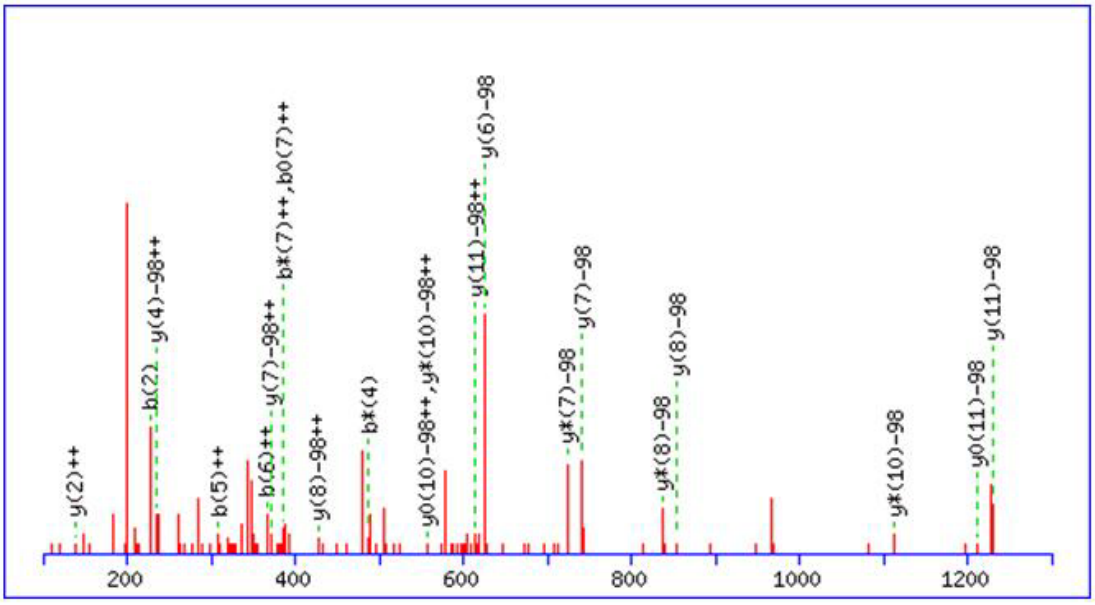

Figure 5-19. MASCOT search results of the phosphopeptide KVEMLDGVpTIVR (m/z 1454.8631) from putative ribosomal protein L9 (gi|12642868) identified in band 12.

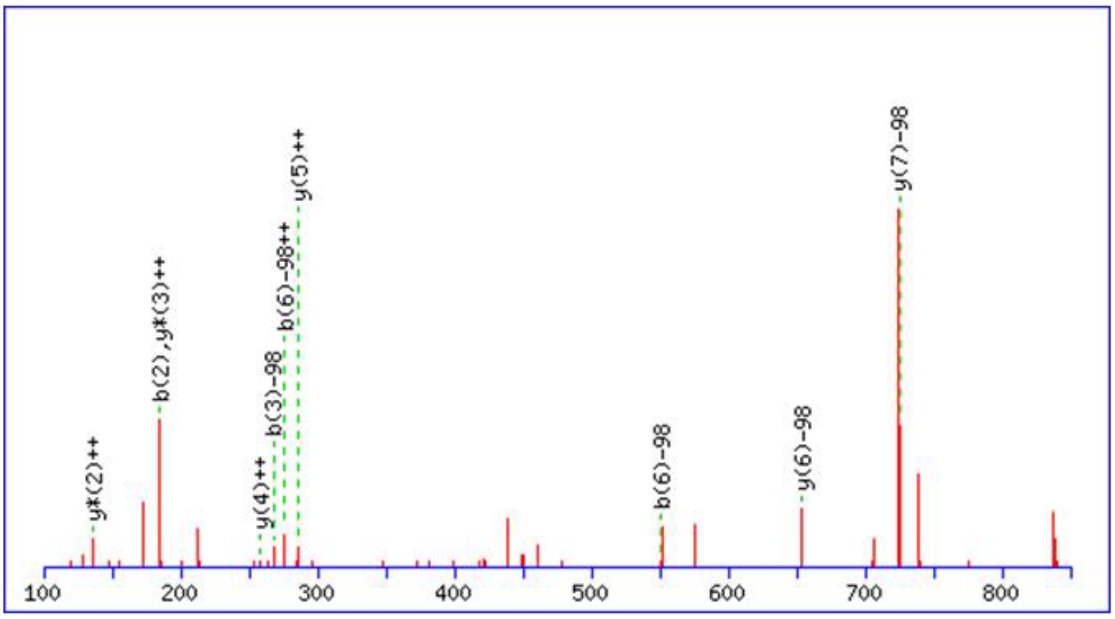

Figure 5-20. MASCOT search results of the phosphopeptide LApTGEPLR (m/z 935.5762) from putative protein (gi|7573368) identified in band 12.

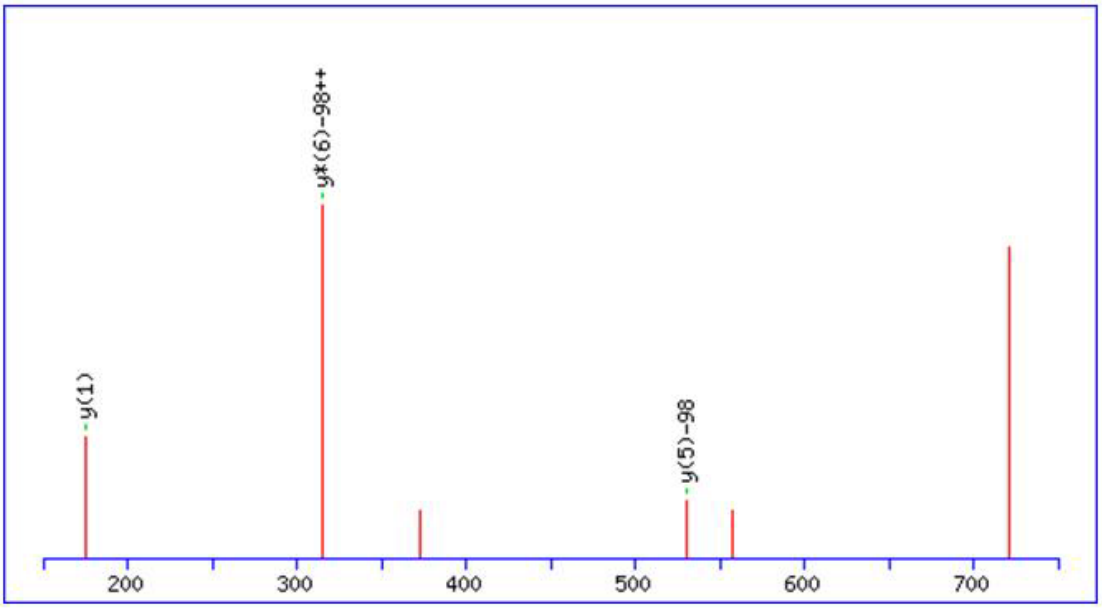

Figure 5-21. MASCOT search results of the phosphopeptide SFGLDSpSQAR (m/z 1146.6510) from putative protein 1 photosystem II oxygen-evolving complex (gi|4835233) identified in band 13.

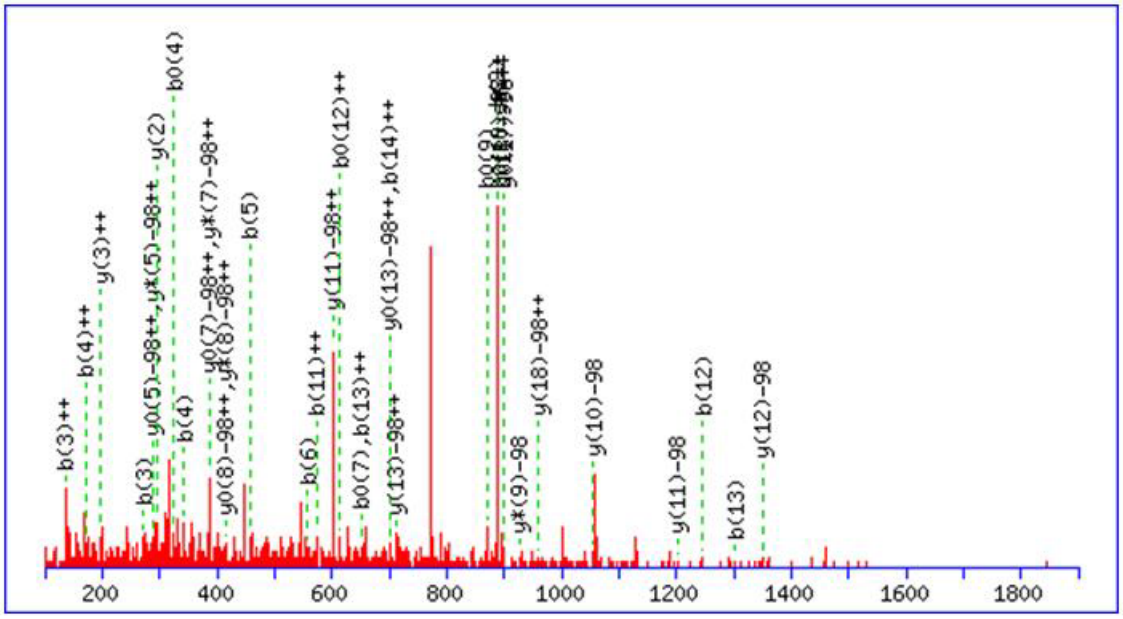

Figure 5-22. MASCOT search results of the phosphopeptide IGTADVLAFFLPGVVpSQVFK (m/z 2187.3049) from unknown protein (gi|3152582) identified in band 14.

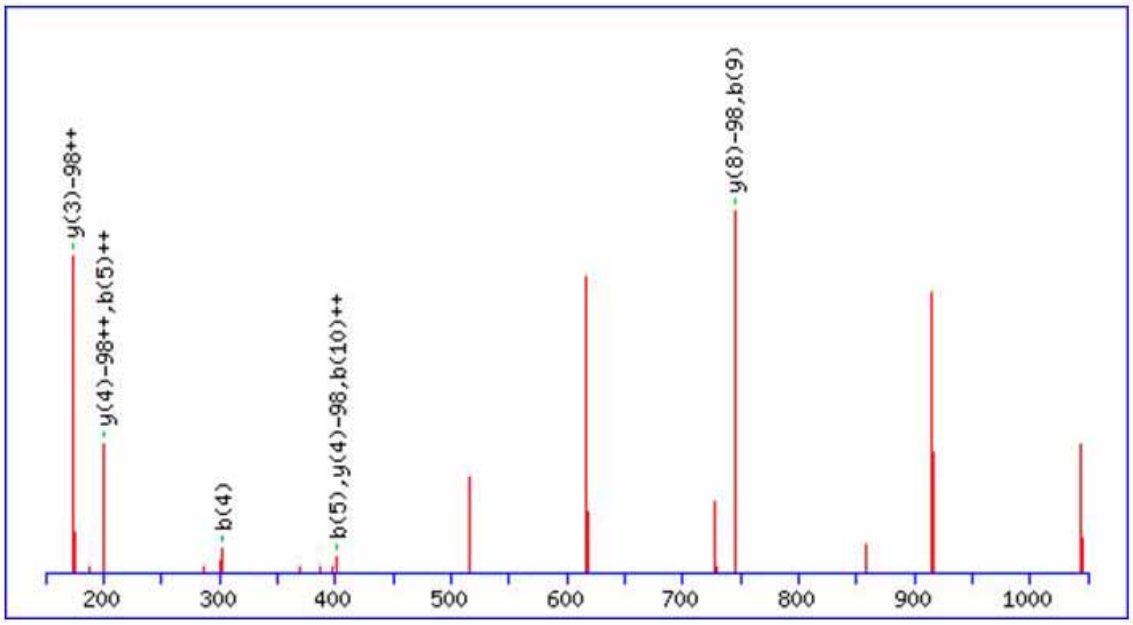

Figure 5-23. MASCOT search results of the phosphopeptide SAGSVGKSAGpSEK (m/z 1243.7179) from putative TNP1-like transposon protein (gi|4734013) identified in band 14.

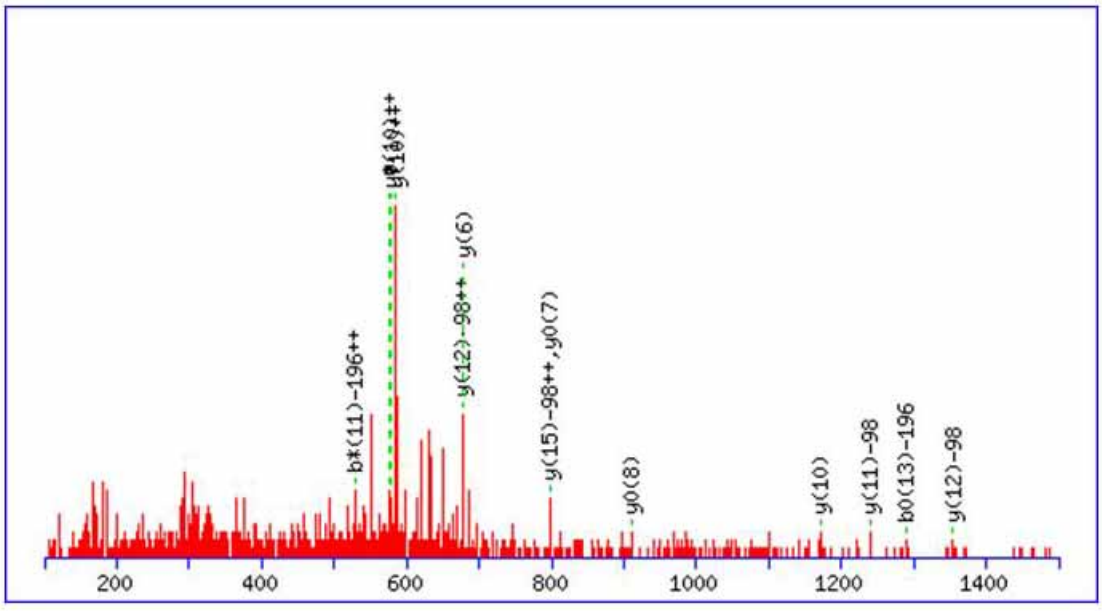

Figure 5-24. MASCOT search results of the phosphopeptide SpSGIALpSSRLHYASPIK (m/z 1946.0432) from peptidylprolyl isomerase ROC4 (gi|6899901) identified in band 15.

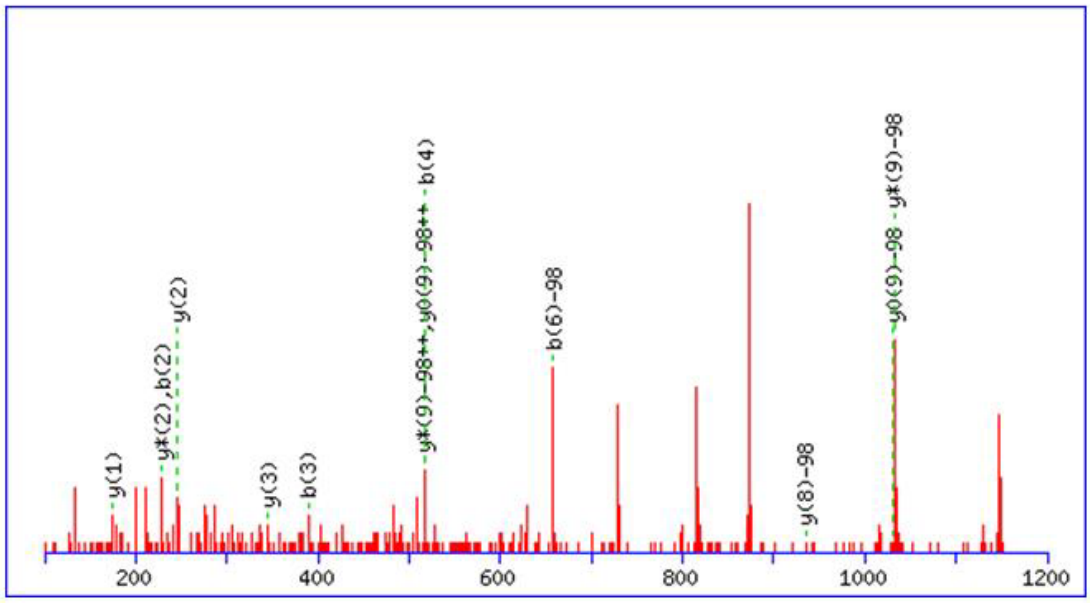

Figure 5-25. MASCOT search results of the phosphopeptide IDCEpSACVAR (m/z 1259.5482) from GAST1-like protein (gi|21618022) identified in band 19.

Table 5-1. Proteins identified from 14-3-3 affinity chromatography of *Arabidopsis thaliana* proteins.

