PROTEOMIC ANALYSIS OF THE FLAVONOID BIOSYNTHETIC MACHINERY IN *ARABIDOPSIS THALIANA*

Nileshwari N. Vaghela

Thesis submitted to the faculty of the Virginia Polytechnic Institute and State University in the partial fulfillment of the requirement for
the degree of

Master of Science
in
Biological Sciences

Brenda S.J. Winkel, Ph.D., Committee Co-chair
Iuliana M. Lazar, Ph.D., Committee Co-chair
Richard F. Helm, Ph.D., Committee Member
Richard A. Walker, Ph.D., Committee Member

August 28, 2007
Blacksburg, Virginia

Keywords: Protein-protein interactions, Enzyme complexes, Affinity chromatography, Flavonoid pathway, Mass spectrometry, Proteomics

PROTEOMIC ANALYSIS OF FLAVONOID BIOSYNTHETIC MACHINERY IN *ARABIDOPSIS THALIANA*

Nileshwari N. Vaghela

Abstract

Work on a wide variety of metabolic pathways indicates that these systems are often, if not always, organized as multienzyme complexes. Enzyme complexes have the potential to increase catalytic efficiency and provide unique mechanisms for the regulation of cellular metabolism. The flavonoid biosynthetic pathway of *Arabidopsis* is an excellent model for studying the organization, localization, and regulation of enzyme complexes at the cellular level. Flavonoids are specialized metabolites that perform many important physiological roles in plants. Protein interactions among several key flavonoid enzymes have been described. Moreover, at least two of the flavonoid enzymes have a dual cytoplasmic/nuclear localization. These results indicate that flavonoid enzymes assemble into one or more distinct complexes at different intracellular locations.

The current study integrates a new technology, mass spectrometry, with well-established affinity chromatography methods to further characterize the organization and composition of the *Arabidopsis* flavonoid enzyme complex. One of the key flavonoid enzymes, chalcone isomerase (CHI), was used in these experiments to detect interacting enzymes. Recombinant thioredoxin (TRX), or TRX-CHI, was produced in *E. coli*, then purified by metal affinity chromatography, and covalently coupled to an activated resin, Affi-Gel 10. Extracts prepared from 4-day-old wild type or CHI-deficient lines of *Arabidopsis* were then passed over the column and the bound proteins eluted with sodium dedocyl sulfate (SDS). This eluate was then subjected to a liquid

chromatography (LC) - mass spectrometry (MS) protocol developed for the analysis of complex peptide mixtures. An Agilent LC system coupled with an LTQ-MS instrument (Thermo Electron, San Jose, CA) was used for this purpose. Data analysis was performed with the Bioworks software package. Different optimization strategies were performed to improve the affinity chromatography, sample preparation, and the LC separation method. A novel approach has been developed for the MS analysis of biological samples containing contaminants such as salts and detergents.

Protein extracts prepared from wild type Landsburg and mutant *tt5*(86) were analyzed by LC-MS/MS. A total of 491 proteins were identified for Landsburg and 633 for *tt5*(86) extracts using a combination of data filters and p-value sorting. All detected proteins had $p<0.001$ and most were identified by at least 2 unique peptides.

ACKNOWLEDGMENTS

I wish to express my deepest gratitude to my mentors, Dr. Brenda S.J. Winkel and Dr. Iuliana M. Lazar, for giving me the opportunity to be part of the fascinating world of science and molecular biology and proteomics research. I thank them for encouraging me in many aspects of my research project, from challenging me in developing ideas, planning and executing experiments, to interpreting and presenting results.

I would like to extend my thanks to my committee members, Dr. Richard F. Helm and Dr. Richard A. Walker for their time, contributions, and recommendations. I appreciate their willingness to serve on my research committee.

My sincere thanks to my fellow lab mates in the Winkel laboratory, Melissa Ramirez, Jonathan Watkinson, Peter Bowerman, and Kevin Crosby, for their cheerful team spirit and support. In addition, I would also like to thank Hetal Sarvaiya and Dr. Jung Hae Yoon for their help with mass spectrometric research.

My deepest gratitude belongs to my family for being endlessly supportive and caring. My sister-in-law, Nirali for helping me and motivating during my graduate career. I would like to thank my husband, Nishant for always believing in me and helping me reach my goals. Without him I would not be where I am today. This work is dedicated to him.

A special thanks to my friend, Dr. Margret S. Rodrigues, for her help in reviewing my thesis and motivating me to reach my goal.

CONTENTS

List of Figures

Chapter 1

Chapter 2

Chapter 3

List of Tables

Chapter 3

Abbreviations

2D-PAGE: two-dimensional polyacrylamide gel electrophoresis

ANR: Anthocyanidin reductase

ANS: Anthocyanidin synthase

ATP: Adenosine triphosphate

$CaCl_2$: Calcium chloride

CHAPS: 3-[3-(Cholamidopropyl)dimethylammonio]-1-proanesulfonate

CHI: Chalcone isomerase

CHS: Chalcone synthase

CI: Chemical ionization

DFR: Dihydroflavonol reductase

DNA: Deoxyribonucleic acid

ESI: Electrospray ionization

F3H: Flavanone-3-hydroxylase

F3'H: Flavonoid 3'-hydroxylase

FAB: Fast atom bombardment

FLS: Flavonol synthase

FTICR: Fourier transform ion cyclotron resonance

HEPES: 4-(2-hydroxyethyl)-1-piperazine ethanesulfonic acid

HPLC: High performance liquid chromatography

IgG: Immunoglobulin G

ICAT: Isotope-coded affinity tagging

LC: Liquid chromatography

LTQ: Linear quadrupole ion trap instrument

MALDI: Matrix assisted laser desorption ionization

MOPS: 4-Morpholinepropanesulfonic acid

MS: Mass spectrometry

MOPS: 4-Morpholinepropanesulfonic acid

m/z: mass to charge ratio

RPLC: Reverse phase liquid chromatography

SDS-PAGE: Sodium dodecylsulfate polyacrylamide gel electrophoresis

TAP: Tandem affinity purification

TCA: Tricarboxylic acid

tds: tannin deficient seeds

TRX: Thioredoxin

tt: Transparent testa

TOF: Time of flight

Chapter 1
Introduction

Multienzyme Complexes

An enzyme complex is a group of two of more enzymes interacting with each other to catalyze related metabolic reactions. This type of organization can offer a number of important advantages to metabolic pathways. The formation of enzyme complexes has the potential to increase catalytic efficiency by coordinating communication among the participating enzymes (Orosz et al., 2000). Enzyme complexes can also help prevent the escape of intermediates from the pathway (Spivey and Merz, 1989; Welch, 1977). Work on enzymes of the Krebs TCA cycle, glycolysis, fatty acid oxidation, and other metabolic systems shows that enzyme complexes help to maintain high local substrate concentrations, separate anabolic and catabolic reactions, and maintain stereospecificity (Mathews, 1993; Ovádi and Srere, 1996). However, much remains to be learned about the composition and assembly of most of these systems.

The organization of metabolic pathways as multienzyme complexes has been an important area of study in cellular metabolism. Some of the first experiments providing evidence supporting this idea were described almost 50 years ago (Kempner and Miller, 1968; Zalokar, 1960). These experiments showed that the cytosol contains very little if any free protein that is not associated with membranes, organelles, or the cytoskeleton. Since then, a variety of experimental approaches including co-immunoprecipitation (Halper and Srere, 1977), fluorescence anisotropy (Tompa et al., 1987), affinity chromatography (Persson and Srere, 1992), electrophoretic techniques (Beeckmans et al., 1989), and countercurrent distribution procedures (Backman and Johansson, 1976) have provided evidence suggesting that many, if not all, metabolic pathways exist as highly organized enzyme complexes.

The Flavonoid Biosynthetic Pathway of *Arabidopsis thaliana*

The flavonoid pathway of plants is a powerful model for studying metabolic organization at the cellular level. Flavonoids are secondary (or specialized) metabolites derived from phenylalanine and acetyl CoA (Fig.1.1) that have roles in plant reproduction, growth, and survival. Flavonoids are indicators of maturation and dormancy in seeds (Debeaujon et al.,

2003). The compounds are responsible for pigmentation of flowers and other plant tissues that are used to attract pollinators (Weiss, 1991) . Efforts have been made to engineer flower color via this pathway (Meyer et al., 1987; Morita et al., 2005). Flavonoids also have nutritional and medicinal properties. Flavonoids are important antioxidant and anticancer components of the human diet (Havsteen, 2002; Rice-Evans, 2001; Vom Endt et al., 2002) and efforts have been made to metabolically engineer vegetable crops such as soybean and tomato to produce higher amounts of these compounds (Muir et al., 2001; Yu et al., 2003).

Genes encoding all seven enzymes of the central flavonoid pathway have been cloned in *Arabidopsis thaliana:* chalcone synthase (CHS), chalcone isomerase (CHI), flavanone-3-hydroxylase (F3H), flavonoid 3'-hydroxylase (F3'H), flavonol synthase (FLS), dihydroflavonol reductase (DFR), anthocyanidin synthase (ANS), and anthocyanidin reductase (ANR) (Fig. 1.2) (Burbulis et al., 1996; Pelletier et al., 1999; Pelletier et al., 1997; Pelletier and Shirley, 1996; Shirley et al., 1992). These enzymes are all encoded by single-copy genes, except FLS, although even in this case it appears that only one isoform has catalytic activity (Owens et al., *in preparation*) containing mutations in all of the single-copy genes have been identified based on the resulting partial or complete loss of pigments from the coat (testa) of seeds and are known as *transparent testa (tt)* or *tannin-deficient seeds* (*tds)* (Abrahams et al., 2003; Feinbaum and Ausubel, 1988; Schoenbohm et al., 2000; Shirley et al., 1992) (Fig. 1.2). T-DNA and Tn-induced knockout lines have been identified for many of these genes, including all six members of the FLS gene family (Gälweiler et al., 1998).

The concept that the flavonoid pathway might be organized as an enzyme complex was first proposed by Helen Stafford (1981). Biosynthesis of flavonoids appears to occur in the cytoplasm via a complex that is assembled at the endoplasmic reticulum and/or in as-yet-uncharacterized electron-dense particles (Burbulis and Winkel-Shirley, 1999; Hrazdina and Jensen, 1992; reviewed in Winkel, 2004). Recent studies done in the Winkel lab show that at least two of the flavonoid enzymes have dual cytoplasmic/nuclear localization (Saslowsky et al., 2005). Specific interactions between flavonoid enzymes have been detected in a variety of ways including two-hybrid assays and surface plasmon resonance (Burbulis and Winkel-Shirley, 1999; Dana et al., *submitted*). The results of affinity chromatography and coimmunoprecipitation have

provided evidence that these associations also occur in plant cells. Recently, a model for the CHS-CHI bienzyme complex was generated based on small angle neutron scattering experiments (Dana et al., *submitted*).

The proposed study attempts to take previous work done in the Winkel laboratory to the next level by using new technologies to probe the entire spectrum of proteins that may comprise the flavonoid enzyme complex. In this study, the affinity chromatography will be used to enrich for the proteins that specifically interact with flavonoid enzymes and these will then be identified by mass spectrometry.

Figure 1.1 Phenotypic changes induced by transparent testa *(tt)* mutation. Change in *Arabidopsis* seed coat pigmentation in *transparent testa (tt)* lines due to mutations in specific flavonoid enzyme genes. Some of the examples of these phenotypes include *tt4* and *tt5* mutants, which produce yellow seeds, while *tt9* and *tt10* mutants produce pale brown seeds.

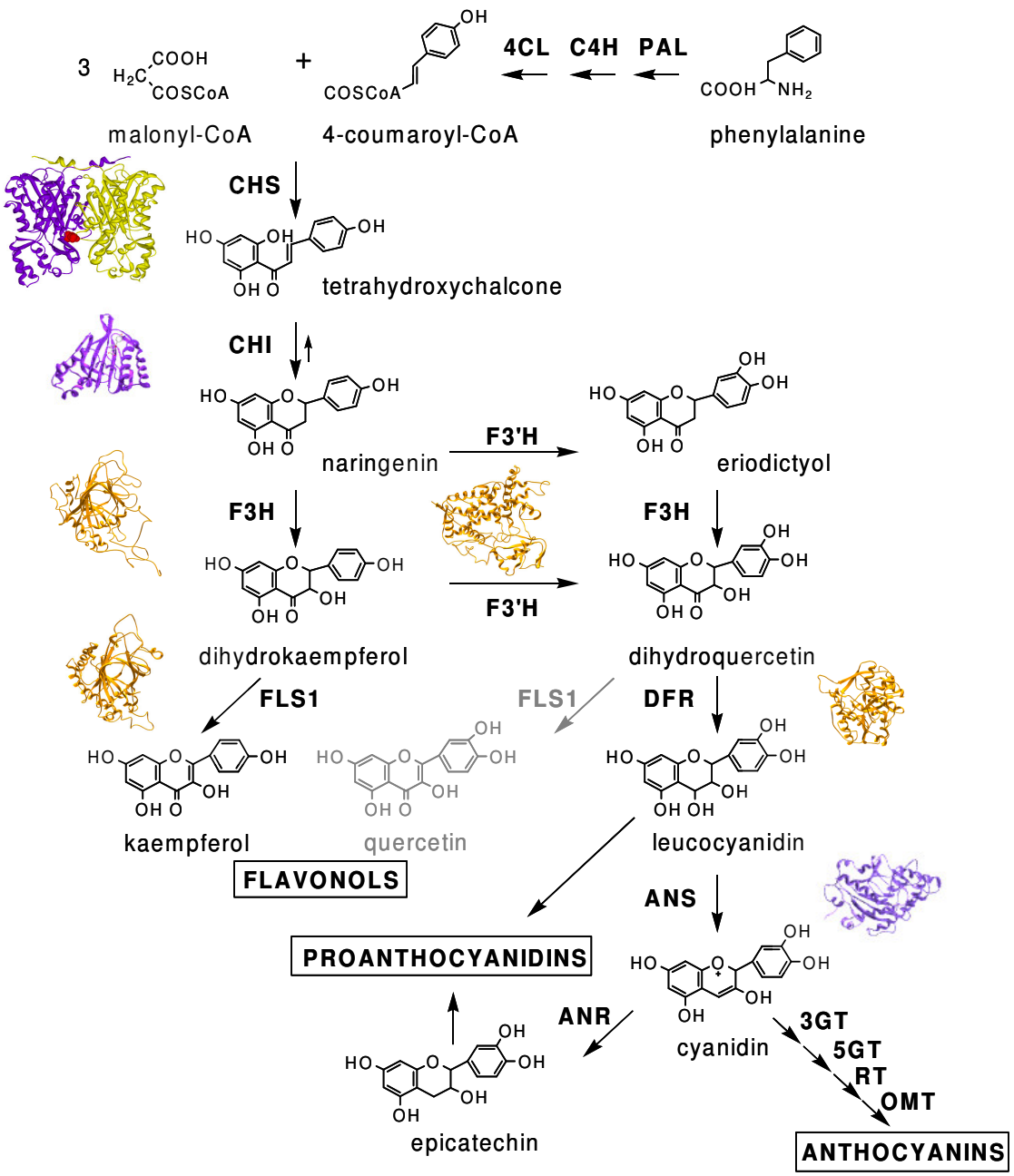


Figure 1.2. Schematic of the flavonoid pathway of *Arabidopsis*. Structures represent the three enzymes in the pathway for which crystal structures have been solved, CHS and CHI (both from *Medicago sativa*), and anthocyanidin synthase (ANS) (from *Arabidopsis*). Structures in orange are homology models of dihydroflavonol 4-reductase (DFR), flavanone 3-hydroxylase (F3H), flavonoid 3' hydroxylase (F3'H), and flavonol synthase (FLS1) (Dana et al., *submitted;* Owens et al., *in preparaion*) No structure is yet available for anthocyanidin reductase (ANR). The step showed in gray is the minor reaction catalyzed by FLS1 (Owens et al., *in preparaion*).

Mass Spectrometry

Mass spectrometry is the most versatile and comprehensive analytical technique currently available for molecular mass determination, structure elucidation, quantification of complex samples at trace levels, and mixture analysis (Larsen and Roepstorff, 2000). A brief introduction of fundamental mass spectrometry concepts is given here.

Components of a mass spectrometer. The mass spectrometer is a detector used for analyzing ions in the gas phase. The instrument consists of an ion source, ion optic elements to transfer the ions from the source to the analyzer, a mass analyzer to measure the mass-to-charge ratio (m/z) of the ionized components, and a detector to count the number of ions at each m/z value (Gross, 2007). A schematic of the main processes occurring within a mass spectrometer is given in Fig. 1.3.

Ion source. This component of the mass spectrometer produces ions by ionizing neutral molecules using various mechanisms such as electron ejection, electron capture, proton transfer, and charge exchange. Chemical ionization (CI), fast atom bombardment (FAB), electrospray ionization (ESI), and matrix assisted laser desorption ionization (MALDI) are the most common ionization techniques used for the analysis of biological molecules. In recent years, a variety of technical developments have completely changed the field of mass spectrometry by making possible the analysis of new classes of compounds. Chemical ionization made it possible to obtain the molecular weights of low mass, volatile organic compounds. Fast atom bombardment was one of the most rapid and convenient methods for analyzing thermally labile compounds like peptides and small proteins. Electrospray ionization represents an important development that made possible the analysis of high molecular weight components such as proteins (Jebanathirajah and Coleman, 1998). Matrix-assisted laser desorption ionization is one of the latest development that enables the fast detection and structural characterization of peptides and proteins (Karas and Hillenkamp, 1988).