Function or location	gi number	Provisional name
Protein synthesis, processing, trafficking	gi|4733972	14-3-3 protein (grf15), putative
	gi|21618266	14-3-3 protein GF14omega (grf2)
	gi|21553476	14-3-3-like protein GF14 iota (general regulatory factor 12)
	gi|2129595	14-3-3 protein homolog GF14 upsilon chain
	gi|166717	GF14 psi chain
	gi|1531629	GF14 mu
	gi|1256534	GF14 chi chain
	gi|12247993	putative 14-3-3 protein GF14epsilon
	gi|15222146	eukaryotic translation initiation factor, putative (EIF4B5)
	gi|295789	elongation factor 1-alpha
	gi|23397095	putative chloroplast translation elongation factor EF-Tu precursor
	gi|21592448	cytosolic cyclophilin ROC3
	gi|6899901	peptidylprolyl isomerase ROC4*
	gi|38454142	peptide methionine sulfoxide reductase-like protein
	gi|804950	cysteine synthase*
	gi|6850835	cysteine synthase AtcysC1
	gi|16305	glycine rich protein
	gi|16301	glycine rich protein
	gi|6899935	ubiquitin extension protein (UBQ5)
	gi|6572081	non-specific lipid transfer protein
	gi|26451353	putative non-specific lipid transfer protein nLTP
	gi|4056469	Strong similarity to gb|M95166 ADP-ribosylation factor
	gi|23198346	adenylate translocator
	gi|16160	adenosine nucleotide translocator
	gi|21280861	putative cytosolic factor protein
	gi|2058282	atranbp1a
Signaling/stress	gi|23197658	apospory-associated protein C-like protein
	gi|20148361	possible apospory-associated like protein
	gi|30794013	putative calcium-binding protein, calreticulin
	gi|1429207	annexin
	gi|7572922	copper homeostasis factor
	gi|7269278	hsp 70-like protein
	gi|7269269	HSP90-like protein
	gi|1906826	heat shock protein
	gi|8953699	4-nitrophenylphosphatase-like
	gi|34485583	extracellular calcium sensing receptor
	gi|7939542	AIG2 protein-like (avirulence induced gene protein)
	gi|21593920	signal recognition particle 19 kDa protein subunit, putative
	gi|21436247	putative pathogenesis-related PR-1 protein
	gi|8778701	Probable phosphopantothenoylcysteine decarboxylase
	gi|2765081	g5bf

*Identified as being phosphorylated by MASCOT search engine

Table 5-1. Continued.

Function or location	gi number	Provisional name
Metabolism	gi|497788	glutathione S-transferase
	gi|23197738	glutathione S-transferase
	gi|2266412	glutathione S-transferase
	gi|6730003	Chain A, glutathione S-transferase in complex with herbicide
	gi|27363352	glutathione S-transferase*
	gi|7263568	sedoheptulose-bisphosphatase precursor
	gi|166702	glyceraldehyde 3-phosphate dehydrogenase A subunit
	gi|6721173	glyceraldehyde-3-phosphate dehydrogenase C subunit (GapC)
	gi|336390	glyceraldehyde 3-phosphate dehydrogenase B subunit
	gi|19699140	putative glyceraldehyde-3-phosphate dehydrogenase
	gi|23198084	phosphoglycerate kinase, putative
	gi|21536853	phosphoglycerate kinase, putative
	gi|12644295	Phosphoglycerate kinase, chloroplast precursor
	gi|1022805	phosphoglycerate kinase
	gi|30102502	aminomethyltransferase-like precursor protein
	gi|21689621	p-nitrophenylphosphatase-like protein
	gi|21360517	myrosinase-associated protein like
	gi|11242	fructose-bisphosphatase
	gi|16226653	fructose-bisphosphate aldolase like protein
	gi|7529717	fructose bisphosphate aldolase-like protein
	gi|14334740	putative fructose bisphosphate aldolase
	gi|7267660	H+-transporting ATP synthase-like protein
	gi|5708095	ATP synthase gamma chain, chloroplast precursor
	gi|5881679	ATPase alpha subunit
	gi|6522554	peroxidase
	gi|405611	peroxidase
	gi|16173	L-ascorbate peroxidase
	gi|9279611	peroxiredoxin Q-like protein
	gi|7529720	peroxiredoxin-like protein
	gi|11994212	glycolate oxidase
	gi|22531128	glycolate oxidase
	gi|9755746	formate dehydrogenase (FDH)
	gi|8809648	ripening-related protein-like; contains similarity to pectinesterase
	gi|8778996	Contains similarity to ferredoxin-NADP+ reductase and contains an oxidoreductase FAD/NAD-binding PF|00175 domain
	gi|871992	thioglucosidase
	gi|871990	thioglucosidase
	gi|7671423	microbody NAD-dependent malate dehydrogenase
	gi|3929649	mitochondrial NAD-dependent malate dehydrogenase

*Identified as being phosphorylated by MASCOT search engine

Table 5-1. Continued.

Function or location	gi number	Provisional name
Metabolism	gi|7529712	monodehydroascorbate reductase (NADH)-like protein
	gi|21536569	NADH:ubiquinone oxidoreductase-like protein
	gi|7076787	cytosolic triosephosphate isomerase
	gi|21593477	putative triosephosphate isomerase
	gi|6735305	beta-1, 3-glucanase 2 (BG2)
	gi|4741197	aldose 1-epimerase-like protein
	gi|23197622	phosphoribulokinase precursor
	gi|22327649	glycosyl hydrolase family 1 protein
	gi|21593796	glutamate-ammonia ligase (EC 6.3.1.2) precursor, chloroplast
	gi|19171469	isocitrate dehydrogenase
	gi|11761812	glutathione dependent dehydroascorbate reductase precursor
	gi|887939	Gibberellin-regulated protein GAST1 protein homolog
	gi|21618022	Gibberellin-regulated protein GAST1-like protein*
Photosynthesis/chloroplast proteins	gi|295792	ribulose bisphosphate carboxylase
	gi|1944432	ribulosebisphosphate carboxylase
	gi|16194	ribulose bisphosphate carboxylase
	gi|23505787	ribulose bisphosphate carboxylase small chain 1b precursor (RuBisCO small subunit 1b)
	gi|13926229	ribulose bisphosphate carboxylase, small subunit protein
	gi|5881702	large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase
	gi|15450379	Rubisco activase
	gi|7267731	chlorophyll a/b-binding protein-like
	gi|6522530	chlorophyll a-b binding protein 4 precursor homolog
	gi|16374	chlorophyll a/b binding protein (LHCP AB 180)
	gi|13265501	chlorophyll a/b-binding protein
	gi|6403491	putative chlorophyll a/b-binding protein
	gi|7327812	chlorophyll a/b-binding protein CP29
	gi|12642854	putative photosystem II type I chlorophyll a/b binding protein
	gi|430947	PSI type III chlorophyll a/b-binding protein
	gi|23296426	putative photosystem II type I chlorophyll a/b binding protein
	gi|10176952	photosystem I reaction centre subunit psaN precursor
	gi|7269762	putative photosystem I reaction center subunit II precursor
	gi|7269730	photosystem I subunit PSI-E-like protein*
	gi|5732201	photosystem I subunit II precursor
	gi|5732205	photosystem I subunit IV precursor
	gi|12642864	putative photosystem I subunit V precursor
	gi|2924280	PSI 9kD protein

*Identified as being phosphorylated by MASCOT search engine

Table 5-1. Continued.

Function or location	gi number	Provisional name
Photosynthesis/chloroplast proteins	gi|9759370	photosystem II stability/assembly factor HCF136
	gi|7268925	photosystem II oxygen-evolving complex protein 3-like
	gi|1076373	photosystem II oxygen-evolving complex protein 2 – (fragment)
	gi|21592906	23 kDa polypeptide of oxygen-evolving comlex (OEC)
	gi|22571	33 kDa oxygen-evolving protein
	gi|10177538	33 kDa polypeptide of oxygen-evolving complex
	gi|4835233	putative protein 1 photosystem II oxygen-evolving complex*
	gi|6175179	putative ribose 5-phosphate isomerase
	gi|99698	glutamate-ammonia ligase (EC 6.3.1.2), cytosolic (clone lambdaAtgskb6)
	gi|6016709	carbonic anhydrase, chloroplast precursor
	gi|28973241	putative thioredoxin-m
	gi|4741960	Lhcb6 protein*
	gi|4741944	Lhcb2 protein
	gi|20465661	putative ferredoxin-NADP+ reductase
	gi|7270198	putative component of cytochrome B6-F complex
	gi|5881707	cytochrome f
	gi|20465723	putative cytochrome c oxidase subunit
	gi|7267158	putative fibrillin
	gi|20148241	putative fibrillin
Ribosomal	gi|2244857	ribosomal protein
	gi|166858	ribosomal protein
	gi|5881760	ribosomal protein L2
	gi|2791998	ribosomal protein L4
	gi|7267628	putative L5 ribosomal protein
	gi|21537351	putative ribosomal protein L7
	gi|7270565	putative ribosomal protein L8*
	gi|6562271	ribosomal protein L8 homolog
	gi|12642868	putative ribosomal protein L9*
	gi|468771	ribosomal protein L12
	gi|16497	Plastid ribosomal protein CL15
	gi|550544	ribosomal protein L16
	gi|12484209	putative ribosomal protein L18
	gi|33589762	ribosomal protein like (L19)
	gi|3482935	Putative ribosomal protein L21
	gi|2654122	ribosomal protein L23a
	gi|21554529	ribosomal protein L27, putative
	gi|19310757	putative ribosomal protein L28
	gi|10177580	ribosomal protein L32
	gi|23296539	putative ribosomal protein S1
	gi|6598334	putative ribosomal protein S4
	gi|2662469	ribosomal protein S6*
	gi|2224751	ribosomal protein S6*

*Identified as being phosphorylated by MASCOT search engine

Table 5-1. Continued.

Function or location	gi number	Provisional name
Ribosomal	gi|166867	ribosomal protein S11 (probable start codon at bp 67)
	gi|7270073	ribosomal protein S6-like*
	gi|7269981	ribosomal protein S11-like
	gi|7267097	putative ribosomal protein S13
	gi|5881753	ibosomal protein S15
	gi|17978783	S18.A ribosomal protein
	gi|21592469	30S ribosomal protein S20
	gi|6706420	40S ribosomal protein S2 homolog
	gi|7671467	40S ibosomal protein S6
	gi|21592577	40S ribosomal protein S8-like
	gi|6682246	putative 40S ribosomal protein S23
	gi|7572932	40S ribosomal protein S26 homolog
	gi|21436345	putative 50S ribosomal protein L27
	gi|21593299	chloroplast 50S ribosomal protein L31, putative
	gi|7529726	60S ribosomal protein-like
	gi|7413634	60S ribosomal protein-like
	gi|23397084	putative 60S ribosomal protein
	gi|22135870	60S ribosomal protein L2*
	gi|9795602	Putative 60S ribosomal protein L6
	gi|21592925	putative 60S ribosomal protein L6
	gi|23308199	60S ribosomal protein L7A
	gi|7362767	60S ribosomal protein L7A protein
	gi|10176857	60S ribosomal protein L13
	gi|6522565	60S ribosomal protein L13, BBC1 protein
	gi|21593767	60S ribosomal protein L13a
	gi|6642654	putative 60S ribosomal protein L13A
	gi|21594923	putative 60S ribosomal protein L17
	gi|30102898	60S ribosomal protein L18A
	gi|17104643	putative 60S ribosomal protein L35
	gi|21592412	putative 60S ribosomal protein L35
	gi|21593754	60S ribosomal protein L37a
	gi|23507773	60S ribosomal protein L38-like protein
Nuclear	gi|20466610	putative nucleosome assembly protein
	gi|99684	DNA-binding protein
	gi|7573443	mRNA binding protein precursor-like
	gi|12325357	RNA-binding protein, putative
	gi|16320	Histone H1-2
	gi|16314	histone H1-1
	gi|1617013	histone H2B like protein
	gi|4914322	putative transcription factor
	gi|19548045	putative SET protein, phospatase 2A inhibitor
	gi|18377868	fibrillin precursor-like protein
	gi|9663025	DIP2 protein
	gi|9663023	DIP1 protein
	gi|6730705	Putative phosphatase 2A inhibitor
	gi|15219846	Nuf2 family protein (Myosin-like protein)*
Protease	gi|1354272	aspartic proteinase

*Identified as being phosphorylated by MASCOT search engine

Table 5-1. Continued.

Function or location	gi number	Provisional name
Nuclear	gi|21593086	ATP-dependent Clp protease proteolytic subunit ClpR4, putative*
	gi|27754300	putative cysteine proteinase inhibitor
Protease	gi|23397070	putative cysteine proteinase AALP
Kinase	gi|7378615	protein kinase-like
	gi|7268618	receptor serine/threonine kinase-like protein
Membrane	gi|7268821	endomembrane-associated protein
	gi|9759532	outer membrane lipoprotein-like
	gi|1143394	V-type proton-ATPase
	gi|16380	laminin receptor homologue*
	gi|16323452	putative thaumatin protein
Putative proteins	gi|7635455	putative protein
	gi|7573368	putative protein*
	gi|7362762	putative protein
	gi|7287993	putative protein
	gi|7269995	putative protein
	gi|7269839	putative protein
	gi|7269521	putative protein
	gi|7269388	putative protein
	gi|7267543	putative protein
	gi|7019666	putative protein
	gi|6562320	putative protein
	gi|5541681	putative protein
	gi|4886277	putative protein
	gi|10045563	putative protein
Unnamed/ hypothetical/ unknown Proteins	gi|9758664	unnamed protein product
	gi|8809633	unnamed protein product
	gi|9755652	hypothetical protein
	gi|6017109	hypothetical protein
	gi|4185132	hypothetical protein
	gi|20198164	hypothetical protein
	gi|7485430	hypothetical protein At2g37660 [imported]
	gi|3513730	hypothetical protein
	gi|6566279	hypothetical protein RF12
	gi|9795608	Unknown protein
	gi|9755448	Unknown protein
	gi|7658343	unknown protein
	gi|6041839	unknown protein
	gi|28973215	unknown protein
	gi|28393989	unknown protein
	gi|22136800	unknown protein
	gi|21450872	unknown protein
	gi|21436055	unknown protein
	gi|15294226	unknown protein
	gi|14334418	unknown protein
	gi|3193285	unknown protein

*Identified as being phosphorylated by MASCOT search engine

Table 5-1. Continued.

Function or location	gi number	Provisional name
Unnamed/ hypothetical/ unknown Proteins	gi|3152582	unknown protein*
	gi|30023784	unknown protein
	gi|23505947	unknown protein
	gi|23505937	unknown protein
	gi|13265523	unknown protein
	gi|21592865	unknown
	gi|21592536	unknown*
	gi|21536786	unknown
	gi|4454459	expressed protein
Other	gi|5734756	Similar to SOUL Protein
	gi|4734013	putative TNP1-like transposon protein*
	gi|20259257	putative 33 kDa secretory protein
	gi|17682	Wilm's tumor suppressor homologue
	gi|15222811	leucine-rich repeat family protein

*Identified as being phosphorylated by MASCOT search engine

CHAPTER 6
HIGH-THROUGHPUT PHOSPHOPROTEOMICS OF ARABIDOPSIS THALIANA

Two-dimensional gel electrophoresis (2-DE) has been exploited over the past 20 or more years for the large-scale analysis of proteins. Several reviews have provided a detailed overview of this separation technique along with associated sample preparation.[82-84] Briefly, 2-DE involves the separation of proteins by displacement in two dimensions oriented at right angles to one another. In the first dimension, proteins are separated by isoelectric focusing (separation by charge, pI) and in the second dimension by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) (separation by molecular weight).

Although this technique was developed in the mid-1970s,[85-87] its real expansion as a useful technique had to wait for the development of microanalytical techniques that were able to identify proteins at the small amounts available in 2-DE. Initially, Edman sequencing was the microanalytical technique used for these identifications, however, mass spectrometry has recently become the method of choice due to its higher sensitivity as compared with Edman sequencing and its capability to characterize almost any post-translational modification.[84] This combination of 2-DE with mass spectrometry resulted in an increase in popularity of this approach in the early 1990s, coining the word "proteomics," the large-scale study of proteins.

Despite its popularity, 2-DE is associated with numerous technical difficulties and inadequacies. First, the reproducibility of the same sample from gel to gel may vary. Second, although the resolution of 2-DE seems impressive, it is still not sufficient for the

enormous amounts of proteins present in the sample which can result in comigration of several proteins in the same spot. Third, due to the enormous chemical diversity of proteins and their very divergent expression in cells and tissues, protein samples may consist mainly of very abundant proteins, resulting in poor representation of low abundance proteins. Fourth, recovery of hydrophobic proteins and large molecular weight proteins tend to be difficult due to solubility problems. Finally, the 2-DE technique requires a degree of technical proficiency and high quality protein sample preparation is a necessity due to dramatic interference in separation if contaminants are present. 2-DE is also notoriously difficult to automate resulting in limited throughput.