Mass analyzers. There are four main types of mass analyzers currently in use: ion trap, time-of-flight (TOF), quadrupole, and Fourier transform ion cyclotron resonance (FTICR). Each

analyzer is different in design and performance, with its own strengths and weaknesses. For proteomics research, the essential qualities of an analyzer are sensitivity, resolution, mass accuracy, and the ability to generate informative ion mass spectra for peptide fragments. Depending upon the application, different combinations of an ion source and mass analyzer can be used.

The most commonly-used mass analyzers in contemporary proteomics research are ion trap analyzers because they are sensitive, robust, and relatively inexpensive. The disadvantage of ion traps is that they provide relatively low mass accuracy. On the other hand, ion traps are the only mass analyzers that enable multiple steps of ion fragmentation.

Separation systems. To enhance the capacity of the mass spectrometer to analyze complex mixtures, various analytical separation systems are interfaced with mass spectrometry detectors. High performance liquid chromatography (HPLC) is most commonly used for the analysis of peptide/protein components. In HPLC, the analytes are forced through a column filled with a stationary phase by a liquid/mobile phase at high pressure; the differential interaction between the sample components and the stationary phase results in the separation of the components as they are eluting from the column. Another separation technique widely used for protein analysis is two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) (Chuong et al., 2004). In 2D-PAGE, proteins are first separated based on their isoelectric point and then by their molecular weight. Spots on the gel are digested with trypsin, extracted from the gel and then analyzed either directly by mass spectrometry or by HPLC-MS. Recently, new techniques such as isotope-coded affinity tagging (ICAT) have been developed for protein quantitation.

Tandem mass spectrometry. Peptide amino acid sequence information can be obtained with a mass spectrometer by fragmenting the ions of interest and determining the fragmentation pattern using a technique called MS/MS or tandem MS. Selected ions are trapped/isolated in the analyzer for a fixed time. The trapped ions are then accelerated by electrical fields and fragmented by collision with noble gases present in side the trap. The fragmented ions are then sequentially ejected from the trap for mass analysis and sequence determination.

The peptide dissociation pathway depends on the collision energy. In most mass spectrometry applications, low-energy CID is used (~0.1eV). At low collision energy, fragmentation occurs mainly along the peptide backbone bonds, while at high collision energy fragmentation may occur at the side-chains of amino acids, as well (Johnson et al., 1987). Fragmentation patterns also depend on the chemical and physical properties of the amino acids and their sequence in the peptide.

Data analysis. For the identification of protein components in a sample, the experimental tandem mass spectra (MS/MS) are compared to the theoretical mass spectra that are computer-generated from a protein sequence database (Fig. 1.4) (Sadygov et al., 2004). Commercially-available software tools identify the top protein matches to a specific peptide sequence. Some of the most widely-used programs are SEQUEST, MASCOT, and X! Tandem. For the research described in this thesis, the sequence assignment and database searches were performed by the SEQUEST algorithm that included in the Bioworks software package provided with the LTQ mass spectrometer by Thermo Electron (San Jose, CA).

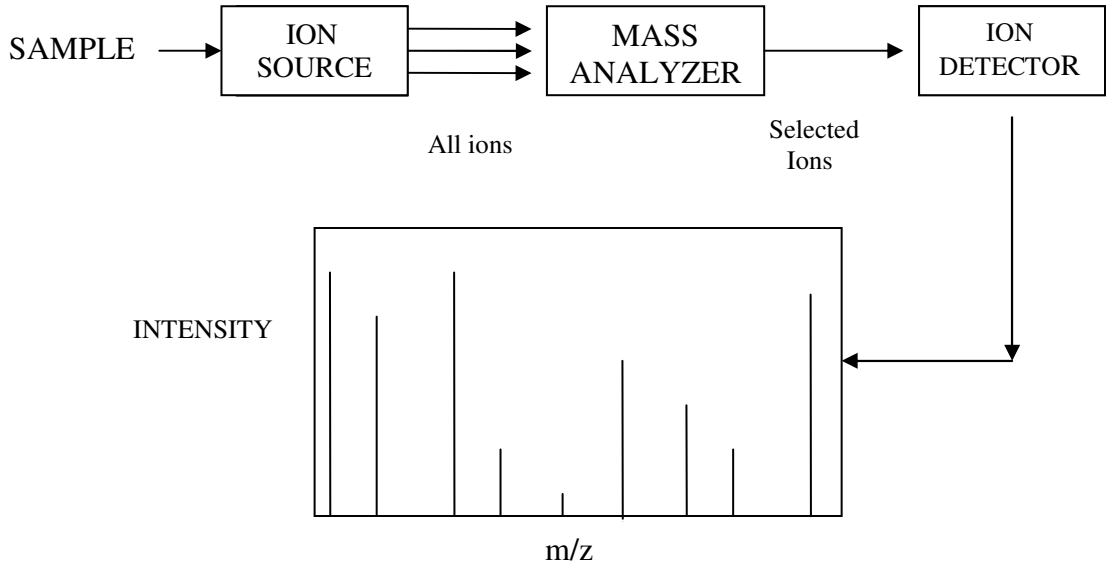


Figure 1.3. Schematic representation of the main processes occurring within a mass spectrometer.

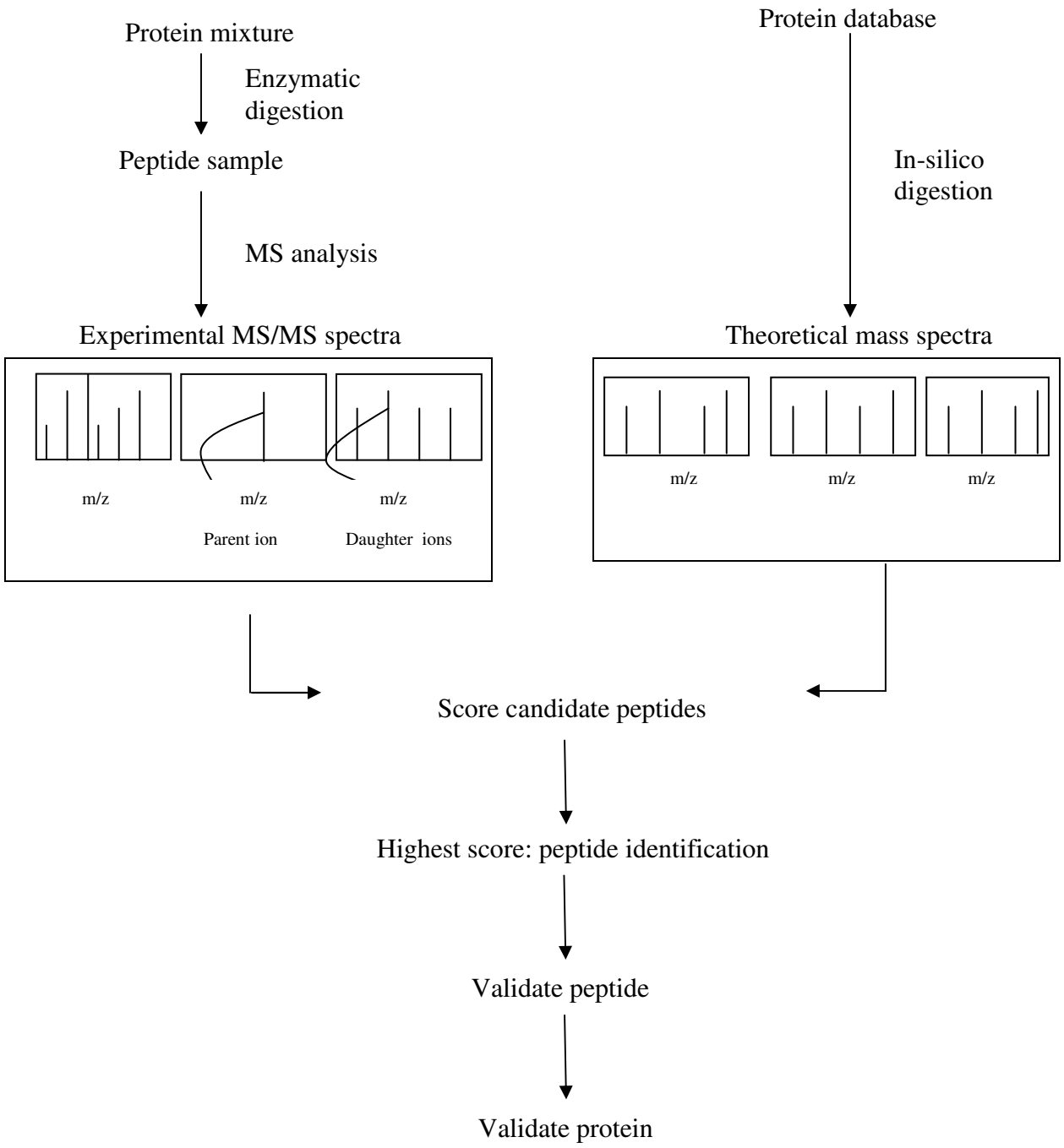


Figure 1.4. Outline of the protein identification process. Protein mixtures are digested with trypsin. The resulting peptides are analyzed by MS/MS to obtain experimental sequence information. Experimental data are compared with theoretical data generated by programs like SEQUEST and MASCOT. The best peptide matches identify a specific protein. Validation software confirms whether the peptide and protein identifications are true or false.