Regardless of these problems, 2-DE has proven to be a very useful tool with regards to separating complex protein mixtures for comparative studies as well as for the identification of post-translational modifications such as phosphorylation. The basis of recognition of modified proteins in 2-DE is simple. If a protein becomes phosphorylated, one of the separation parameters (pI or MW) must be altered in the modified form. Traditionally, identification of a phosphorylation modification on a protein was determined by a shift in a protein's pI. However, the recent development of the sensitive fluorescent Pro-Q® Diamond Phosphoprotein Gel Stain has provided a means of selectively staining phosphorylated proteins in polyacrylamide gels. The detected proteins may then be excised, digested with trypsin, and analyzed by mass spectrometry for protein identification and phosphorylation site mapping.

As in any approach, preparation of large quantities of samples can be a major bottleneck. Alleviating this problem has become possible with the development of robotics for post-electrophoretic preparations prior to mass spectrometric analysis.

Included in these preparations are gel spot excision and digestion. Among the various robotic instrumentation available for spot picking and digestion are the Investigator™ ProPic™ and the Investigator™ ProGest™.

The Investigator™ ProPic™ comes integrated with three key processes: high-resolution gel imaging, choices in image analysis with HT PC Analyzer Software and HT Analyzer 2-D Evolution, and protein spot cutting. Gel imaging is executed in a fully light tight enclosure with an interchangeable UV light source for imaging of fluorescent stained gels via CCD-based imaging technology. Once imaged, the HT Analyzer Software is then used for selecting spots for excision by generating a picklist automatically or by manual point and click methods. Gel plugs of 1.8 mm are then excised by using gel hydration, cutting and gentle vacuum extraction, and transferred to 96-well plates. Utilizing the ProPic's fully enclosed processing environment minimizes sample handling and the risk of keratin contamination. Also, with the automation of the instrument during picking and transferring of the spots, the user's time is freed to attend to other tasks.

Once the spot picking process is completed, the 96-well plate can then be directly transferred to the Investigator™ ProGest™ for in-gel digestion of all samples simultaneously. The Investigator™ ProGest™ comes equipped with automated protocols and temperature-controlled reactions that save time and increase reproducibility. Customization of the protocols is also available. The automation of the tedious manual procedures such as wash steps significantly decreases the introduction of human errors, increasing reproducibility. Additionally, sample processing is performed by using a concentric dual needle design that delivers liquid reagents and pressurized nitrogen to

each well. Waste removal and sample recovery without aspiration steps is possible due to the use of pierced reaction plates. This omission of aspiration steps during sample processing eliminates the risk of sample cross-contamination or gel plug carry-over. Overall, the use of the ProGest system with its fully enclosed processing environment minimizes sample handling, the risk of keratin contamination, cross-contamination among samples, and increases throughput.

The results presented here demonstrate the utility of the phosphoprotein stain with the automation of robotics as a means for high-throughput phosphoproteomic analysis of *Arabidopsis thaliana*.

Experimental Methods

Materials and Instruments

TRIzol® Reagent was obtained from Invitrogen (Carlsbad, CA). The 2-D Cleanup Kit and PlusOne 2-D Quant Kit and immobilized linear gradient strips (pH 3-11) were purchased from GE Healthcare (Piscataway, NJ). 8-16% Tris-glycine polyacrylamide gels were from Bio-Rad (Hercules, CA). Equilibration Buffers I and II were purchased from Genomic Solutions (Ann Arbor, MI). Pro-Q® Diamond Phosphoprotein Gel Stain, PeppermintStick™ Phosphoprotein Molecular Weight Markers, and Colloidal Blue Gel Stain were from Invitrogen. Sequencing-grade modified trypsin was purchased from Promega (Madison, WI). All other solvents were obtained from Fisher Scientific (Fairlawn, NJ).

Isoelectric focusing was performed on an IPGphor system (GE Healthcare, Piscataway NJ). SDS-PAGE was performed on a Protean 2 (BioRad, Hercules, CA). Fluorescent imaging of gels was acquired with a Typhoon 8600 variable mode scanner (GE Healthcare, Piscataway, NJ). Automated spot excision and digestion were carried out

with the Investigator™ ProPic™ and the Investigator™ ProGest™ from Genomic Solutions. A FAMOS autosampler from LC Packings (Sunnyvale, CA) was used for automated sample loading. Capillary rpHPLC separation of protein digests was performed on a 15 cm x 75 um i.d. PepMap C18 column from LC Packings (San Francisco, CA) in combination with an Ultimate Capillary HPLC System (LC Packings, San Francisco, CA). Inline mass spectrometric analysis was accomplished by a hybrid quadrupole time-of-flight instrument (QSTAR, Applied Biosystems, Foster City, CA) equipped with a nanoelectrospray source.

Sample Preparation

Proteins were extracted from mature *Arabidopsis thaliana* leaves with TRIZOL® Reagent by grinding the leaves with a pre-chilled mortar and pestle and liquid nitrogen. Ground leaves were then transferred to a chilled Corex centrifuge tube and 5 mL of TRIZOL® Reagent was added for every 500 mg plant leaves. Ground leaves were allowed to sit in the TRIZOL® Reagent for 5 minutes at room temperature after which the sample was centrifuged at 10,900 x*g* for 5 minutes at 4^0C. The supernatant was then transferred to a fresh Corex tube, centrifuged for another 5 minutes, and the resulting supernatant transferred to an Oakridge tube. One mL of chloroform was then added and the tube shaken vigorously by hand for 15 seconds then allowed to stand at room temperature for 2 minutes. The sample was centrifuged at 10,900 x*g* for 15 minutes and the upper (aqueous) layer completely removed. 1.5 mL of absolute ethanol was then added to the lower phase (phenol-chloroform phase) and the sample mixed by inversion followed by incubation for 3 minutes at room temperature. Centrifugation of the sample at 483 x*g* was followed by the transfer of the supernatant to a fresh Corex tube. Three volumes of ice cold acetone was added to the supernatant followed by centrifugation for

2 minutes at 2,860 x*g*. The supernatant was then decanted, the acetone wash was repeated, and the pellet was air dried. The protein pellet was then solubilized with an 8M UTCEB rehydration buffer containing 8M urea, 2M thiourea, 4% CHAPS, 0.2% SDS, 100 mM DTT, and 0.5% pH 3-11 IPG buffer. Protein quantitation was then performed with the PlusOne 2-D Quant Kit according to the manufacturer's protocol.

2-Dimensional Gel Electrophoresis

For IEF in the first dimension, 400 uL (1.432 mg) of the protein sample was combined with 1.2 µL pH 3-11 IPG buffer and 2µL Orange G dye. The solution was then applied on an immobilized linear gradient strip (18 cm, pH 3-11) for rehydration overnight. IEF was then performed using the following three steps: 500 V for 6 hours, 1000 V for 8 hours, and 8000 V for 5 hours. The strip was then removed for equilibration (reduction and alkylation) prior to separation in the second dimension. First, incubation of the strip with 15 mL Equilibration Buffer I (6M urea, 130 mM DTT, 30% glycerol, 45 mM Tris base, 1.6% SDS, 0.002% Bromophenol Blue, pH 7) for 30 minutes was executed for reduction of the proteins. The buffer was then decanted and 15 mL Equilibration Buffer II (6M urea, 135 mM iodoacetamide, 30% glycerol, 45 mM Tris base, 1.6% SDS, 1.6% Bromophenol Blue, pH 7) was then added for another 30 minutes of incubation for alkylation of the proteins. The second dimension was then performed by running on an 18x18 cm 8-13% Tris-Glycine-SDS-PAGE gel. The gel was run at 10 mA for 1 hour followed by 24 mA for 5 hours at 10 ^{0}C.

Protein Visualization and Analysis

Pro-Q staining was performed according to the manufacturer's protocol for visualization of phosphorylated proteins. The fluorescent stained gel was then imaged with the Typhoon scanner at 532/560 nm excitation/emission. Following this, the gel was

then counter-stained with Colloidal Blue Gel Stain for total protein staining and the image captured with an office scanner. Both the Pro-Q and Colloidal images were then superimposed for identification of the protein spots for analysis, that is, protein spots from the Pro-Q image were identified in the Colloidal image for analysis.

Automated Spot Picking and Digestion

The Colloidal stained gel was then imaged with the Investigator™ ProPic™ and the HT Analyzer Software was then used for selecting spots for excision by generating a picklist by manual point and click methods. Gel plugs of 1.8 mm were then excised and transferred to 96-well plates for digestion with the Investigator™ ProGest™ according to one of the automated protocols. The digested samples were then collected and placed in the FAMOS autosampler for mass spectrometric analysis.

Amino Acid Sequencing by nanoESI QqTOF MS Analysis

MS/MS experiments for peptide sequencing were performed by loading 10 μL of each sample at 10 μL/min for 5 minutes with a FAMOS autosampler onto the C18 precolumn for desalting and concentration of the sample. The switching-valve position was then changed and the trapped peptides were back flushed and separated on the C18 nano column operated at a flow rate of 200 nL/min with a gradient of 5% to 60% acetonitrile over 30 minutes. MS/MS data were acquired by Information Dependent Acquisition (IDA) mode.

Protein Identification

Fragment ion data generated by Information Dependent Acquisition (IDA) via the QSTAR were searched against the NCBI nr sequence database using the Mascot (Matrix Science, Boston, MA) database search engine. Probability-based MOWSE scores above

the default significant value were considered for protein identification in addition to validation by manual interpretation of the MS/MS data.

Results and Discussion

Protein Visualization and Analysis

Pro-Q staining of the 2-D gel revealed at least 31 very intense protein spots (Figure 6-1) and a large number of weakly stained spots after background subtraction of unphosphorylated proteins in the standard markers that were co-electrophoresced on the gel. Following this, the gel was counter-stained with Colloidal Protein Stain (Figure 6-2) for easy visualization of protein spots, however, several spots were very faintly stained. Some spots that were intensely stained by the Pro-Q were faintly stained by the Colloidal Blue, suggesting that these are highly phosphorylated proteins. Protein spots were then chosen for analysis by superimposing the Pro-Q and Colloidal images of the gel.

Automated Spot Picking and Digestion

The gel was then imaged with the spot picker and a pick list was created for excision of protein spots (Figure 6-3). Operation of the Investigator^{TM} ProPic^{TM} resulted in the excision of only 18 spots. The remaining 13 spots were cut by the robot but not removed from the gel, therefore, manual excision was performed for these spots. Automated tryptic digestion by the Investigator^{TM} ProGest^{TM} resulted in the digestion of all samples, however, the final sample volumes varied, 2 of the 31 samples were placed in the wrong wells, and 3 samples had some of the sample placed on top of the plate beside the well instead of in the well.

Protein Identification

LC/MS/MS analysis of the tryptic digests of the individual spots resulted in the identification of 2 or more proteins per spot due to co-migration of proteins on the gel.

Causes of co-migration are overloading of the gel and inadequate separation. Several solutions to the latter problem have become available, such as fractionation of the sample prior to 2-DE or using zoom gels which separate the proteins over smaller pI ranges. It should also be noted that many proteins were found in multiple spots. In 2-D analysis, multiple spots are presumably due to post-translational modifications, degradation of proteins *in-vivo* or *in vitro*, or expression of differential isoforms derived from different genes.[88]

Overall, LC/MS/MS analysis of the tryptic peptides obtained from the excised gel spots revealed the identities of at least 138 proteins that were identified by the MASCOT search engine with at least 2 peptides with significant scores (Table 6-1). Identified proteins were grouped according to their biological function or location. These 138 proteins can be considered as putative phosphoproteins of *Arabidopsis thaliana*, however, verification by phosphorylation site mapping proved to be difficult for several possible reasons. First, identification of the site could have been missed by database searching due to insufficient fragmentation during sequencing. Second, proteins at the acidic end of the gel were not separated very well resulting in the identification of several proteins in one gel spot, which could result in insufficient amounts of the phosphorylated protein for phosphorylation site mapping. Third, since spots of interest were manually selected by superimposing the Pro-Q stained image and Colloidal image, it is possible that the excised spot was not on the correct target since gel plugs of only 1.8 mm were excised. Also, since the Pro-Q stain is more sensitive than the Colloidal stain, several spots were of very low intensity making spot picking somewhat difficult. This problem could be alleviated if spots could be excised directly from the Pro-Q imaged gel, however, this

capability was not available. Figures 6-4 through 6-12 show MS/MS spectra of possible phosphopeptides identified by the MASCOT search engine.

Of these 138 proteins identified, it is expected that only some of these proteins are truly phosphorylated since some spots were found to contain several proteins, that is, only one of the several proteins could be phosphorylated. However, since phosphorylation site mapping did not give the identification of the phosphorylation sites for most of the identified proteins it is difficult to say which of these 138 proteins are being detected by the Pro-Q stain. Despite the fact that the phosphorylation site was not identified it is still strongly believed that many of these proteins are indeed phosphorylated due to the fact that they were detected by the Pro-Q stain and also that many of the proteins were found in several spots running side by side which is a key indicator of a protein that has been post-translationally modified.

Overall, the results show that with the use of robotics for spot excision and proteolytic digestion in combination with selective staining of phosphorylated proteins, phosphoproteome analysis can become high-throughput, however, improvements need to be made with the robotic instrumentation to prevent loss of samples. Also, the development of a filter for imaging Pro-Q stained gels with the spot picker for direct excision would also increase throughput and decrease possible picking errors.

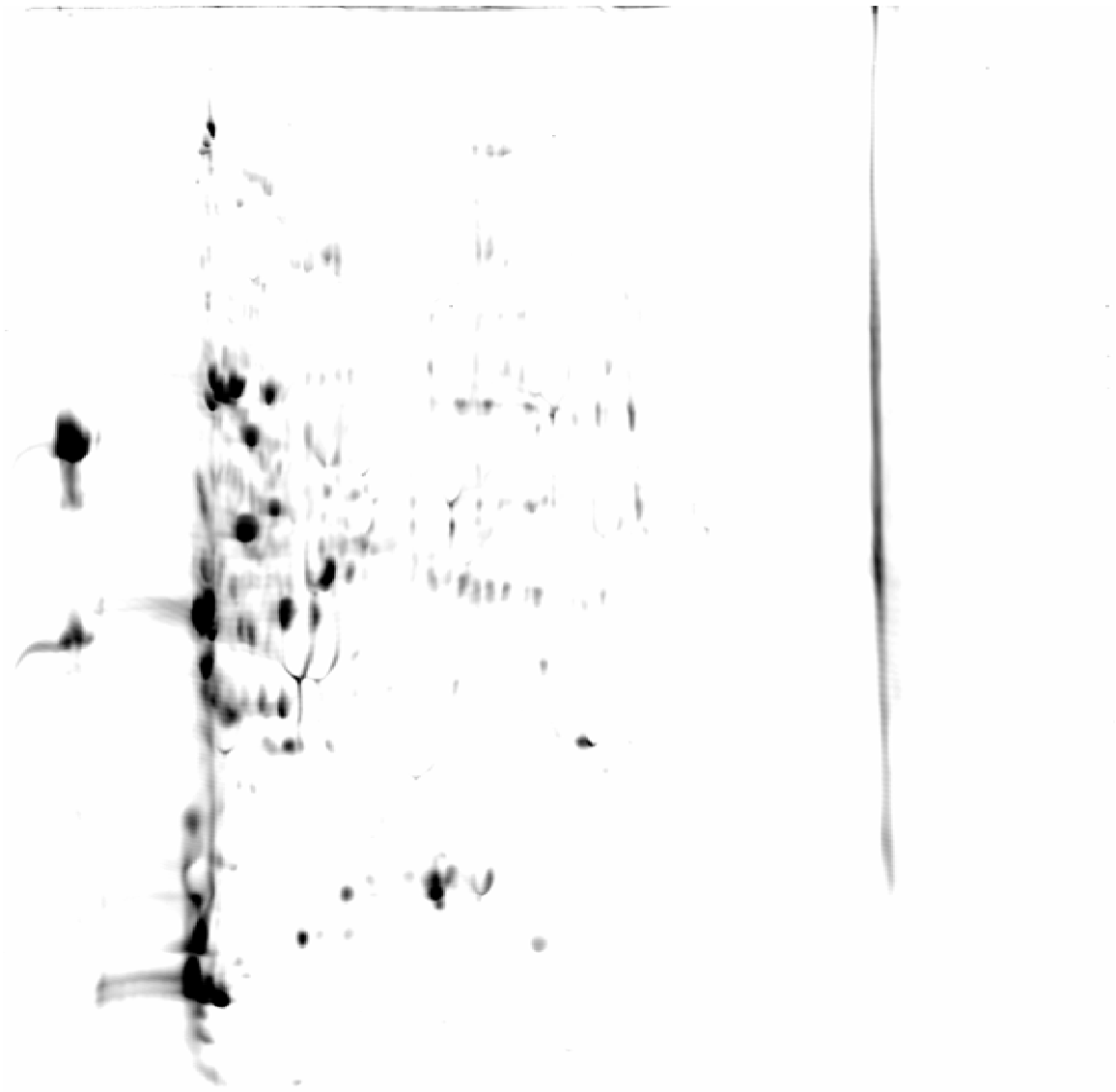

Figure 6-1. Pro-Q® Diamond Phosphoprotein Gel Stain image indicating potential phosphorylated proteins from a two-dimensional gel electrophoresis separation of *Arabidopsis thaliana* protein extract.