Protein Profiling by Mass Spectrometry

Mass spectrometry is a powerful tool for protein profiling. With advances in genomics information and the development of new analytical tools, proteomics is also becoming an important tool for the study of many aspects of plant physiology and biochemistry. To date, most large-scale analyses have been carried out using the model plant, *Arabidopsis*. These analyses have been performed with isolated organelles, membrane systems, and subcellular structures (Peck, 2005). Studies include protein profiling of the chloroplast (Brugiere et al., 2004; Friso et al., 2004; Froehlich et al., 2003; Huber et al., 2004; Kleffmann et al., 2004; Peltier et al., 2002; Peltier et al., 2004), mitochondria (Brugiere et al., 2004; Lister et al., 2004; Millar et al., 2005), peroxisomes (Fukao et al., 2002), vacuoles (Carter et al., 2004; Shimaoka et al., 2004), the plasma membrane (Alexandersson et al., 2004; Borner et al., 2005; Ephritikhine et al., 2004; Marmagne et al., 2004; Shimaoka et al., 2004), cell wall (Borderies et al., 2003; Boudart et al., 2005; Chivasa et al., 2002) and cytosolic ribosomes (Li et al., 2005). Studies have also been performed in rice, including a systematic proteomic analysis of rice leaf, root, and seed tissue using two-dimensional gel electrophoresis followed by tandem mass spectrometry and multidimensional protein identification technology, that allowed the detection and identification of 2,528 unique proteins (Koller et al., 2002).

Protein-Protein Interaction and Proteomics

Many cellular functions performed by proteins depend on, or are enhanced by, protein-protein interactions. Enzymes of metabolic pathways that form complexes provide good examples of this. Along with expression patterns, interactions between proteins can provide clues about the function and regulation of individual proteins. Analysis of protein complexes is one of the areas where mass spectrometry-based proteomics is having an impact (Uetz and Hughes, 2000; Werhahn et al., 2001). In this case the protein of interest is often used as an affinity reagent to isolate the proteins that bind to it. An advantage of mass spectrometry-based proteomics over yeast two-hybrid and chip-based approaches is the use of fully processed and modified proteins as a bait to attract interacting proteins (Turnbull et al., 2004). Moreover,

multicomponent complexes can be isolated and analyzed in a single mass spectrometric application. However, mass spectrometry-based methods combined with affinity chromatography tend to detect only a subset of the occurring protein interactions since many biologically-relevant interactions are transient, or require a specific cellular environment in order to take place.

Millar et al. (2005) have used immobilized ATP affinity chromatography to identify the soluble nucleotide binding proteome in *Arabidopsis*. In this study, a range of highly-enriched proteins were identified from the mitochondrial proteome, including proteins known to contain nucleotide binding domains. Similarly, a DNA replication/repair protein network was identified in wheat by proliferating cell nuclear antigen affinity chromatography followed by liquid chromatography and tandem MS (Toueille et al., 2007). Analysis of complexes formed by human topoisomerase I detected 29 new and confirmed seven previously-identified protein partners by affinity chromatography and mass spectrometry (Czubaty et al., 2005).

Affinity-based proteomics is widely used to study protein interactions in a wide variety of organisms. Previous analysis of the yeast mitochondrial proteome by liquid chromatography and mass spectrometry identified 546 proteins (Prokisch et al., 2004). Novel proteins interacting with the cytosolic glucocorticoid receptor were identified by immunoaffinity chromatography and MALDI-TOF-MS (Hedman et al., 2006). The study reported 27 new candidates for the glucocorticoid complex. Similarly, 24,540 potential protein interactions were identified in a large-scale mapping of human protein-protein interactions performed by immunoprecipitation and LC-ESI-MS/MS (Ewing et al., 2007).

Tandem affinity purification (TAP) tagging is another approach developed for purification and isolation of protein complexes (Rigaut et al., 1999). In this method, two different tags are fused to the protein of interest in an expression construct and then the construct is introduced in the host cell or organism (Puig et al., 2001). Cell extracts are prepared and two specific affinity purification steps, typically using IgG beads and calmodulin beads, are performed in tandem to recover the fusion protein and interacting proteins. The recovered material is then analyzed by techniques such as immunoblot analysis or mass spectrometry. A

large scale protein-protein interaction study was recently carried out in yeast using TAP technology (Gavin et al., 2002). In this study, a large number of multiprotein complexes were detected and further analyzed by mass spectrometry. TAP tagging has also been used to isolate and characterize protein complexes that include translation elongation factor in *Plasmodium falciparum* (Takebe et al., 2007). Recently, a TAP-based platform was used to study the cell cycle interactome in *Arabidopsis* (Van Leene et al., 2007).

Proposed Study

Significance. The flavonoid biosynthetic pathway of *Arabidopsis* is an excellent experimental model for the study of enzyme complexes. Interactions between flavonoid pathway enzymes have previously been detected by two-hybrid assays, affinity chromatography, and co-immunoprecipitation (Burbulis and Winkel-Shirley, 1999). Recent evidence shows that at least two of the enzymes, CHS and CHI, have dual cytoplasmic/nuclear localization (Saslowsky et al., 2005). Moreover, a structural model for the CHS-CHI bienzyme complex was recently developed based on small angel neutron scattering (Dana et al., *submitted*). Though major steps of the flavonoid pathway have been well studied, the precise composition of the flavonoid enzyme complex remains unknown. In this study, affinity chromatography will be coupled with mass spectrometry in an effort to detect all possible enzymes or proteins involved in flavonoid metabolism. Crude soluble protein prepared from 5-day-old *Arabidopsis* seedlings, which express high levels of flavonoid enzymes, will be passed over recombinant CHS and CHI protein immobilized on Affi-Gel 10 resin as described previously (Burbulis and Winkel-Shirley, 1999). Affi-Gel 10 is an activated immunoaffinity support that allows for covalent binding of protein. The resin consists of a N-hydroxysuccinimide ester of a derivatized crosslinked agarose gel bead support. The chemical structure of the bead provides a 10-atom spacer arm. Upon addition of ligand, the N-hydroxysuccinimide is displaced and a stable amide bond is formed. The proposed study may thus confirm known interactions among flavonoid enzymes and may also identify novel components of the flavonoid metabolic network.

Specific research objectives. The goal of my research has been to analyze components of the flavonoid enzyme complex by affinity chromatography and mass spectrometry. This work had the following specific objectives:

1. To produce and purify recombinant protein.
2. Perform affinity chromatography experiments.
3. Analyze interacting proteins by ESI mass spectrometry.

Acknowledgments: The project was supported by grants from National Science Foundation (MCB-0445878) to B.S.J.W. and the Virginia Tech Graduate Research Development Program to N.V.

Chapter 2

Proteomic Analysis of the Flavonoid Biosynthetic Machinery of *Arabidopsis thaliana*

SUMMARY

This thesis describes an effort to integrate mass spectrometry with a previously-established affinity chromatography method to extend the analysis of protein complexes associated with the flavonoid biosynthetic pathway of *Arabidopsis thaliana* (Burbulis and Winkel-Shirley, 1999). For the affinity chromatography approach described in this chapter, the flavonoid enzyme, chalcone isomerase (CHI), was expressed in *E. coli* as a fusion to the carboxy terminus of thioredoxin (TRX). TRX was also expressed alone for use in control experiments. The recombinant proteins were used to prepare an affinity column using Affi-Gel 10 resin, to which soluble protein extracted from 4-day-old seedlings of Landsburg and the CHI null mutant, *tt5*(86) was applied. The resulting samples were then used to test the feasibility of using a proteomics approach to identify the full array of proteins recovered by affinity chromatography, as explained in Chapter 3.

Keywords: Protein-protein interactions, Enzyme complexes, Affinity chromatography, Flavonoid Pathway

INTRODUCTION

The flavonoid biosynthetic pathway is a well-studied metabolic system that exists exclusively in plants. Products of the pathway are secondary metabolites, derived from phenylalanine and acetyl CoA (Stafford, 1990), that perform a diverse set of functions. Flavonoids are known for their roles as the major red, blue, and purple pigmentation that are used to attract pollinators and seed dispersers (Weiss, 1991). The products also provide UV protection (Li et al., 1993), and microbial defense (Stafford, 1990). Animals, including humans, consume large quantities of flavonoids in fruits and vegetables. The antioxidant and anticancer properties of flavonoids are advantageous for animal and human health (Harborne and Williams, 2000).

It is believed that many metabolic pathways exist as networks of interacting enzymes that enhance the efficiency of metabolism in diverse ways. For example, the formation of enzyme complexes helps to separate anabolic and catabolic reactions (Hrazdina and Jensen, 1992). This organization allows the accumulation of high local substrate concentrations and the channeling of intermediates (Winkel, 2004). The Winkel laboratory uses the flavonoid biosynthetic pathway of *Arabidopsis thaliana* as a model to understand the assembly of enzyme complexes. Interaction among several flavonoid enzymes, demonstrated by yeast two-hybrid, affinity chromatography, and coimmunoprecipitation, showed that the enzymes of the pathway interact with each other in highly specific manner (Burbulis and Winkel-Shirley, 1999). Also, two key enzymes of the pathway, chalcone synthase (CHS) and CHI, co-localize in both the cytoplasm and nuclei of specific cells at the root tip and in leaves (Saslowsky et al., 2005).

In the proposed research, an effort was made to elucidate the organization of the flavonoid pathway using affinity directed mass spectrometry (Figure 2.1). One of the flavonoid enzymes, CHI, was used as a bait to attract interacting proteins from crude extract prepared from 4-day-old wild type Landsburg and mutant *tt5*(86) seedlings, a CHI gene mutant. An effort was then made to identify the captured proteins using an ESI LC-MS/MS method, as described in Chapter 3.

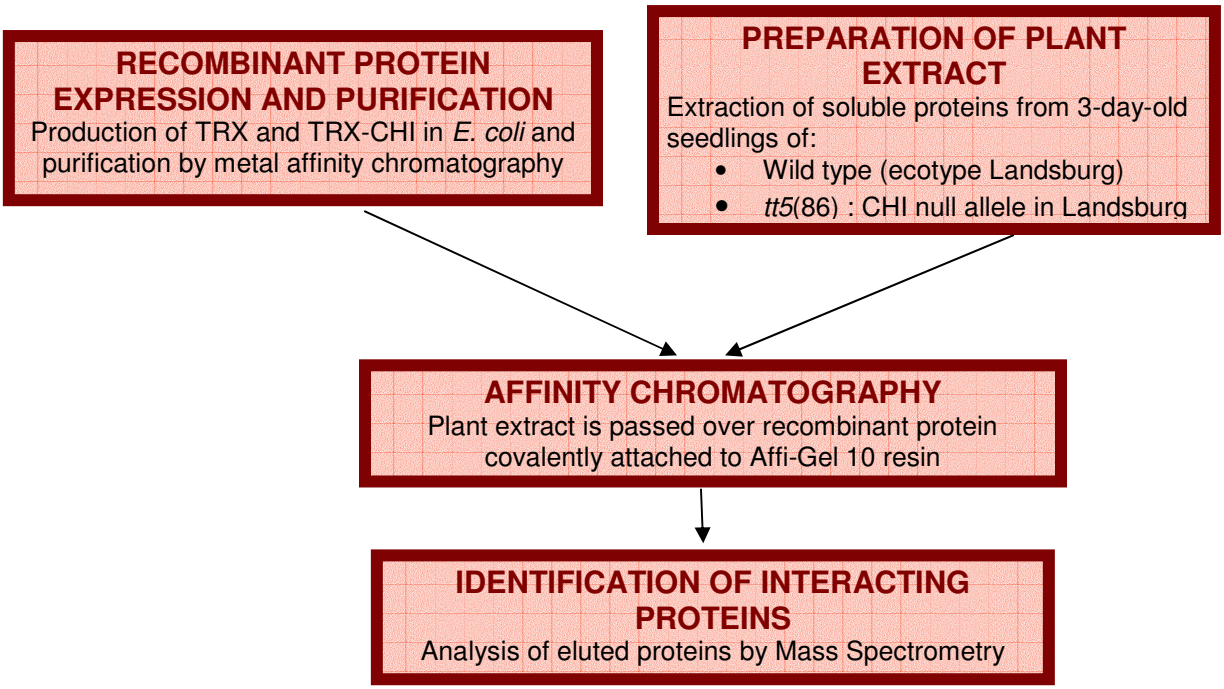


Figure 2.1: Schematic diagram of experimental steps performed for affinity chromatography and mass spectrometry. One of the key enzymes of the flavonoid pathway, CHI, was produced in *E.coli* as a fusion to TRX (TRX-CHI). The enzyme was purified using metal affinity chromatography. Similarly, recombinant TRX was expressed in *E. coli* and purified by metal chromatography for use as a negative control. Soluble protein extracts were prepared from 4-day-old Landsburg and *tt5*(86) *Arabidopsis* seedlings. In affinity chromatography experiments, recombinant TRC-CHI or TRX was used as a bait to attract protein partners present in the Landsburg and *tt5*(86) soluble protein extracts. The interacting partners were analyzed by ESI LC-MS/MS.

Materials and Methods

Recombinant Protein Expression

Recombinant protein was expressed following a previously-published protocol (Burbulis and Winkel-Shirley, 1999). Two key flavonoid enzymes, CHS and CHI, were expressed as fusions to the carboxyl terminus of thioredoxin (TRX) using the pCD1 vector (Dana et al., *submitted*). TRX was also expressed from the pCD1 vector. The plasmids were introduced in *E. coli* strain BL21(DE3) pLysS (Novagen). Protein expression was induced by the addition of dioxane-free isopropyl-β-D-thiogalactoside (Fisher Scientific, New Jersey) to a final concentration of 0.5 mM (1 ml of 0.5 M per liter of culture), followed by incubation at room temperature, 200 rpm for 4 h. The cells were collected by centrifugation at 5000 RPM in a JA-10 rotor at 4°C for 15 min and then were stored at -20°C. The cell pellets were resuspended in 33 ml of lysis buffer (50 mM Hepes, pH 8.0, 150 mM NaCl, 10% glycerol, 1% Tween-20, 10 mM 2-mercaptoethanol) per liter of culture, and then sonicated (6 × 10 sec) on ice. Pellets were collected by centrifugation at 14,000 RPM in a JA-17 rotor at 4°C for 40 min.

Recombinant Protein Purification Using Metal Affinity Chromatography

Recombinant protein was purified from the supernatant using TALON metal affinity resin (BD Biosciences CLONTECH) according to the manufacturer's instructions. The column was prepared using 1 ml of resin, which was equilibrated with 10 bed volumes of lysis buffer. The sample was applied to the column, followed by 10 bed volumes of wash buffer (50 mM HEPES, pH 8.0, 50 mM NaCl, 10 mM Immidazole, 10% glycerol, 10 mM 2-mercaptoethanol). Bound protein was eluted with elution buffer (50 mM Hepes, pH 8.0, 150 mM NaCl, 250 mM imidazol, 10% glycerol, 10 mM 2-mercaptoethanol) using the following volumes: 0.5 ml, 1.5 ml, 1 ml, 1 ml, and 1 ml. Successful expression and purification was verified using sodium dodecylsulfate-polyacrylamide gel electrophoresis (SDS-PAGE). Eluted recombinant protein was dialyzed against 100 mM MOPS, pH 7.5, and 80 mM $CaCl_2$ at 4°C for 4 hr using Slide-A-Lyzer dialysis cassettes (10,000 molecular weight cutoff; Pierce). A Bradford assay (Bio-Rad) was performed to determine the final concentration of the resulting protein preparations.

Plant Material

Arabidopsis seedlings, Landsburg wild type and *tt5*(86) (Shirley et al., 1995), were grown on MS-agar medium as described previously (Kubasek et al., 1992). Flavonoid expression was induced by shifting seedlings to constant white light (100 μE/m^2), 22°C for 4 d. Seedlings were collected, washed with distilled water, and gently blotted dry on Whatman # 1 papers. Seedlings were used fresh, or placed in liquid nitrogen and then stored at -80°C.

Preparation of Affinity Columns

Purified recombinant TRX or TRX-CHI were covalently attached to Affi-Gel 10 activated resin (Bio-Rad, Hercules, CA) in 100 mM HEPES, pH 7.5 and 80 mM $CaCl_2$ for 4 h at 4°C according to the manufacturer's instructions. The coupling reaction was quenched by adding ethanolamine to 100 mM and incubating for 1 hr at 4°C. The column was washed three times with coupling buffer (100 mM MOPS, pH 7.5, 80 mM $CaCl_2$. For each wash, the absorbance at 280 nm was measured to detect the presence of unbound proteins. Based on this criterion, protein was detected in first two washes, but not in the third wash. The column was then washed twice with elution buffer (65 mM Tris, pH 6.8, 2% SDS, 15% glycerol) and stored at 4°C in PBS buffer containing 0.2% sodium azide until used for affinity chromatography experiments. Protein concentrations were detected before and after passage over the column to measure coupling efficiency relative to the manufacturer's stated capacity of 35 mg of protein per ml of resin.