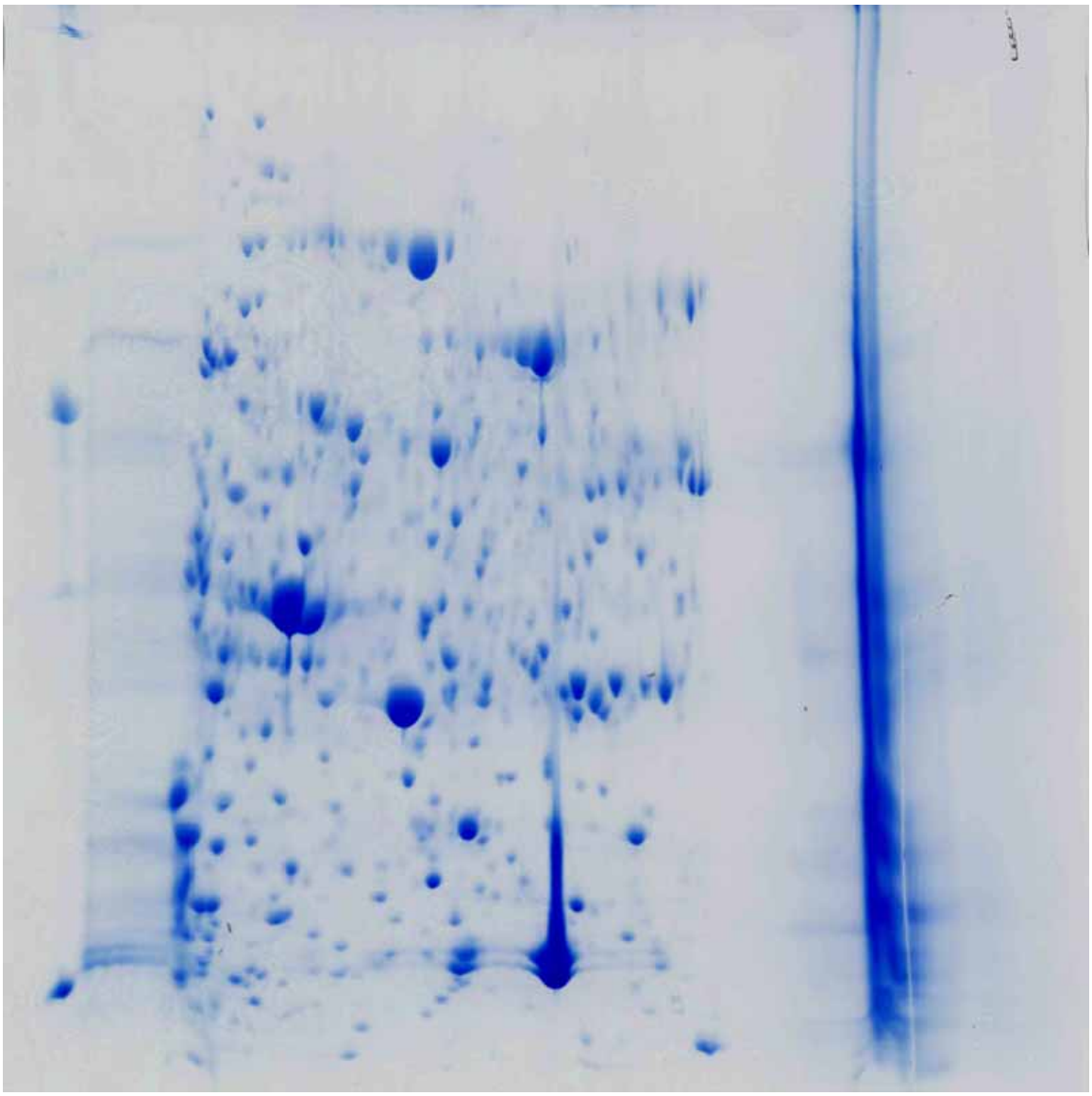

Figure 6-2. Colloidal Gel Stain image indicating all proteins from a two-dimensional gel electrophoresis separation of *Arabidopsis thaliana* protein extract.

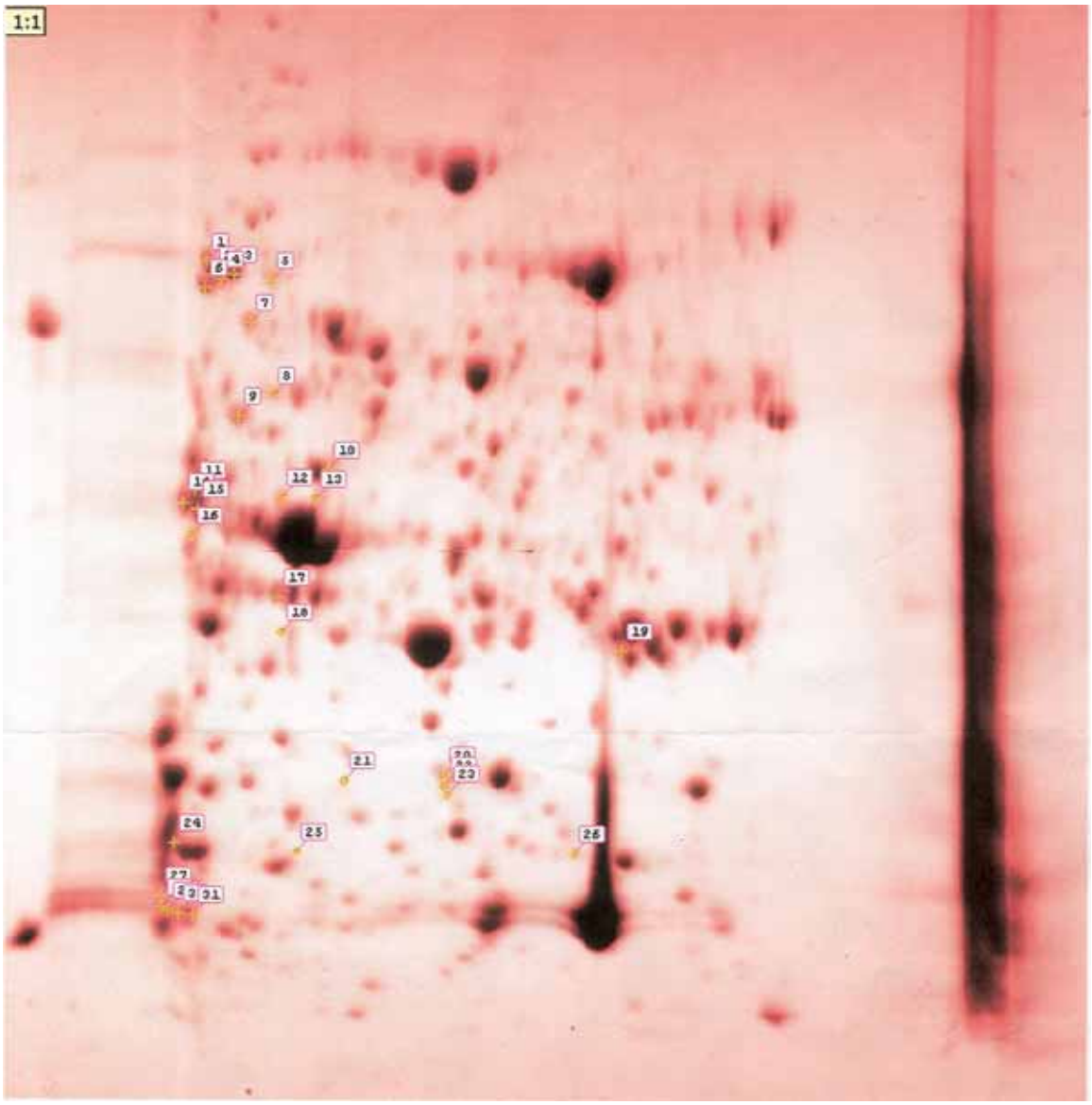

Figure 6-3. Investigator™ ProPic™ image of Colloidal stained gel and corresponding spots for excision.

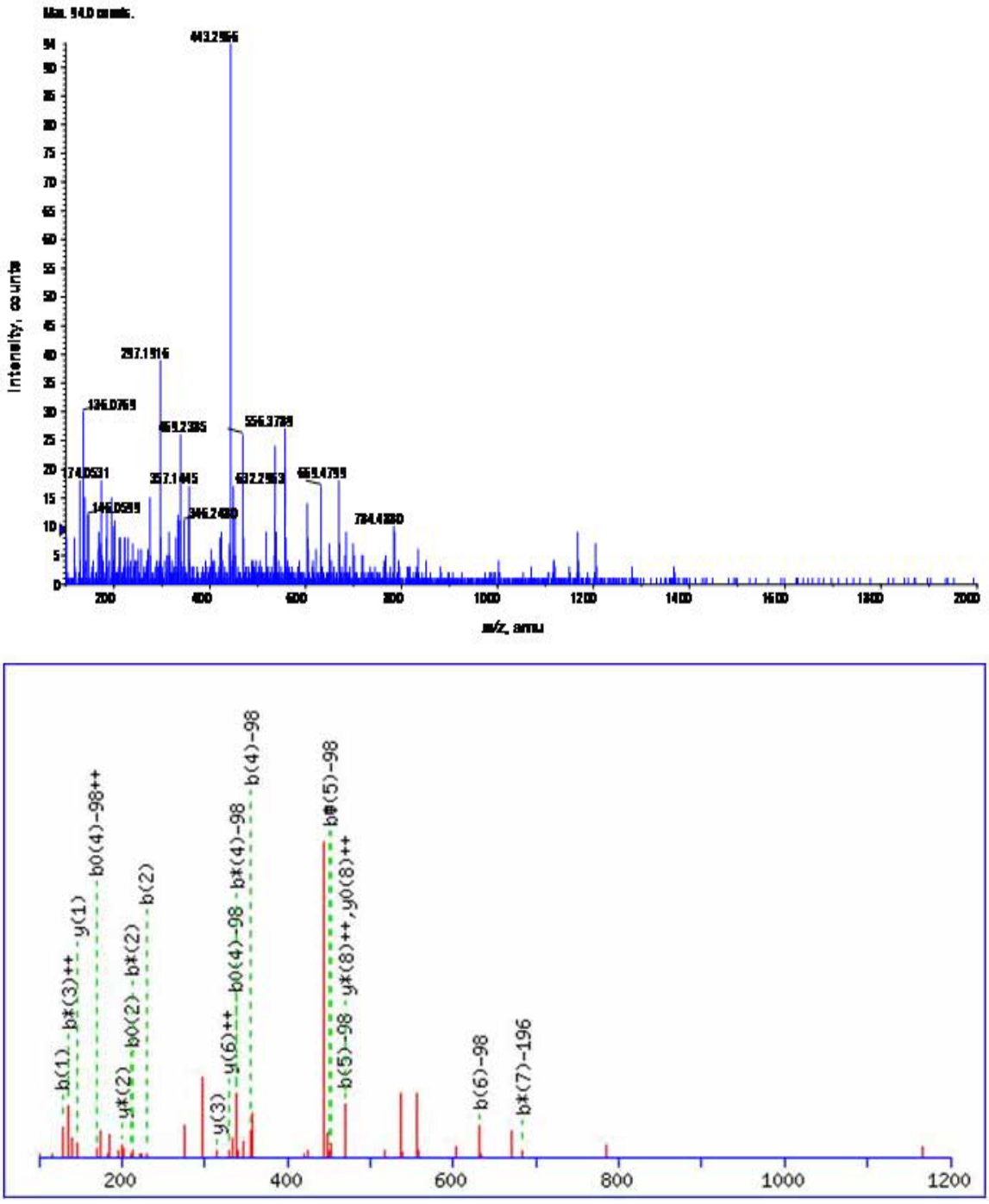

Figure 6-4. MS/MS spectrum and corresponding MASCOT search results of the phosphopeptide QTGpSLYpSDWDLLPAK (*m/z* 1852.9480) from the unknown protein (gi|30725696) identified in spot 1.

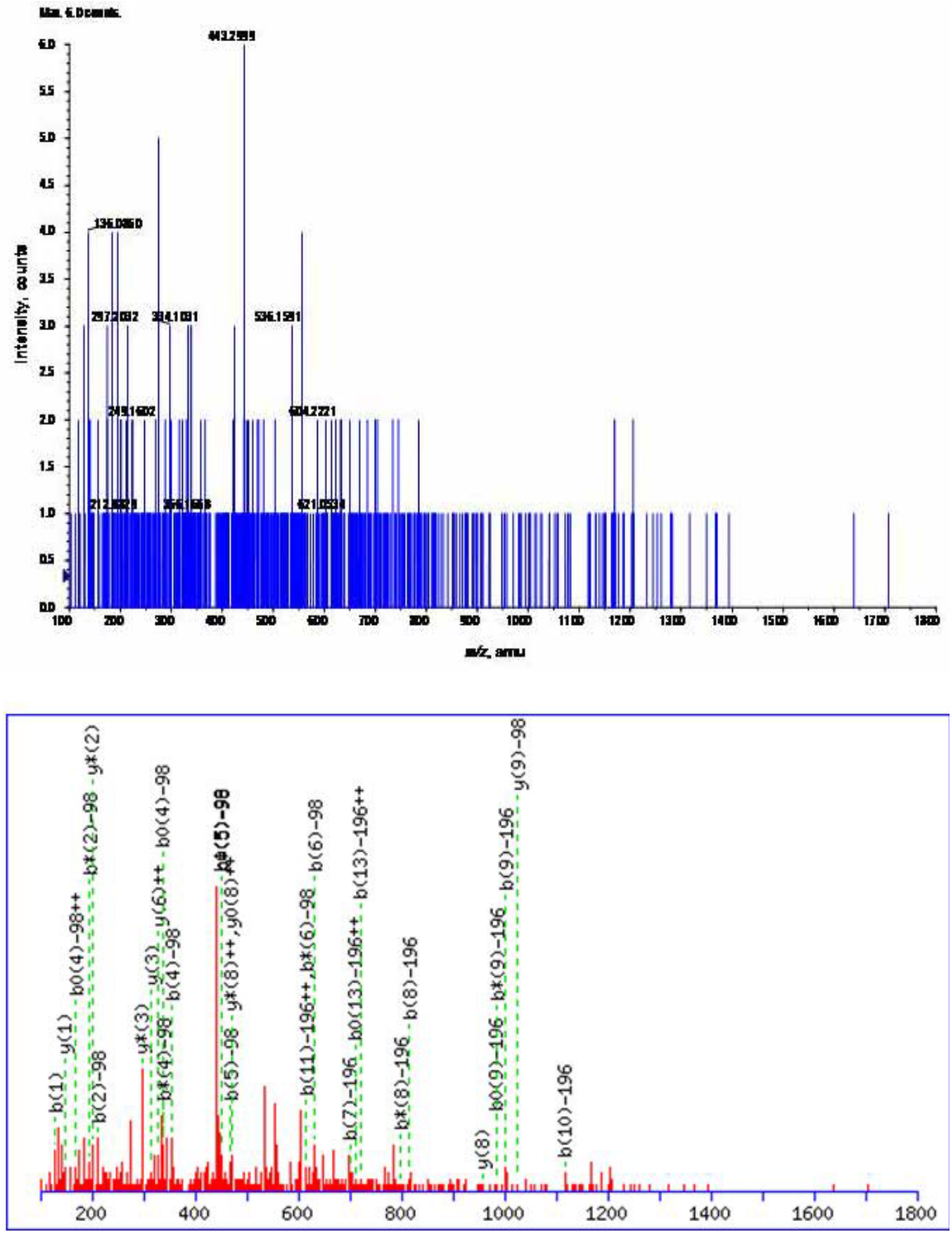

Figure 6-5. MS/MS spectrum and corresponding MASCOT search results of the phosphopeptide QpTGSLYpSDWDLLPAK (*m/z* 1852.9516) from the unknown protein (gi|30725696) identified in spot 2.

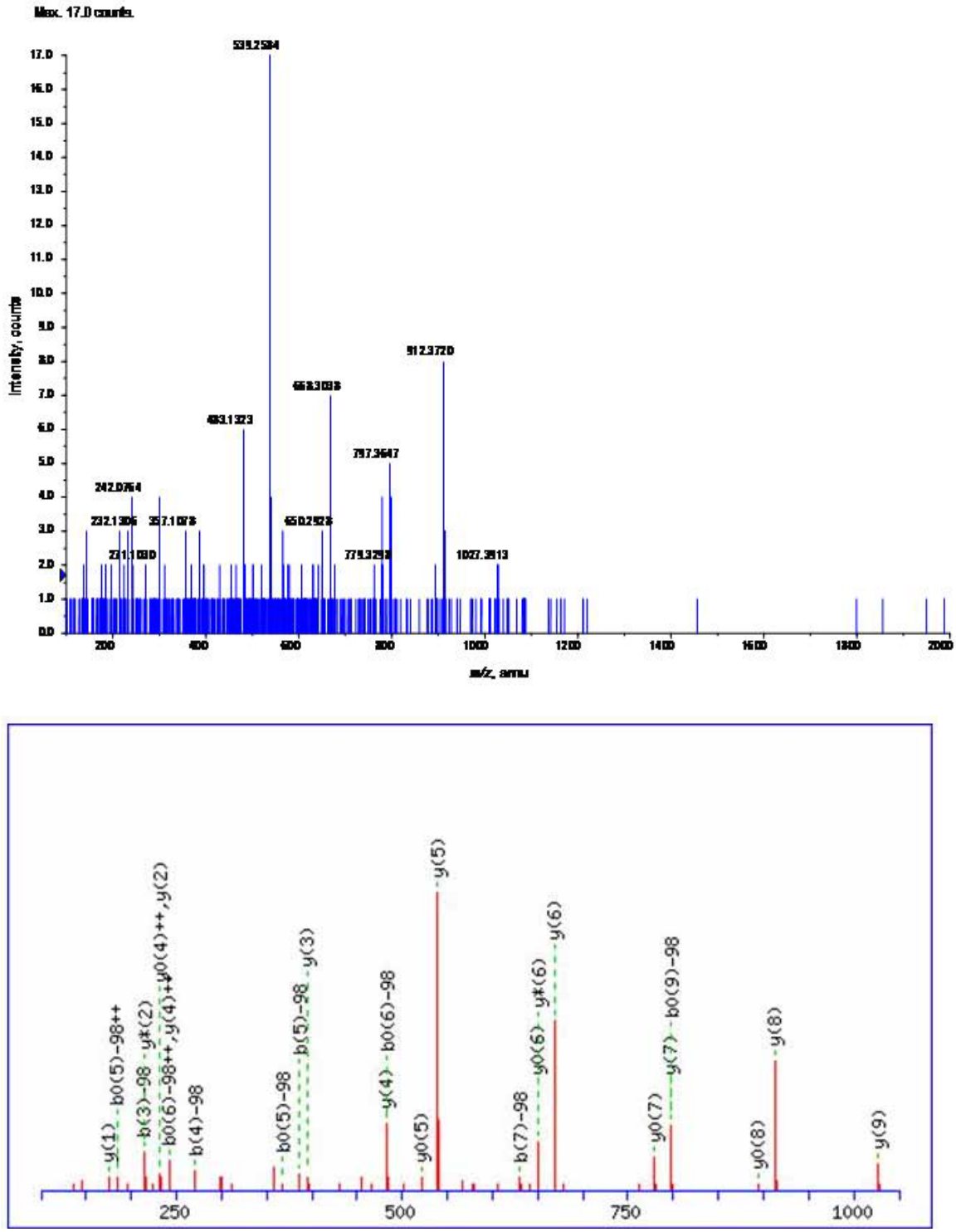

Figure 6-6. MS/MS spectrum and corresponding MASCOT search results of the phosphopeptide SGpSGDDEEGSYGR (*m/z* 1394.4946) from the unknown protein (gi|23308191) identified in spot 3.

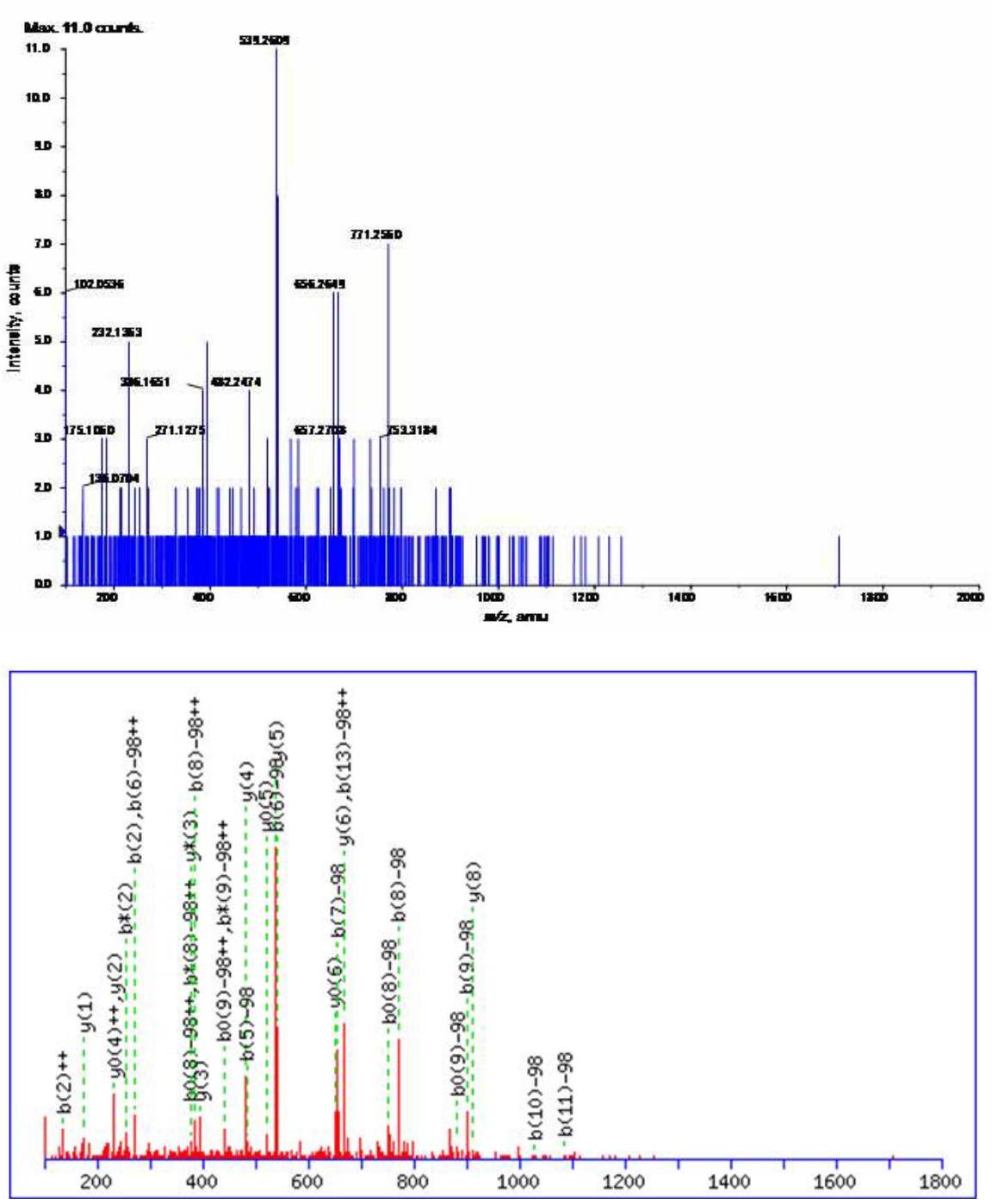

Figure 6-7. MS/MS spectrum and corresponding MASCOT search results of the phosphopeptide NRSGpSGDDEEGSYGR (*m/z* 1665.6520) from the unknown protein (gi|23308191) identified in spot 3.

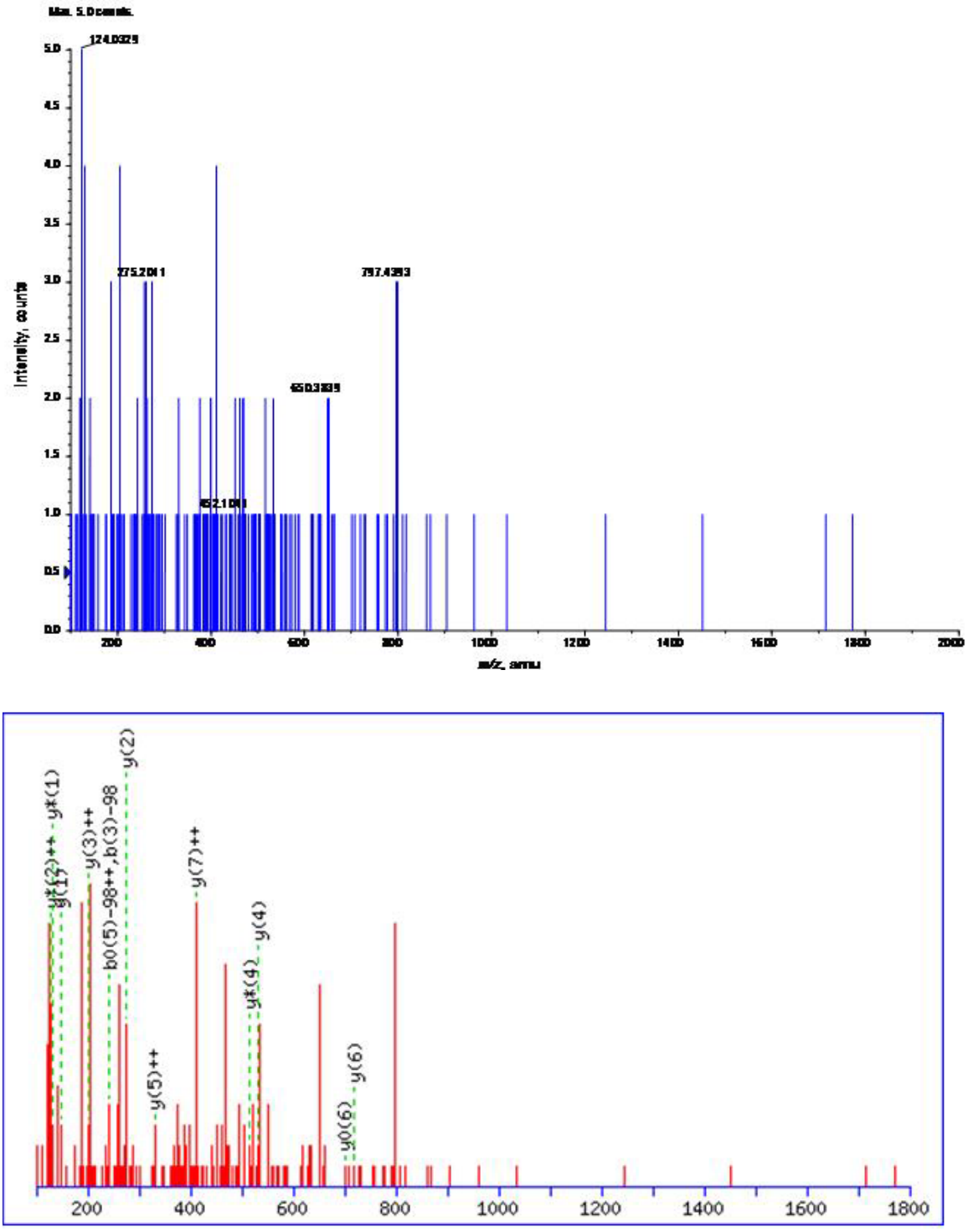

Figure 6-8. MS/MS spectrum and corresponding MASCOT search results of the phosphopeptide pTTGEEEKK (*m/z* 1000.5043) from the low temperature-induced protein (gi|509262) identified in spot 8.

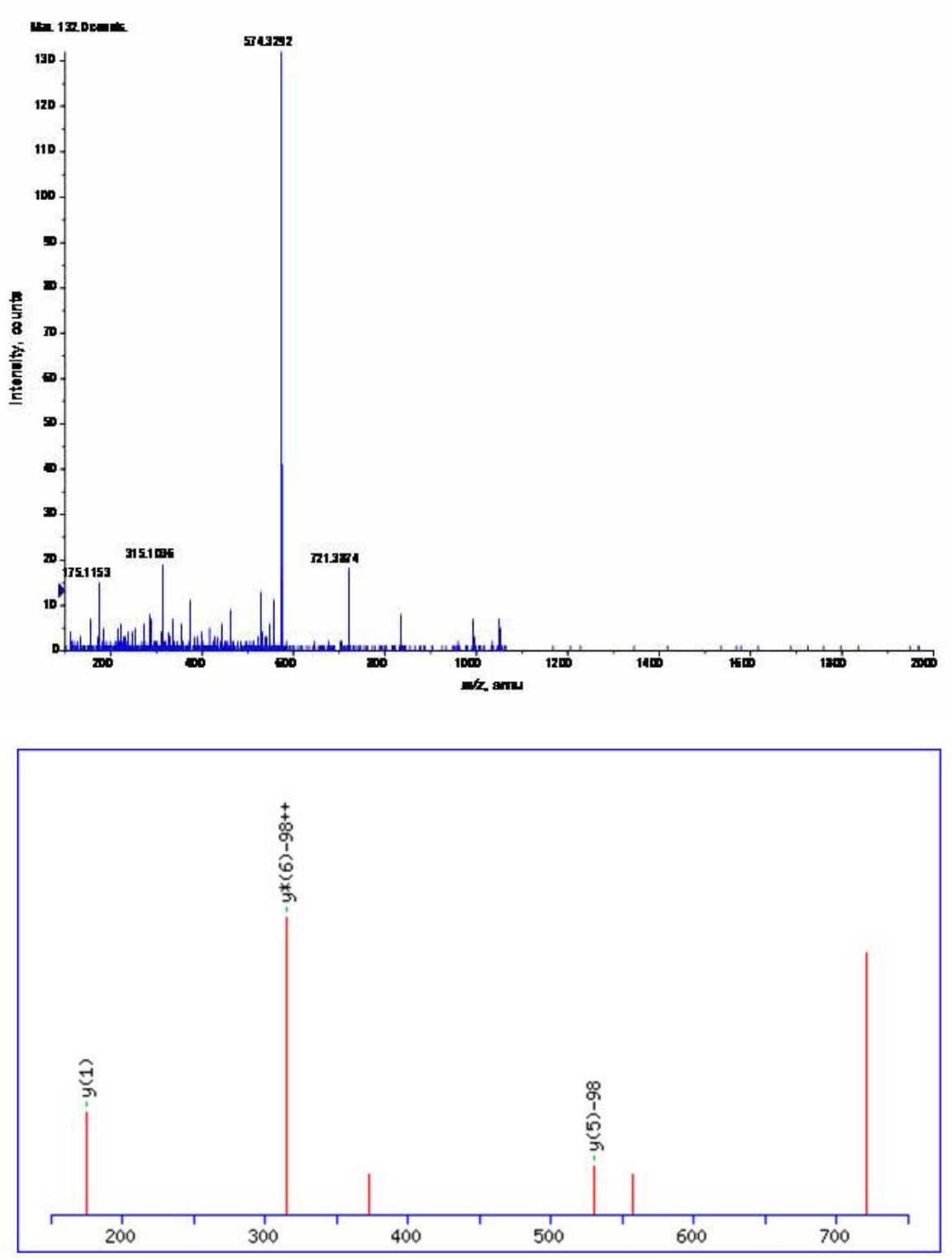

Figure 6-9. MS/MS spectrum and corresponding MASCOT search results of the phosphopeptide SFGLDSpSQAR (*m/z* 1146.6510) from putative protein 1 photosystem II oxygen-evolving complex (gi|4835233) identified in spot 13.