Affinity Chromatography

To prepare soluble plant lysates for affinity chromatography, 4-day-old seedlings were ground to a fine powder in liquid nitrogen as described previously (Burbulis and Winkel-Shirley, 1999). The powder was then suspended in plant lysis buffer (50 mM Tris, pH 7.2, 150 mM NaCl, 1 mM 2-mercaptoethanol, 70 μg/ml DNase I, 0.6% NP-40, 0.6 % 3-[(3-cholamido-propyl) dimethylammonio]-1-propanesulfonate (CHAPS), 1 tablet protease inhibitor cocktail (Roche Diagnostics, IN, USA) at a ratio of approximately 1 ml buffer per gram of powder, and incubated

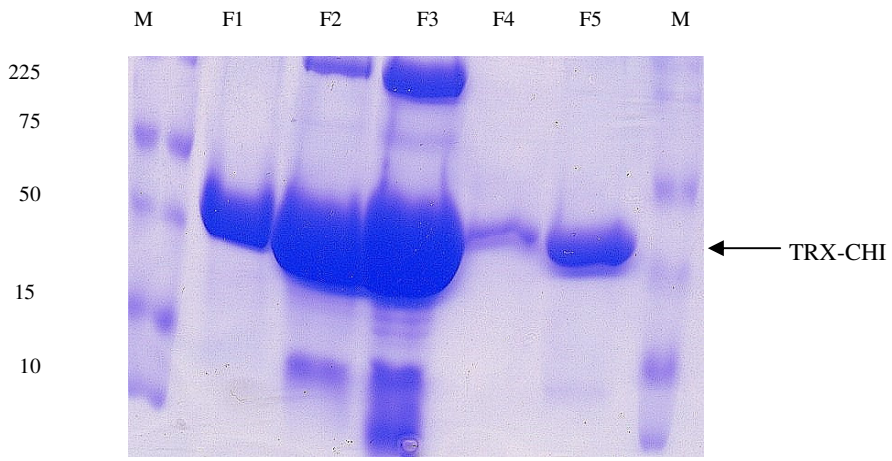
at 25°C for 20 min. The soluble protein fraction was isolated from the insoluble material by centrifugation at 30,000 × g at 20°C for 10 min. For the affinity experiments, 400 µl of soluble plant lysate was mixed with 100 µl of resin coupled with TRX-CHI or TRX and incubated with gentle agitation at 4°C for 2 h. Greater than 90% of binding efficiency was achieved. Approximately 20 nmole protein was bound to resin. After 2 h incubation, the resin was pelleted at 500 ×g for 30 sec and resuspended in 30 volumes of plant lysis buffer without DNase. The resin was pelleted at 500 ×g for 30 sec to remove unbound proteins. Bound proteins were released from the resin in 30 µl of 65 mM Tris pH 6.8, 2 % SDS, 15 % glycerol at 65°C for 5 min and resin was pelleted by centrifugation at 500 × g for 30 sec. The supernatant was collected and analyzed by ESI LC-MS/MS.

RESULTS

Recombinant Protein Expression and Purification

An effort was made to identify proteins that interact with enzymes of the flavonoid pathway by coupling a previously-developed affinity chromatography method (Burbulis and Winkel-Shirley, 1999) with mass spectrometry. One of the flavonoid enzymes, CHI, was used as a bait to attract interacting proteins in the affinity chromatography experiment. TRX was used to prepare an affinity column for use in control experiments. The enzyme was produced in *E. coli,* a convenient and commonly-used expression system for production of recombinant proteins (Baneyx, 1999). It has been shown previously that recombinant CHI produced in *E. coli* is enzymatically active (Cain et al., 1997; Santos et al., 2001). The expressed protein was purified using TALON metal affinity chromatography. TRX was produced using the same method. The purity of the proteins was analyzed by SDS-PAGE (Figure2.2). The yield of purified TRX-CHI and TRX, as determined by the Bio-Rad protein assay, was 10-12 mg/liter of culture and 2-3 mg/liter of culture, respectively.

(A)



(B)

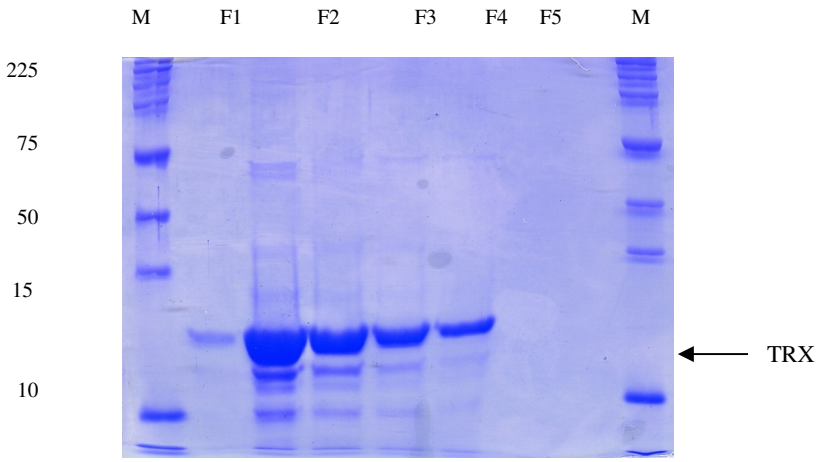


Figure 2.2: SDS-PAGE analysis of purified TRX-CHI (A) and TRX (B). F1 (0.5 ml), F2 (1.5 ml), F3 (1.0 ml), F4 (1.0 ml), and F5 (1.0 ml) represent eluted fractions from metal affinity column. Each lane contains 10 µl of sample. M represents the molecular weight marker, with the sizes of individual marker proteins shown on the left. The predicted sizes of the recombinant proteins are 60 kD for TRX-CHS and 18 kD for TRX.

Affinity Chromatography: Challenges and Optimization

To study protein-protein interactions, affinity chromatography was performed using recombinant flavonoid enzyme and soluble protein extracts prepared from *Arabidopsis* seedlings grown for 4 days under continuous white light in the presence of sucrose, conditions under which flavonoid enzyme accumulation is at high levels. Soluble protein extract prepared from mutant *tt5*(86) was passed over a column consisting of TRX-CHI coupled resin. Protein extract prepared from this mutant line does not contain CHI protein. This *tt5*(86) allele was induced by ionizing radiation, consisted of an inversion within the gene (Shirley et al., 1992). A 272 bp fragment on the same chromosome was transferred to one end of this inversion. Based on affinity experiments done previously, it was known that presence of endogenous CHI protein in wild-type extracts interferes with the interaction between soluble proteins in extract and immobilized TRX-CHI (Burbulis and Winkel-Shirley, 1999). As a control for the specificity of the interactions, experiments using immobilized TRX-CHI and extract from wild type seedlings as well as experiments using resin containing only TRX fusion partner and extracts from both the mutant and wild-type lines were carried out.

A number of experiments were performed to optimize the coupling efficiency. The lysis, wash, and elution buffers for purification of the recombinant proteins were prepared as described previously (Burbulis and Winkel-Shirley, 1999), with the exception that 2-mercaptoethanol was omitted due to concerns that it might interfere with the subsequent coupling procedure, as indicated by the manufacturer of the resin. Moreover, as per the manufacturer's recommendations, the coupling reaction was performed with MOPS (100 mM) and $CaCl_2$ (80 mM) instead of Tris (which can couple to the resin), NaCl, and 2-mercaptoethanol, as in previous experiments. The concentration of CHI was measured before (5-6 mg/ml) and after (3-4 mg/ml) using $O.D._{280.}$ A binding efficiency of approximately 80% was achieved.

As a further optimization procedure, the affinity experiment was performed with different proportions of ligand bound resin and soluble protein extract. To establish a proportion of ligand bound resin and protein extract, various concentrations of ligand bound resin and soluble protein extracts were tested. The initial affinity experiment was performed with 3 mg/ml (25 nmoles

total) of TRX-CHI bound resin and 400 µl of 10 mg/ml soluble protein extract. The eluted proteins were analyzed by both immunoblotting and MS. Preliminary results obtained from MS analysis indicated unspecific binding of proteins. Based on the result of the analysis it was evident that series of optimization experiments should be performed. The affinity chromatography experiment was performed three to four more times using different amounts of ligand-bound resin (10-100 µl at 3-4 mg/ml) and soluble protein extract (400-1000 µl at 10-12 mg/ml).

DISCUSSION

To date, different approaches have been applied to characterize protein complexes and protein-protein interaction networks. The first interaction maps were obtained using yeast two-hybrid systems (Bork, 2002; Fields and Sternglanz, 1994; Parrish et al., 2006). More recently, a combination of affinity purification and mass spectrometry has been used to greatly advance our understanding of protein-complex composition. Recently, high-confidence data were produced for the yeast proteome by this approach (Gavin et al., 2006; Krogan et al., 2006). These studies showed that protein interaction mapping is feasible by affinity chromatography linked to mass spectrometry.

A similar approach was used to study protein interactions in the flavonoid pathway. The flavonoid enzyme complex has been extensively studied in the Winkel lab (Burbulis and Winkel-Shirley, 1999; Saslowsky and Winkel-Shirley, 2001; Winkel-Shirley, 2001). Elucidation of protein interactions is a highly complex task demanding the use of different strategies to capture weak protein interactions. In the current research project, an effort was made to reproduce results of previous affinity chromatography experiments with some modifications in the protocol. Although this objective was not achieved, a new protocol was developed for MS analysis for plant proteomics. The results of the mass spectrometry experiments will be discussed in detail in Chapter 3.