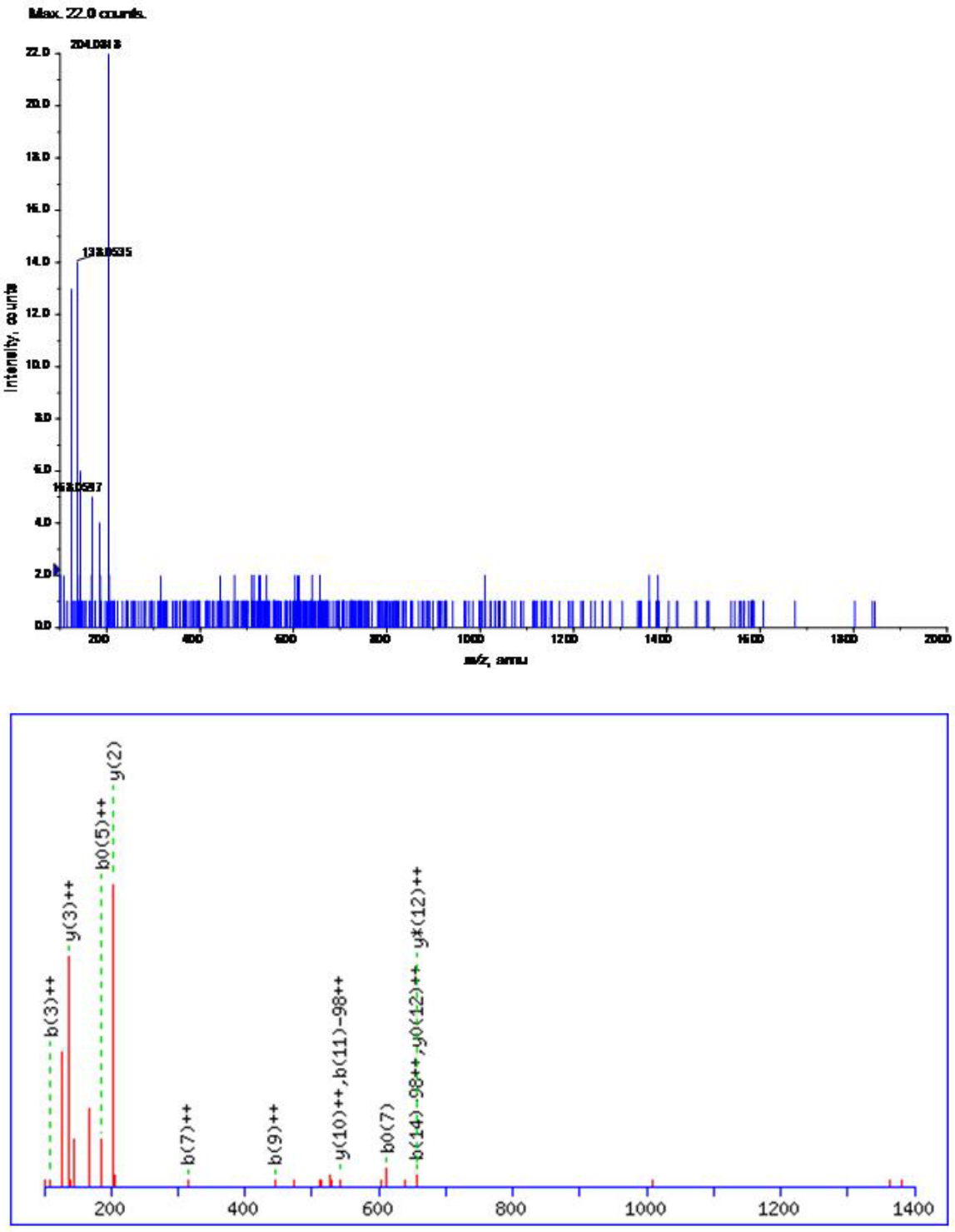

Figure 6-10. MS/MS spectrum and corresponding MASCOT search results of the phosphopeptide GTGTANQCPpTIDGGSETFSFKPGKYAGK (*m/z* 2955.3193) from the 33 kDa polypeptide of oxygen-evolving complex (gi|10177538) identified in spot 14.

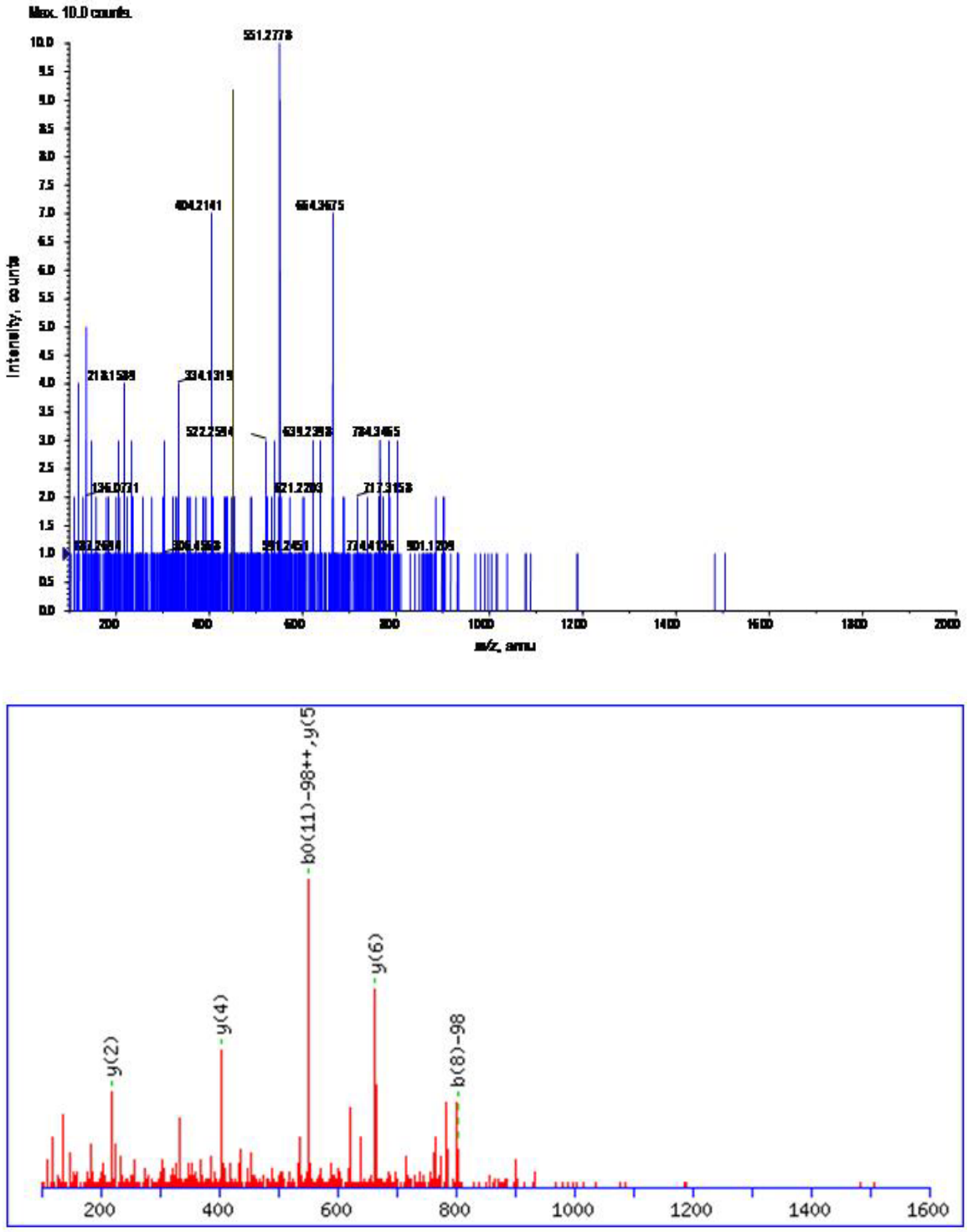

Figure 6-11. MS/MS spectrum and corresponding MASCOT search results of the phosphopeptide SPASDpTYVIFGEAK (*m/z* 1563.7827) from the unknown protein (gi|48310641) identified in spot 15.

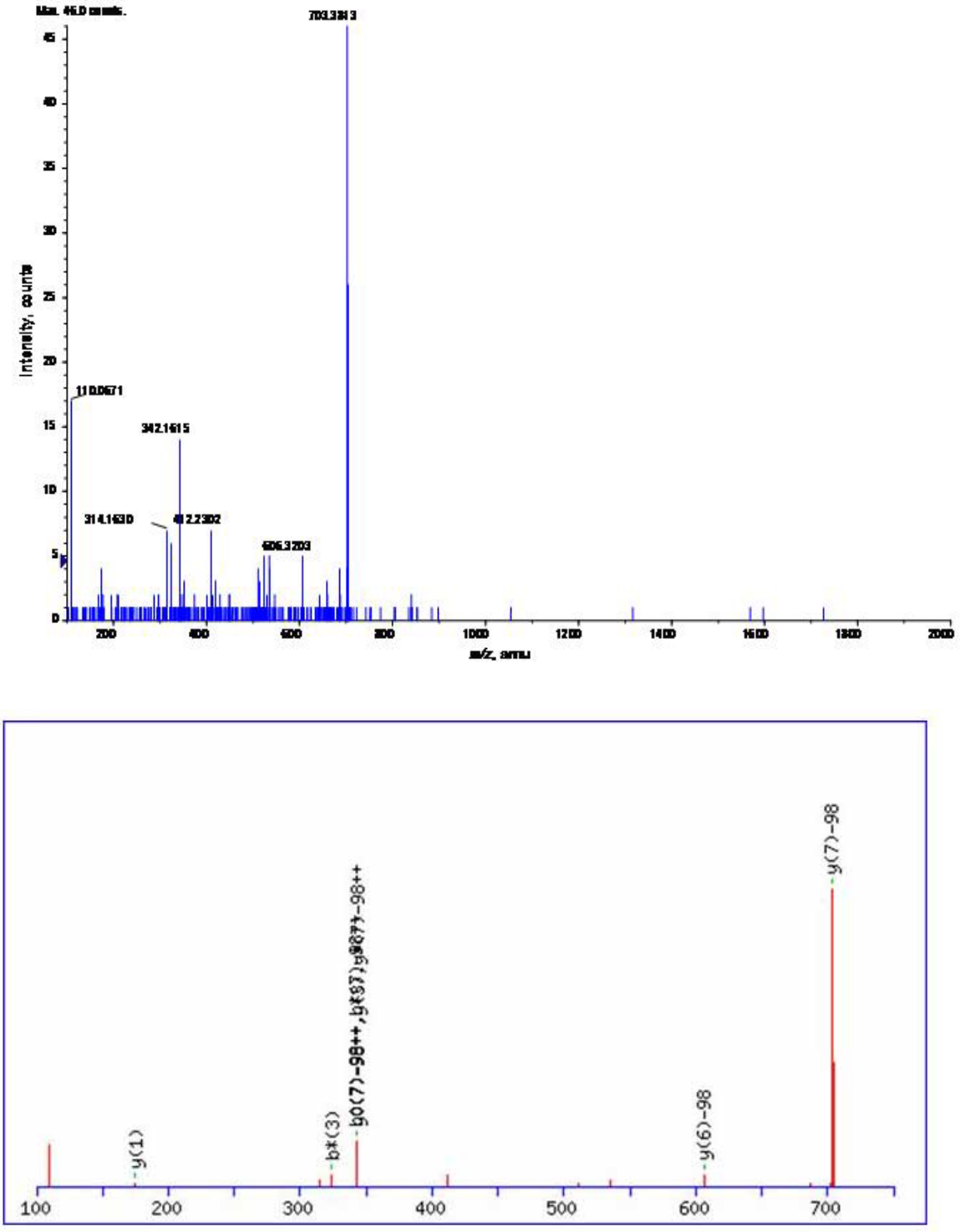

Figure 6-12. MS/MS spectrum and corresponding MASCOT search results of the phosphopeptide RSPpSPPPAR (*m/z* 1043.5207) from the RSZp22 protein (gi|2582645) identified in spot 19.

Table 6-1. List of proteins identified from 2-DE spots.

Function or location	gi number	Provisional name	Spot #
Photosynthesis/ chloroplast	gi|5734518	photosystem I subunit III precursor	24
	gi|4835233	putative protein 1 photosystem II oxygen-evolving complex*	12,13
	gi|23198314	putative chloroplast RNA binding protein precursor	10
	gi|22571	33 kDa oxygen-evolving protein	15,16
	gi|10177538	33 kDa polypeptide of oxygen-evolving complex*	12,13,14
	gi|21592906	23 kDa polypeptide of oxygen-evolving complex (OEC)	18
	gi|16374	chlorophyll a/b binding protein (LHCP AB 180)	10,13,17,18
	gi|12642854	putative photosystem II type I chlorophyll a/b binding protein	17
	gi|6539610	thioredoxin m2	29,30
	gi|28973241	putative thioredoxin-m	28,29,30,31
	gi|9759188	thylakoid lumenal 17.4 kD protein, chloroplast precursor	25
Ribosomal	gi|23296539	putative ribosomal protein S1	7
	gi|21553851	40S ribosomal protein S2	14
	gi|6566279	RF12	15
	gi|6598334	putative ribosomal protein S4	16
	gi|21592577	40S ribosomal protein S8-like	16
	gi|166867	ribosomal protein S11 (probable start codon at bp 67)	23
	gi|21537373	40S ribosomal protein S12-2	20,22,23
	gi|8567795	40S ribosomal protein S17, putative	24
	gi|5881716	ribosomal protein S18	27
	gi|4567232	40S ribosomal protein S25	28,29
	gi|7270903	ribosomal protein S25	28,29
	gi|21553790	50S ribosomal protein L29	24
	gi|10177580	ribosomal protein L32	24
	gi|7268562	ribosomal protein L32-like protein	24,26
	gi|6478915	putative 60S acidic ribosomal protein P0	10
	gi|5923684	putative 60S acidic ribosomal protein, 5' partial	15
	gi|23397084	putative 60S ribosomal protein	23
	gi|21617903	60S acidic ribosomal protein P3	24
	gi|23308199	60S ribosomal protein L7A	16
	gi|16497	Plastid ribosomal protein CL15	19
Metabolism	gi|6056409	glutathione S-transferase	19
	gi|3461818	putative glutathione S-transferase	19
	gi|2266412	glutathione S-transferase	19
	gi|20197312	glutathione S-transferase (GST6)	19
	gi|8885622	N-glyceraldehyde-2-phosphotransferase-like	14
	gi|7572929	alpha-galactosidase-like protein	6
	gi|7263568	sedoheptulose-bisphosphatase precursor	8,20

*Identified as being phosphorylated by MASCOT search engine

Table 6-1. Continued.

Function or location	gi number	Provisional name	Spot #
Metabolism	gi|38603934	S-adenosylmethionine:2-demethylmenaquinone methyltransferase-like	21,23
	gi|6728985	putative S-adenosylmethionine:2-demethylmenaquinone methyltransferase	23
	gi|6175179	putative ribose 5-phosphate isomerase	17
	gi|5834508	thiosulfate sulfurtransferase	10
	gi|2895510	putative pectin methylesterase	19
	gi|21592462	glycine decarboxylase complex H-protein	24
	gi|21554426	biotin carboxyl carrier protein precursor-like protein	12
	gi|1066348	acetyl-CoA carboxylase biotin-containing subunit	12
	gi|23505901	diaminopimelate epimerase - like protein	10
	gi|21689621	p-nitrophenylphosphatase-like protein	12,13
	gi|1944432	ribulosebisphosphate carboxylase, large subunit	1,2,3
	gi|16194	ribulose bisphosphate carboxylase, small subunit	27,28,29,30,31
	gi|13926229	ribulose bisphosphate carboxylase, small subunit protein	27,28,29,30,31
Protein synthesis, processing, trafficking	gi|6686821	elongation factor 1B alpha-subunit	11,15,16
	gi|461073	elongation factor 1 beta, EF-1 beta {internal fragments}	15
	gi|398606	eEF-1beta	15
	gi|7270846	multiubiquitin chain binding protein (MBP1)	2,6
	gi|6899901	peptidylprolyl isomerase ROC4	20,22
	gi|11762200	cyclophilin/peptidylprolyl isomerase ROC4	20
	gi|18398710	nascent polypeptide-associated complex (NAC) domain-containing protein	15
	gi|4115918	similar to nascent polypeptide associated complex alpha chain	11,14
	gi|6561948	alpha NAC-like protein	11
	gi|38454142	peptide methionine sulfoxide reductase-like protein	12,13
	gi|2582645	RSZp22 protein (splicing factor)	19
	gi|487791	GF14omega isoform	14
	gi|1256534	GF14 chi chain	14,15
	gi|12247993	putative 14-3-3 protein GF14epsilon	16
	gi|16974503	eukaryotic initiation factor 5A (eIF-5A) like protein	23
	gi|9295717	initiation factor 5A-4, putative (eIF-5A)	20,22,23
	gi|12083338	putative glycine cleavage system H protein precursor	24

*Identified as being phosphorylated by MASCOT search engine

Table 6-1. Continued.

Function or location	gi number	Provisional name	Spot #
Signaling/stress	gi|21553555	dehydration stress-induced protein	24,30
	gi|6957717	putative RAD23	5
	gi|388259	cor47 (low temperature-induced, ABA, severe water stress)	1,2,4
	gi|509262	lti45 (low temperature-induced)*	8
	gi|21593067	HSP associated protein like	5
	gi|20258909	putative protein phosphatase 2C	7
	gi|16223	calmodulin	27,28
	gi|15217459	calreticulin 2 (CRT2)	6,7
	gi|12643243	Calreticulin 2 precursor	1
	gi|3212877	similar to late embryogenesis abundant proteins	9
	gi|23197658	apospory-associated protein C-like protein	12,13
	gi|21555216	submergence induced protein 2A	18
	gi|11761812	glutathione dependent dehydroascorbate reductase precursor	17
	gi|7269215	putative major latex protein	26
Nuclear	gi|99684	DNA-binding protein	13,14,16
	gi|549975	nucleosome assembly protein I-like protein; similar to mouse nap I	1
	gi|12325357	RNA-binding protein, putative; 35994-37391	18
	gi|30023780	putative RNA-binding protein	16
	gi|20334750	histone H2A like protein	24
Protease	gi|26452816	putative carboxyl-terminal proteinase	12
	gi|10177410	HCF106	16
	gi|23397070	putative cysteine proteinase AALP	17,18
	gi|19548039	Cysteine proteinase RD21a precursor	11,12,15,16
	gi|21595063	putative aspartyl protease	10
	gi|1354272	aspartic proteinase	13
	gi|10177282	20S proteasome subunit PAF1	10
	gi|4887543	ATP-dependent Clp protease subunit ClpP	19
Membrane	gi|7268821	endomembrane-associated protein	10
Cytoskeleton	gi|7268885	tubulin beta-9 chain	5
	gi|16323374	putative tubulin beta-4 chain	5
	gi|20334778	beta tubulin 1, putative	5
Putative Proteins	gi|9955526	putative protein	1
	gi|7594543	putative protein	19
	gi|7340724	putative protein	23
	gi|7287993	putative protein	12,13
	gi|7270957	putative protein	24
	gi|7269995	putative protein	17
	gi|7269580	putative protein	19
	gi|7269388	putative protein	21,23
	gi|7267907	putative protein	20
Putative Proteins	gi|7267473	putative protein (fragment)	3
	gi|6759449	putative protein	6

*Identified as being phosphorylated by MASCOT search engine

Table 6-1. Continued.

Function or location	gi number	Provisional name	Spot #
Unnamed/hypothetical/ unknown proteins	gi|9758428	unnamed protein product	6
	gi|16471	unnamed protein product	7
	gi|15795158	unnamed protein product	16
	gi|7268071	hypothetical protein	7
	gi|25406857	hypothetical protein [imported]	23
	gi|6466955	unknown protein	10
	gi|3128209	unknown protein	6
	gi|29824163	unknown protein	14
	gi|28393989	unknown protein	17
	gi|24417462	unknown	6
	gi|24417358	unknown	20
	gi|24417296	unknown	6
	gi|23296326	unknown protein	8
	gi|21555690	unknown	20,22,23
	gi|21555497	unknown	20,21,22
	gi|21554910	unknown	28,29
	gi|21386997	unknown protein	29,30
	gi|20258893	unknown protein	16
	gi|18377530	unknown protein	12
	gi|17065622	unknown protein	5
	gi|13751875	unknown protein	7
	gi|48310641	unknown protein*	15,16
	gi|30725696	unknown protein*	1,2
	gi|23308191	unknown protein*	2,3,4
	gi|19699303	unknown protein	14,15
	gi|19699114	unknown protein	16
	gi|13265523	unknown protein	9,12,13
	gi|3193303	unknown protein	11,14,15

*Identified as being phosphorylated by MASCOT search engine

CHAPTER 7
RESEARCH OVERVIEW

Mass spectrometry has emerged as the method of choice for phosphoprotein analysis. Presented was the application of several mass spectrometric techniques for phosphorylation analysis. This research was begun by optimizing matrix conditions for MALDI-TOFMS analysis of phosphopeptides and investigation of various enrichment techniques. Once optimized, these enrichment techniques were then applied to various phosphoproteomic projects relating to *Arabidopsis thaliana*.

Project 1, the development of methods for identifying phosphorylated sites of a protein, demonstrated the development of a simple, cost effective method for phosphorylation identification using β-elimination and a MALDI-TOF MS. This method will be useful for laboratories with only a MALDI-TOF-MS instrument, however, sequences will have to be verified once the targets are found due to possible false positives. Application of the developed methods resulted in the identification of numerous phosphorylation sites of a calcium-dependent protein kinase. Attainment of this goal was possible with the use of three different types of mass spectrometers (MALDI-TOF-MS, QIT-MS and QqTOF-MS). Comparison of results obtained by these three instruments gave significant overlap showing that these methods are indeed complimentary. Although numerous autophosphorylation sites were identified, precursor ion scanning indicated the presence of other sites that could be phosphorylated, however these sites remained elusive with the types of analyses attempted. Since the autophosphorylation properties of CDPKs are not fully understood, this work is presented

as a contribution towards their understanding. Compilation of these sites along with identified sites of other CDPKs will aid in their understanding by possibly identifying a common sequence motif among these kinases.

Project 2 demonstrates the implementation of a method for the identification of 28 possible substrates of a CDPK from a complex system. Achievement of this goal was accomplished by the utilization of several newer technologies for phosphorylation analysis that became available during the time of this project. Included were various separation techniques as well as mass spectrometric analysis. The information attained in this project may be used towards the understanding of CDPKs, including their functions based on target substrates as well as their target specificity based on binding motifs. However, additional work has to be performed to achieve this goal because of the lack of phosphorylation site mapping due to insufficient fragmentation of the phosphopeptides after MS/MS analysis. Despite this setback, the information attained may be useful for designing future experiments, that is, the proteins identified may be tested individually as substrates of the kinase for verification.

The application of affinity chromatography and mass spectrometery was then performed for the identification of 14-3-3 interactors from *Arabidopsis thaliana* protein extracts for Project 3. Illustrated in Chapter 5 was the identification of 263 interacting proteins, included among these were several verified phosphorylated proteins. It is expected that a much larger fraction of these identified interactors are phosphorylated, however, phosphorylation sites of these proteins are needed. Further interpretation of these results may lead to an increase in the knowledge of the binding specificity of these interacting proteins with the 14-3-3s which may lead to the discovery of additional

binding motifs for this family of proteins. Since 14-3-3 proteins are known to interact with phosphorylated proteins, the goal of this project was to correlate these identified proteins with substrates of the kinase to see if there was any overlap between the two families of proteins. Table 7-1 shows the proteins that were identified in both sets of analyses. At this time, the biological significance of this information has not been determined, however, this information may be used in the future for understanding the function of the 14-3-3 proteins as well as their overlap with CDPKs.

Finally, Project 4 involved the development of a high-throughput approach for phosphoproteomic analyses. Illustrated was the application of robotic instrumentation for high-throughput phosphoproteomic analysis of a complex *Arabidopsis* protein extract resulting in the identification of 138 proteins, many of which are expected phosphorylated proteins. However, verification of the phosphorylation sites of many of these proteins was not possible. Comparison of the identified potential phosphorylated proteins with that of the 14-3-3 interactors resulted in the identification of 39 proteins (Table 7-2), several of which were verified phosphorylated proteins, increasing our confidence that these proteins are likely to be phosphorylated. The incorporation of robotics in this analysis resulted in less time required by the user for sample preparation, however, improvements still need to be made to the instrumentation due to occasional errors. Also, the addition of a filter specific for the Pro-Q stain will also result in an increase in throughput for this method. This high-throughput approach for phosphorylation analysis may be applied to other projects, freeing the time of the user for other important tasks.

The results of this research are presented as a contribution towards the understanding of the physiological roles of calcium-dependent protein kinases and 14-3-3 proteins. Application of the presented methods may be applied to other kinases in the CDPK superfamily for identification of signaling networks in which each kinase participates, imparting insight into the physiological roles of these kinases. Since these kinases are found in vascular and nonvascular plants, a better understanding of the regulation of key aspects such as cellular function, metabolism, and response to external signals will be gained for a wide range of plants. These findings will have a great impact on biotechnology which will in turn affect important areas such as agriculture. The developed methods may also be applied to other fields of study, for example, studying kinases in disease research.

Table 7-1. List of proteins identified as both CDPK substrates and 14-3-3 interactors.

Function or location	Provisional name
Photosynthesis/chloroplast	Fructose-bisphosphate aldolase
Nuclear	Nucleosome assembly protein
Protein synthesis, processing, trafficking	Peptidyl-prolyl cis-trans isomerase, chloroplast precursor
Metabolism	ATP synthase alpha chain
	Putative H+-transporting ATP synthase
	Glutathione S-transferase like protein
	Glyceraldehyde-3-phosphate dehydrogenase B, chloroplast precursor
	Phosphoribulokinase, chloroplast precursor
	Splice Isoform 2 of Ribulose bisphosphate carboxylase/oxygenase activase,
Signaling/stress	Heat shock cognate 70 kDa protein 2

Table 7-2. Correlating proteins identified in 2-DE and 14-3-3 experiments.

Function or location	gi number	Provisional name
Photosynthesis/ chloroplast	gi|16374	chlorophyll a/b binding protein (LHCP AB 180)
	gi|12642854	putative photosystem II type I chlorophyll a/b binding protein
	gi|21592906	23 kDa polypeptide of oxygen-evolving comlex (OEC)
	gi|22571	33 kDa oxygen-evolving protein
	gi|10177538	33 kDa polypeptide of oxygen-evolving complex*
	gi|4835233	putative protein 1 photosystem II oxygen-evolving complex*
	gi|28973241	putative thioredoxin-m
Ribosomal	gi|23296539	putative ribosomal protein S1
	gi|16497	Plastid ribosomal protein CL15
	gi|10177580	ribosomal protein L32
	gi|6598334	putative ribosomal protein S4
	gi|166867	ribosomal protein S11 (probable start codon at bp 67)
	gi|21592577	40S ribosomal protein S8-like
	gi|23397084	putative 60S ribosomal protein
	gi|23308199	60S ribosomal protein L7A
	gi|6566279	RF12
Metabolism	gi|2266412	glutathione S-transferase
	gi|7263568	sedoheptulose-bisphosphatase precursor
	gi|21689621	p-nitrophenylphosphatase-like protein
	gi|6175179	putative ribose 5-phosphate isomerase
	gi|1944432	ribulosebisphosphate carboxylase
	gi|16194	ribulose bisphosphate carboxylase
	gi|13926229	ribulose bisphosphate carboxylase, small subunit protein
Protein synthesis, processing, trafficking	gi|1256534	GF14 chi chain
	gi|12247993	putative 14-3-3 protein GF14epsilon
	gi|6899901	peptidylprolyl isomerase ROC4*
	gi|38454142	peptide methionine sulfoxide reductase-like protein
Signaling/stress	gi|11761812	glutathione dependent dehydroascorbate reductase precursor
	gi|23197658	apospory-associated protein C-like protein
Nuclear	gi|12325357	RNA-binding protein, putative; 35994-37391
	gi|99684	DNA-binding protein
Protease	gi|1354272	aspartic proteinase
	gi|23397070	putative cysteine proteinase AALP
Membrane	gi|7268821	endomembrane-associated protein
Putative proteins	gi|7269388	putative protein
	gi|7269995	putative protein
	gi|7287993	putative protein
Unknown proteins	gi|13265523	unknown protein
	gi|28393989	unknown protein

*Identified as being phosphorylated by MASCOT search engine

LIST OF REFERENCES

1. Yan, J. X.; Packer, N. H.; Gooley, A. A.; Williams, K. L. *Journal of Chromatography A* **1998**, *808*, 23-41.

2. Steen, H.; Kuster, B.; Mann, M. *Journal of Mass Spectrometry* **2001**, *36*, 782-790.

3. Bennett, K.; Stensballe, A.; Podtelejnikov, A.; Moniatte, M.; Jensen, O. *Journal of Mass Spectrometry* **2002**, *37*, 179-190.

4. Wang, D.; Harper, J. F.; Gribskov, M. *Plant Physiology* **2003**, *132*, 2152-2165.

5. Harmon, A. C.; Gribskov, M.; Gubrium, E.; Harper, J. F. *New Phytologist* **2001**, *151*, 175-183.

6. Harper, J. F.; Bretton, G.; Harmon, A. C. *Annual Review of Plant Biology* **2004**, *55*, 263-288.

7. Harmon, A. C.; Gribskov, M.; Harper, J. F. *Trends in Plant Science* **2000**, *5*, 154-159.

8. Cheng, S. H.; Willmann, M. R.; Chen, H. C.; Sheen, J. *Plant Physiology* **2002**, *129*, 469-485.

9. Chehab, E. W.; Patharkar, O. R.; Hegeman, A. D.; Taybi, T.; Cushman, J. C. *Plant Physiology* **2004**, *135*, 1430-1446.

10. Chaudhuri, S.; Seal, A.; DasGupta, M. *Plant Physiology* **1999**, *120*, 859-866.

11. Saha, P.; Singh, M. *Biochemistry Journal* **1995**, *305* (Pt 1), 205-210.

12. Anil, V. S.; Rao, K. S. *Phytochemistry* **2001**, *58*, 203-212.

13. Harper, J.; Harmon, A. *Nature Reviews. Molecular Cell Biology* **2005**, In Press.

14. Dammann, C.; Ichida, A.; Hong, B. M.; Romanowsky, S. M.; Hrabak, E. M.; Harmon, A. C.; Pickard, B. G.; Harper, J. F. *Plant Physiology* **2003**, *132*, 1840-1848.

15. Lu, S. X.; Hrabak, E. M. *Plant Physiology* **2002**, *128*, 1008-1021.

16. Anil, V. S.; Harmon, A. C.; Rao, K. S. *Plant and Cell Physiology* **2003**, *44*, 367-376.

17. Pical, C.; Fredlund, K. M.; Petit, P. X.; Sommarin, M.; Moller, I. M. *FEBS Letters* **1993**, *336*, 347-351.

18. Kennelly, P. J.; Krebs, E. G. *Journal of Biological Chemistry* **1991**, *266*, 15555-15558.

19. Bylund, D. B.; Krebs, E. G. *Journal of Biological Chemistry* **1975**, *250*, 6355-6361.

20. Huang, J. Z.; Hardin, S. C.; Huber, S. C. *Archives of Biochemistry and Biophysics* **2001**, *393*, 61-66.

21. Hernandez Sebastia, C.; Hardin, S. C.; Clouse, S. D.; Kieber, J. J.; Huber, S. C. *Archives of Biochemistry and Biophysics* **2004**, *428*, 81-91.

22. Roberts, D. M.; Harmon, A. C. *Annual Review of Plant Physiology and Molecular Biology* **1992**, *43*, 375-414.

23. Neumann, G. M.; Thomas, I.; Polya, G. M. *Plant Science* **1996**, *114*, 45-51.

24. Huang, J. Z.; Huber, S. C. *Plant and Cell Physiology* **2001**, *42*, 1079-1087.

25. Loog, M.; Toomik, R.; Sak, K.; Muszynska, G.; Jarv, J.; Ek, P. *European Journal of Biochemistry* **2000**, *267*, 337-343.

26. Sebastia, C. H.; Hardin, S. C.; Clouse, S. D.; Kieber, J. J.; Huber, S. C. *Archives of Biochemistry and Biophysics* **2004**, *428*, 81-89.

27. Moore, B. W.; Perez, V. J. *Physiological and Biochemical Aspects of Nervous Integration* **1976**, 343-359.

28. Wu, K.; Rooney, M. F.; Ferl, R. J. *Plant Physiology* **1997**, *114*, 1421-1431.

29. Ferl, R. J.; Lu, G.; Bowen, B. W. *Genetica* **1994**, *92*, 129-138.

30. Ferl, R. J. *Annual Review of Plant Physiology and Plant Molecular Biology* **1996**, *47*, 49-73.

31. Aitken, A. *Trends in Cell Biology* **1996**, *6*, 341-347.

32. Yaffe, M. B.; Rittinger, K.; Volinia, S.; Caron, P. R.; Aitken, A.; Leffers, H.; Gamblin, S. J.; Smerdon, S. J.; Cantley, L. C. *Cell* **1997**, *91*, 961-971.

33. Muslin, A. J.; Tanner, J. W.; Allen, P. M.; Shaw, A. S. *Cell* **1996**, *84*, 889-897.

34. Wurtele, M.; Jelich-Ottmann, C.; Wittinghofer, A.; Oecking, C. *Embo Journal* **2003**, *22*, 987-994.

35. Fuglsang, A. T.; Borch, J.; Bych, K.; Jahn, T. P.; Roepstorff, P.; Palmgren, M. G. *Journal of Biological Chemistry* **2003**, *278*, 42266-42272.

36. Rubio, M. P.; Geraghty, K. M.; Wong, B. H.; Wood, N. T.; Campbell, D. G.; Morrice, N.; Mackintosh, C. *Biochemistry Journal* **2004**, *379*, 395-408.

37. Mann, M.; Ong, S.; Gronborg, M.; Steen, H.; Jensen, O.; Pandey, A. *Trends in Biotechnology* **2002**, *20*, 261-268.

38. Carr, S. A.; Hemling, M. E.; Bean, M. F.; Roberts, G. D. *Analytical Chemistry* **1991**, *63*, 2802-2824.

39. Siuzdak, G. *Proceedings of the National Academy of Science of the United States of America* **1994**, *91*, 11290-11297.