Acknowledgments: The project was supported by grants from National Science Foundation (MCB-0445878) to B.S.J.W and Graduate Research Development Program to N.V.

Chapter 3

Mass Spectrometric Analysis of Recombinant Proteins, Affinity Eluted sample, and Soluble Extracts From Wild Type Landsburg Ecotype and Mutant *tt5*(86) *Arabidopsis* Seedlings

SUMMARY

In this research, affinity chromatography, liquid chromatography (LC) and mass spectrometry (MS) were used to identify and study the interactions between *Arabidopsis thaliana* flavonoid enzymes and to characterize the proteome of 4-day-old *Arabidopsis* seedlings. Chalcone isomerase (CHI) - one of the flavonoid enzymes, and thioredoxin (TRX) - a fusion protein attached to CHI that was used for control experiments, were expressed in *Escherichia coli* and purified by metal affinity chromatography. These proteins were characterized by LC-MS/MS and were used to prepare affinity columns for pull-down assays with *Arabidopsis* seedling extracts. *Arabidopsis* wild type Landsburg ecotype and mutant *tt5*(86) seedlings were grown on medium containing sucrose and under continuous white light conditions to enhance the expression of flavonoid genes. Under the stated conditions, wild type Landsburg seedlings accumulate high levels of flavonoid enzymes, which should be useful for proteomic analysis of flavonoid enzyme interactions. The *tt5*(86) mutant lacks the CHI enzyme (Shirley et al, 1995) and has previously been shown to have elevated levels of other flavonoid enzymes (Pelletier et al., 1999). This mutant has also been used in affinity chromatography experiments to enhance the recovery of proteins interacting with immobilized recombinant CHI.

A reversed-phase liquid chromatography electrospray ionization tandem MS (RPLC-ESI-MS/MS) method was developed to evaluate the purity of CHI and CHS, and to analyze the proteome of 4-day-old *Arabidopsis* seedlings. Protein identifications were performed by searching the tandem mass spectra against *Arabidopsis* and *E. coli* protein databases, respectively. CHI and TRX were found as top matching proteins with p-values less than 0.001 (the p-value being the probability of a random match). The total number of proteins ($p<0.001$) detected in the Landsburg and mutant *tt5*(86) extracts was 491 and 633, respectively. Protein detection limits were determined to be in the ~4-40 fmol range.

Keywords: *Arabidopsis* seedlings, proteomics, mass spectrometry.

INTRODUCTION

In recent years, mass spectrometry has emerged as an important tool for the identification and quantification of proteins in complex cellular extracts and for mapping protein-protein interactions (Dreger, 2003; Fu et al., 2005). Such proteomic studies can help with confirming protein localization and function in plant and mammalian systems. Because *Arabidopsis thaliana* was the first plant for which an annotated genome sequence was made public (2000) , most plant proteomic studies to date have been performed with this model system. The protein contents of different *Arabidopsis* organelles such as the chloroplast, mitochondria, nucleus, and vacuoles have been investigated. In these studies, 690 proteins have been identified in the chloroplast (Kleffmann et al., 2004), 114 proteins in the mitochondria (Brugiere et al., 2004), and 217 proteins in the nucleus (Pendle et al., 2005) by tandem mass spectrometric analysis.

Shotgun proteomics that involves MS detection is often used to identify the protein complement of cells (Pocsfalvi et al., 2006), and has been successfully used in several studies to detect *Arabidopsis* proteins, as well. Work done by Charmont *et al*. (2005) to identify apoplastic proteins in *Arabidopsis* seedlings resulted in the identification of 44 secreted proteins. In another study, proteomic analysis of germinating *Arabidopsis* seeds and young seedlings by 2D gel electrophoresis and matrix assisted laser desorption ionization time-of-flight (MALDI-TOF) or TOF/TOF-MS resulted in the identification of a total of 293 proteins present at all stages, and 226 polypeptides involved in different metabolic pathways (Arts and Hollman, 2005).

In the proposed study affinity chromatography and mass spectrometry techniques were applied to perform protein profiling and interaction studies. This research had two main objectives. The first objective involved the proteomic characterization of recombinant CHI, TRX, and protein complexes isolated on CHI-affinity columns. The purity of recombinant TRX and TRX-CHI was assessed by LC-ESI-MS/MS. These proteins were then further used to prepare affinity columns for pull-down assays with *Arabidopsis* extracts. Relevant samples eluted from the affinity column were analyzed by MS to identify the proteins that interact with the flavonoid enzyme, CHI. The second objective involved the proteomic profiling of an

Arabidopsis protein extract. Soluble protein extracts prepared from 4-day-old wild type Landsburg and mutant *tt5*(86) seedlings were analyzed by LC-ESI-MS/MS to detect proteins present in each extract. The goal was further extended to create an LC separation method for the analysis of complex protein mixtures.

MATERIALS AND METHODS

Recombinant Protein Expression.

CHI was expressed as a fusion to the carboxyl terminus of TRX in pCD1 vector (Pelletier et al., 1999). TRX was also expressed without CHI from the pCD1 vector without an insert. These were expressed in *E. coli* strain BL21(DE3) (Novagen), using 1 ml of 0.5M dioxane-free isopropyl-β-D-thiogalactopyranoside (IPTG) (Fisher Scientific, New Jersey) per liter of mid-log phase culture and incubating at room temperature, 200 rpm for 4 h. Pellets were collected by centrifugation at 5,000 rpm in a JA-10 rotor at 4°C for 15 min and then stored at -20°C. The cell pellets from 1 liter of culture were resuspended in 33 ml lysis buffer (50 mM HEPES, 150 mM NaCl, 10% glycerol, 1% Tween-20, 10 mM 2-mercaptoethanol), and then sonicated (6 × 10 sec) on ice. Cell debris were collected at 14,000 rpm at 4°C for 40 min in the JA-17 rotor.

Recombinant Protein Purification Using Metal Affinity Chromatography

Recombinant proteins were purified from the supernatant by TALON metal affinity chromatography. Affinity resin was purchased from BD Biosciences CLONTECH. As per the manufacturer's instructions, columns were prepared using 1 ml resin and equilibrated with 10 volumes of lysis buffer. The *E. coli* crude protein sample (the supernatant) was loaded on the column and washed with 10 volumes of wash buffer (50 mM HEPES, 50 mM NaCl, 10 mM imidazole, pH 8.0, 10% glycerol, 10 mM 2-mercaptoethanol). The bound protein was eluted with elution buffer (50 mM HEPES, 150 mM NaCl, 250 mM imidazol, 10% glycerol, 10 mM 2-mercaptoethanol) in the following order: 0.5 bed volume (b.v.) , 1.5 b.v., 1 b.v., 1 b.v, and 1 b.v. Successful protein expression and purification was verified using sodium dodecylsulfate

polyacrylamide gel electrophoresis (SDS-PAGE) and LC-ESI-MS/MS. The eluted recombinant protein was dialyzed against 100 mM MOPS, pH 7.5, 80 mM $CaCl_2$ at 4°C for 4 hrs in a Slide-A-Lyzer dialysis cassette (10,000 molecular weight cutoff, Pierce, Rockford, IL).

Mass Spectrometric Analysis of Recombinant TRX and TRX-CHI

To enable MS characterization, the detergents and salts from the purified recombinant protein solutions were removed by acetone precipitation. Protein fractions collected following metal affinity chromatography were placed in acetone compatible tubes and treated with cold (-20°C) acetone (4X the sample volume). The samples were vortexed and incubated for 60 min at -20°C. After incubation, the samples were pelleted at 13,000 × g for 10 min. The supernatant was decanted and the pellets rinsed twice with cold acetone (-20°C) for 5 min at 13,000 × g. Next, the acetone was evaporated with a vacuum centrifuge. For MS analysis, the pellets recovered from 1 ml of recombinant protein sample recovered after metal affinity chromatography were dissolved in 25 µl denaturing solution (5M urea, 52 mM Tris-HCl, and 5 mM 2,3-dithiothreitol (DTT)), incubated for 45 min at 60°C, cooled to room temperature, and diluted 9X with 55 mM NH_4HCO_3. The proteins were digested by adding 1µg of trypsin (substrate: enzyme ratio ~ 50:1) and incubating at 37°C overnight. The digestion reaction was quenched by the addition of 10 µl trifluoroacetic acid (TFA) per ml of tryptic digest solution.

Plant Material

Arabidopsis Landsburg wild type and *tt5*(86) seedlings were grown on MS-sucrose-agar medium as described previously (Kubasek et al., 1992). Flavonoid expression was induced by germinating the seeds in constant white light (100 µE/m^2) at 22°C for three days. Seedlings were collected, washed with distilled water, and gently blotted dry on Whatman #1 paper. Seedlings were used fresh, or quick-frozen in liquid nitrogen and then stored at -80°C.