40. Hillenkamp, F.; Karas, M.; Beavis, R. C.; Chait, B. T. *Analytical Chemistry* **1991**, *63*, 1193A-1203A.

41. Beavis, R. C.; Chaudhary, T.; Chait, B. T. *Organic Mass Spectrometry* **1992**, *27*, 156-158.

42. Yates, J. R., 3rd. *Journal of Mass Spectrometry* **1998**, *33*, 1-19.

43. Guilhaus. *Journal of Mass Spectrometry* **1995**, *30*, 1519-1532.

44. March, R. E. *Journal of Mass Spectrometry* **1997**, *32*, 351-369.

45. March, R. E. *Rapid Communications in Mass Spectrometry* **1998**, *12*, 1534-1554.

46. Chernushevich, I. V.; Loboda, A. V.; Thomson, B. A. *Journal of Mass Spectrometry* **2001**, *36*, 849-865.

47. Yost, R. A.; Boyd, R. K. *Methods of Enzymology* **1990**, *193*, 154-200.

48. Aebersold, R.; Mann, M. *Nature* **2003**, *422*, 198-207.

49. Saravanan, R. S.; Rose, J. K. *Proteomics* **2004**, *4*, 2522-2532.

50. Lauber, W. M.; Carroll, J. A.; Dufield, D. R.; Kiesel, J. R.; Radabaugh, M. R.; Malone, J. P. *Electrophoresis* **2001**, *22*, 906-918.

51. Yan, J. X.; Harry, R. A.; Spibey, C.; Dunn, M. J. *Electrophoresis* **2000**, *21*, 3657-3665.

52. Miller, M. D., Jr.; Acey, R. A.; Lee, L. Y.; Edwards, A. J. *Electrophoresis* **2001**, *22*, 791-800.

53. Patton, W. F. *Electrophoresis* **2000**, *21*, 1123-1144.

54. Wilm, M.; Shevchenko, A.; Houthaeve, T.; Breit, S.; Schweigerer, L.; Fotsis, T.; Mann, M. *Nature* **1996**, *379*, 466-469.

55. Hunt, D. F.; Yates, J. R., 3rd; Shabanowitz, J.; Winston, S.; Hauer, C. R. *Proceedings of the National Academy of Science of the United States of America* **1986**, *83*, 6233-6237.

56. Yates, J. R., 3rd; Eng, J. K.; McCormack, A. L.; Schieltz, D. *Analytical Chemistry* **1995**, *67*, 1426-1436.

57. Yates, J. R., 3rd; Eng, J. K.; McCormack, A. L. *Analytical Chemistry* **1995**, *67*, 3202-3210.

58. Yates, J. R., 3rd; Morgan, S. F.; Gatlin, C. L.; Griffin, P. R.; Eng, J. K. *Analytical Chemistry* **1998**, *70*, 3557-3565.

59. Mann, M.; Wilm, M. *Analytical Chemistry* **1994**, *66*, 4390-4399.

60. Perkins, D. N.; Pappin, D. J.; Creasy, D. M.; Cottrell, J. S. *Electrophoresis* **1999**, *20*, 3551-3567.

61. McLachlin, D.; Chait, B. *Current Opinion in Chemical Biology* **2001**, *5*, 591-602.

62. Ficarro, S.; McCleland, M.; Stukenberg, P.; Burke, D.; Ross, M.; Shabanowitz, J.; Hunt, D.; White, F. *Nature Biotechnology* **2002**, *20*, 301-305.

63. Raska, C.; Parker, C.; Dominski, Z.; Marzluff, W.; Glish, G.; Pope, R.; Borchers, C. *Analytical Chemistry* **2002**, *74*, 3429-3433.

64. Stensballe, A.; Andersen, S.; Jensen, O. *Proteomics* **2001**, *1*, 207-222.

65. Cao, P.; Stults, J. *Rapid Communications in Mass Spectrometry* **2000**, *14*, 1600-1606.

66. Zhou, W.; Merrick, A.; Khaledi, M.; Tomer, K. *Journal of the American Society for Mass Spectrometry* **2000**, *11*, 273-282.

67. Posewitz, M.; Tempst, P. *Analytical Chemistry* **1999**, *71*, 2883-2892.

68. Knight, Z.; Schilling, B.; Row, R.; Kenski, D.; Gibson, B.; Shokat, K. *Nature Biotechnology* **2003**, *21*, 1047-1054.

69. Oda, Y.; Nagasu, T.; Chait, B. *Nature Biotechnology* **2001**, *19*, 379-382.

70. McLachlin, D.; Chait, B. *Analytical Chemistry* **2003**, *75*, 6826-6836.

71. Adamczyk, M.; Gebler, J.; Wu, J. *Rapid Communications in Mass Spectrometry* **2001**, *15*, 1481-1488.

72. Goshe, M.; Veenstra, T.; Panisko, E.; Conrads, T.; Angell, N.; Smith, R. *Analytical Chemistry* **2002**, *74*, 607-616.

73. Molloy, M.; Andrews, P. *Analytical Chemistry* **2001**, *73*, 5387-5394.

74. Andersson, L.; Porath, *Journal of Analytical Biochemistry* **1986**, *154*, 250-254.

75. Neville, D. C.; Rozanas, C. R.; Price, E. M.; Gruis, D. B.; Verkman, A. S.; Townsend, R. R. *Protein Science* **1997**, *6*, 2436-2445.

76. Andersson, L. *Journal of Chromatography A* **1991**, *539*, 327-334.

77. Goshe, M.; Conrads, T.; Panisko, E.; Angell, N.; Veenstra, T.; Smith, R. *Analytical Chemistry* **2001**, *73*, 2578-2586.

78. Asara, J. M.; Allison, J. *Journal of the American Society for Mass Spectrometry* **1999**, *10*, 35-44.

79. Rutschmann, F.; Stalder, U.; Piotrowski, M.; Oecking, C.; Schaller, A. *Plant Physiology* **2002**, *129*, 156-168.

80. Hegeman, A. D.; Harms, A. C.; Sussman, M. R.; Bunner, A. E.; Harper, J. F. *Journal of the American Societry for Mass Spectrometry* **2004**, *15*, 647-653.

81. Hakes, D.; Dixon, J. *Analytical Biochemistry* **1992**, *202*, 293-298.

82. Gorg, A.; Obermaier, C.; Boguth, G.; Harder, A.; Scheibe, B.; Wildgruber, R.; Weiss, W. *Electrophoresis* **2000**, *21*, 1037-1053.

83. Lilley, K. S.; Razzaq, A.; Dupree, P. *Current Opinion in Chemical Biology* **2002**, *6*, 46-50.

84. Rabilloud, T. *Proteomics* **2002**, *2*, 3-10.

85. Klose, J. *Humangenetik* **1975**, *26*, 231-243.

86. MacGillivray, A. J.; Rickwood, D. *European Journal of Biochemistry* **1974**, *41*, 181-190.

87. O'Farrell, P. H. *Journal of Biological Chemistry* **1975**, *250*, 4007-4021.

88. Nam, M. H.; Heo, E. J.; Kim, J. Y.; Kim, S. I.; Kwon, K. H.; Seo, J. B.; Kwon, O.; Yoo, J. S.; Park, Y. M. *Proteomics* **2003**, *3*, 2351-2367.

BIOGRAPHICAL SKETCH

Camille Strachan was born in St. Andrew, Jamaica on March 17, 1978. She is the daughter of Marjorie and John Strachan, who also parented her 3 brothers, Wayne, Kenroy and Stephen. Throughout her years at Immaculate Conception High School, she aspired to be become a doctor and consequently took 4 ½ years of chemistry, biology and mathematics in order to achieve this dream. Encouraging her along this path was her chemistry teacher, Ms. Bowen who always tried to convince Camille that electrons and atoms do exist even though they can't be seen! Despite her not being convinced at the time, she enjoyed the excitement of chemistry labs when she could see the wonderful colors of compounds formed in Inorganic Lab or the changes in color of the Bunsen burner flame when testing various metals. Also sharing words of encouragement was Sister Helen Rose who taught her Plant Biology. Even though she enjoyed learning about plants, Camille never thought she would ever use this information again. Little did she know how useful this information would be to her future studies! During this time she also enjoyed extra-curricular activities such as modern dancing and playing the piano, violin and clarinet.

At age 18, she decided to broaden her horizons by attending school in another country so she left the sunny island of Jamaica to attend Florida Atlantic University (FAU) in Boca Raton, FL as an honors student in the 5-year master's program in Physical Therapy. At the beginning of her second semester she began tutoring Chemistry and Algebra in the Office of Multicultural Affairs and remained with this position throughout

her studies at FAU. Unfortunately, during her sophomore year of studies she experienced the sudden loss of her beloved father. During her time at FAU she took several psychology, biology, chemistry, and anatomy and physiology classes. At the end of her junior year, her passion for chemistry took over and Camille decided to become a chemistry major. In May 2000, she graduated CUME LAUDE from FAU with a Bachelor of Arts in Chemistry and minor in Psychology.

In Fall 2000, she enrolled in the Analytical Division of the Chemistry Department at the University of Florida as an Alumni Fellow. During this time she was introduced to mass spectrometry by Dr. Richard Yost and decided to continue her research in this field. Working under the supervision of Dr. James Winefordner, she entered a collaboration project funded by NSF with Dr. Alice Harmon from the Botany Department and Dr. Nancy Denslow from the Protein Core Facility. With the help of both Drs. Harmon and Denslow, she was able to learn many sample handling techniques as well as some of the biology behind the projects, fulfilling her desire to do both chemistry and biology. During her last year of research she was funded by a Dow Chemical Fellowship as well as a supplemental grant funded by NSF. Upon completion of her projects, she received her Doctor of Philosophy degree in chemistry in August 2005. Camille has been awarded a two year postdoctoral fellow position with Centocor, a Johnson & Johnson company.

55. Hunt, D. F.; Yates, J. R., 3rd; Shabanowitz, J.; Winston, S.; Hauer, C. R. *Proceedings of the National Academy of Science of the United States of America* **1986**, *83*, 6233-6237.

56. Yates, J. R., 3rd; Eng, J. K.; McCormack, A. L.; Schieltz, D. *Analytical Chemistry* **1995**, *67*, 1426-1436.

57. Yates, J. R., 3rd; Eng, J. K.; McCormack, A. L. *Analytical Chemistry* **1995**, *67*, 3202-3210.

58. Yates, J. R., 3rd; Morgan, S. F.; Gatlin, C. L.; Griffin, P. R.; Eng, J. K. *Analytical Chemistry* **1998**, *70*, 3557-3565.

59. Mann, M.; Wilm, M. *Analytical Chemistry* **1994**, *66*, 4390-4399.

60. Perkins, D. N.; Pappin, D. J.; Creasy, D. M.; Cottrell, J. S. *Electrophoresis* **1999**, *20*, 3551-3567.

61. McLachlin, D.; Chait, B. *Current Opinion in Chemical Biology* **2001**, *5*, 591-602.

62. Ficarro, S.; McCleland, M.; Stukenberg, P.; Burke, D.; Ross, M.; Shabanowitz, J.; Hunt, D.; White, F. *Nature Biotechnology* **2002**, *20*, 301-305.

63. Raska, C.; Parker, C.; Dominski, Z.; Marzluff, W.; Glish, G.; Pope, R.; Borchers, C. *Analytical Chemistry* **2002**, *74*, 3429-3433.

64. Stensballe, A.; Andersen, S.; Jensen, O. *Proteomics* **2001**, *1*, 207-222.

65. Cao, P.; Stults, J. *Rapid Communications in Mass Spectrometry* **2000**, *14*, 1600-1606.

66. Zhou, W.; Merrick, A.; Khaledi, M.; Tomer, K. *Journal of the American Society for Mass Spectrometry* **2000**, *11*, 273-282.

67. Posewitz, M.; Tempst, P. *Analytical Chemistry* **1999**, *71*, 2883-2892.

68. Knight, Z.; Schilling, B.; Row, R.; Kenski, D.; Gibson, B.; Shokat, K. *Nature Biotechnology* **2003**, *21*, 1047-1054.

69. Oda, Y.; Nagasu, T.; Chait, B. *Nature Biotechnology* **2001**, *19*, 379-382.

70. McLachlin, D.; Chait, B. *Analytical Chemistry* **2003**, *75*, 6826-6836.

71. Adamczyk, M.; Gebler, J.; Wu, J. *Rapid Communications in Mass Spectrometry* **2001**, *15*, 1481-1488.

72. Goshe, M.; Veenstra, T.; Panisko, E.; Conrads, T.; Angell, N.; Smith, R. *Analytical Chemistry* **2002**, *74*, 607-616.

73. Molloy, M.; Andrews, P. *Analytical Chemistry* **2001**, *73*, 5387-5394.

74. Andersson, L.; Porath, *Journal of Analytical Biochemistry* **1986**, *154*, 250-254.

75. Neville, D. C.; Rozanas, C. R.; Price, E. M.; Gruis, D. B.; Verkman, A. S.; Townsend, R. R. *Protein Science* **1997**, *6*, 2436-2445.

76. Andersson, L. *Journal of Chromatography A* **1991**, *539*, 327-334.

77. Goshe, M.; Conrads, T.; Panisko, E.; Angell, N.; Veenstra, T.; Smith, R. *Analytical Chemistry* **2001**, *73*, 2578-2586.

78. Asara, J. M.; Allison, J. *Journal of the American Society for Mass Spectrometry* **1999**, *10*, 35-44.

79. Rutschmann, F.; Stalder, U.; Piotrowski, M.; Oecking, C.; Schaller, A. *Plant Physiology* **2002**, *129*, 156-168.

80. Hegeman, A. D.; Harms, A. C.; Sussman, M. R.; Bunner, A. E.; Harper, J. F. *Journal of the American Societry for Mass Spectrometry* **2004**, *15*, 647-653.

81. Hakes, D.; Dixon, J. *Analytical Biochemistry* **1992**, *202*, 293-298.

82. Gorg, A.; Obermaier, C.; Boguth, G.; Harder, A.; Scheibe, B.; Wildgruber, R.; Weiss, W. *Electrophoresis* **2000**, *21*, 1037-1053.

83. Lilley, K. S.; Razzaq, A.; Dupree, P. *Current Opinion in Chemical Biology* **2002**, *6*, 46-50.

84. Rabilloud, T. *Proteomics* **2002**, *2*, 3-10.

85. Klose, J. *Humangenetik* **1975**, *26*, 231-243.

86. MacGillivray, A. J.; Rickwood, D. *European Journal of Biochemistry* **1974**, *41*, 181-190.

87. O'Farrell, P. H. *Journal of Biological Chemistry* **1975**, *250*, 4007-4021.

88. Nam, M. H.; Heo, E. J.; Kim, J. Y.; Kim, S. I.; Kwon, K. H.; Seo, J. B.; Kwon, O.; Yoo, J. S.; Park, Y. M. *Proteomics* **2003**, *3*, 2351-2367.

BIOGRAPHICAL SKETCH

Camille Strachan was born in St. Andrew, Jamaica on March 17, 1978. She is the daughter of Marjorie and John Strachan, who also parented her 3 brothers, Wayne, Kenroy and Stephen. Throughout her years at Immaculate Conception High School, she aspired to be become a doctor and consequently took 4 ½ years of chemistry, biology and mathematics in order to achieve this dream. Encouraging her along this path was her chemistry teacher, Ms. Bowen who always tried to convince Camille that electrons and atoms do exist even though they can't be seen! Despite her not being convinced at the time, she enjoyed the excitement of chemistry labs when she could see the wonderful colors of compounds formed in Inorganic Lab or the changes in color of the Bunsen burner flame when testing various metals. Also sharing words of encouragement was Sister Helen Rose who taught her Plant Biology. Even though she enjoyed learning about plants, Camille never thought she would ever use this information again. Little did she know how useful this information would be to her future studies! During this time she also enjoyed extra-curricular activities such as modern dancing and playing the piano, violin and clarinet.

At age 18, she decided to broaden her horizons by attending school in another country so she left the sunny island of Jamaica to attend Florida Atlantic University (FAU) in Boca Raton, FL as an honors student in the 5-year master's program in Physical Therapy. At the beginning of her second semester she began tutoring Chemistry and Algebra in the Office of Multicultural Affairs and remained with this position throughout

her studies at FAU. Unfortunately, during her sophomore year of studies she experienced the sudden loss of her beloved father. During her time at FAU she took several psychology, biology, chemistry, and anatomy and physiology classes. At the end of her junior year, her passion for chemistry took over and Camille decided to become a chemistry major. In May 2000, she graduated CUME LAUDE from FAU with a Bachelor of Arts in Chemistry and minor in Psychology.

In Fall 2000, she enrolled in the Analytical Division of the Chemistry Department at the University of Florida as an Alumni Fellow. During this time she was introduced to mass spectrometry by Dr. Richard Yost and decided to continue her research in this field. Working under the supervision of Dr. James Winefordner, she entered a collaboration project funded by NSF with Dr. Alice Harmon from the Botany Department and Dr. Nancy Denslow from the Protein Core Facility. With the help of both Drs. Harmon and Denslow, she was able to learn many sample handling techniques as well as some of the biology behind the projects, fulfilling her desire to do both chemistry and biology. During her last year of research she was funded by a Dow Chemical Fellowship as well as a supplemental grant funded by NSF. Upon completion of her projects, she received her Doctor of Philosophy degree in chemistry in August 2005. Camille has been awarded a two year postdoctoral fellow position with Centocor, a Johnson & Johnson company.