Plant Lysate

To prepare soluble plant lysate for affinity chromatography, seedlings were ground to a fine powder in liquid nitrogen as described previously (Burbulis and Winkel-Shirley, 1999). Approximately 1 g of ground material was suspended in 1 ml plant lysis buffer [(50 mM Tris, pH 7.2, 150 mM NaCl, 1 mM 2-mercaptoethanol, 70 µg/ml DNase I, 0.6% NP-40, 0.6 % 3-[(3-cholamido-propyl) dimethylammonio]-1-propanesulfonate (CHAPS)] containing 1 tablet protease inhibitor cocktail (Roche Diagnostics, IN, USA) and incubated at 25°C for 20 min. The soluble protein fraction was isolated by pelleting the insoluble material at 30,000 × g at 20°C for 10 min.

Mass Spectrometric Analysis of Wild Type Landsburg and Mutant *tt5*(86).

Soluble plant lysates were treated with urea (5 M final concentration) and DTT (4.5 mM final concentration) to reduce disulfide bonds. To avoid the possible scenario of incomplete alkylation and the generation of byproducts, the alkylation step was omitted, and digestion was performed immediately after DTT treatment. The samples were denatured for 1 h at 60ºC, cooled at room temperature, diluted 8X with 55 mM NH_4HCO_3, and digested with trypsin (substrate:enzyme ratio ~50:1 w/w) overnight at 37°C. The digestion reaction was quenched by adding 10 µl TFA per ml tryptic digest solution. The samples were then desalted using a SPEC-PTC18 solid phase extraction cartridge (Varian Inc., Lake Forest, CA). The SPEC cartridge was rinsed with 50 µl wetting solution (CH_3CN/H_2O, 50:50, v/v) and then with 50 µl equilibrating solution (0.1% TFA in H_2O). The entire digest was passed through the cartridge multiple times to allow for maximum peptide adsorption. The cartridge was rinsed with 50 µl wash solution (CH_3OH/H_2O/TFA, 5:95:1, v/v). The peptides were eluted with multiple additions of 50 µl elution solution I (CH_3CN/H_2O/TFA, 60:40:0.1, v/v) and once with elution solution II (CH_3CN/H_2O/TFA, 80:20:0.1, v/v). The samples were then concentrated to ~100 µl final volume with a vacuum centrifuge. Next, the sample was cleaned with a SPEC-SCX cartridge (Varian Inc., Lake Forest, CA). Samples were diluted by adding ~200 µl of 0.1 M CH_3COOH. The SCX cartridge was rinsed with ~250 µl wetting solution (CH_3OH) and then with 2 x 250 µl equilibrating solution (0.1 M CH_3COOH). The entire sample was passed through the cartridge

multiple times. The cartridge was washed with 100 µl wash solution [CH_3OH/CH_3COOH (0.1 M), 50:50, v/v]. Peptides were eluted with 100 µl elution buffer (CH_3OH/NH_4OH, 98:2, v/v). The sample was brought to dryness in a Speedvac to remove the organic solvents. The sample was re-dissolved in 100 µl H_2O containing 0.1% TFA.

HPLC And Mass Spectrometer Set Up

A high performance liquid chromatography system (Agilent Technologies, Palo Alto, CA) and an LTQ ion trap mass spectrometer (Thermo Electron Corp., San Jose, CA) were used to analyze the protein components in the various biological samples (Sarvaiya et al., 2006). The LC system was coupled with the LTQ-MS thorough a home-built ESI source. Nano-LC columns (RPC18, 5µm, 100 µm i.d. x 10 cm long, prepared in-house) were used for sample separation prior to MS detection. The eluent flow generated by the HPLC pumps was split only during sample elution, but not during sample loading, enabling thus the loading of the entire sample on the separation column.

HPLC

Reversed phase columns (100 µm i.d. × 10 cm) were prepared in-house with 5 µm Zorbax SB-C18 packing material (Agilent Technologies). The LC separation columns were fitted with 1 cm long (20 µm i.d. × 90 µm o.d.) nanospray emitters. Solvent A consisted of H_2O/CH_3CN (95:5 v/v) + 0.01% TFA, and solvent B of CH_3CN/ H_2O (80:20 v/v) + 0.01% TFA. An eluent gradient from 0 to 60% B at ~150-170 nl/min was used for sample elution. Digested peptides were analyzed by injecting 8 µl sample on the HPLC column.

ESI-MS/MS

For the mass spectrometric analysis of samples, a data-dependent acquisition method was used, i.e., one MS scan was followed by one zoom scan and one MS^2 on the top five most intense peaks. Protein database searches were performed with the BioWorks 3.2 software (Thermo Electron Corp., San Jose, CA) against *Arabidopsis* and *E. coli* databases. Only fully

tryptic fragments (maximum allowed missed cleavage sites was two) were considered for peptide matching. Chemical and posttranslational modifications were not allowed, and the capability to match one peptide sequence to multiple references within the database was disabled. Data filtering typically included three sets of filters: filter 1 - Xcorr vs. charge state (Xcorr=1.9 for z=1, Xcorr=2.2 for z=2, and Xcorr=3.8 for z≥3), filter 2 - different peptides, and filter 3 - Top 1 protein matches. Confident peptide identifications were considered those that were characterized by $p<0.001$ (the p-value being the probability of a random match). For easier interpretation, the p-values shown in each table are given as log values, i.e., -10 x (log p). Thus, a $p > 30$ corresponds to a p-value < 0.001 and a probability of false positive interpretations <0.1% (Rosner, 2006).

RESULTS

1. MS Analysis of Recombinant Proteins and Affinity Chromatography Samples

Analysis of Recombinant Proteins

TRX-CHI and TRX were analyzed by LC-ESI-MS/MS. In each experiment, 8 µl of protein digest solution [~6 µg of TRX-CHI (M.W. 45 KDa) and ~3 µg of TRX (M.W. 18 KDa)] were injected in the LC system. To detect contaminating *E. coli* proteins, the MS raw files were searched against an *E. coli* protein database appended with the amino acid sequence of CHI. CHI (from TRX-CHI) was detected with $p<0.001$ (Table 3.1). CHI was matched by 18 unique peptides with $p < 0.001$ (Table 3.2), corresponding to an amino acid (AA) sequence coverage of 70.52% (Figure 3.1). A tandem mass spectrum of a representative CHI peptide is shown in Figure 3.2.

TRX was independently identified as the third top-ranking protein in the list, also with $p < 0.001$ (Table 3.1). There were 11 unique peptides that matched the protein (Table 3.3.), accounting for a sequence coverage of 81.65% (Figure 3.3). The display of TRX in Table 3.1 as a *Salmonella* instead of an *E.coli* TRX sequence is an artifact of the database search process. The

two sequences are identical, but the software displays only the first matching entry from the database. TRX was attached to CHI as a fusion protein to increase stability and solubility (Yasukawa et al., 1995). A representative tandem mass spectrum of a TRX peptide is shown in Figure 3.4. Apart from the two proteins of interest, 23 *E. coli* proteins with $p < 0.001$ were found in the sample. These proteins are contaminants that were observable in the SDS-PAGE experiment as well (Figure 2.3). None of the proteins had multiple histidines, but were most probably retained on the affinity column by non-specific binding. In addition, TRX alone was also analyzed by LC-ESI-MS/MS, and was again identified with $p < 0.001$ by four matching peptides.

Table 3.1: Mass spectrometry analysis of recombinant TRX-CHI. The database search was performed against an *E. coli* protein database appended with the CHI protein amino acid sequence. The data filtering parameters were: (1) Xcorr vs. charge state, (2) Different peptides, and (3) Top 1 matches reported only. CHI was the top matching protein (p-values in the table are expressed as -10 x log p).

	Reference	P (pro)	Sf	Score	Coverage	Peptide (Hits)
1	**chalcone-flavanone isomerase / chalcone isomerase**	**3.00E+02**	**17.81**	**200.32**	**70.52**	**20 (2000 0 0)**
2	hypothetical protein [Escherichia coli K12]	3.00E+02	2.89	30.34	37.32	3 (3 0 0 0 0)
3	thioredoxin [Salmonella enterica subsp. Enterica serovari]	1.44E+02	9.31	110.34	81.65	11 (1100 0 0)
4	putative ligase [Escherichia coli K12]	1.40E+02	2.79	30.27	14.44	3 (3 0 0 0 0)
5	cyclic AMP receptor protein [Escherichia coli O157:H7 33]	1.29E+02	7.32	90.27	48.57	9 (9 0 0 0 0)
6	Lactose operon repressor	1.28E+02	3.76	40.41	21.94	4 (4 0 0 0 0)
7	orf, hypothetical protein [Escherichia coli O157:H7DL933]	1.27E+02	0.99	10.33	13.04	1 (1 0 0 0 0)
8	respiratory NADH dehydrogenase [Escherichia coli K12]	1.25E+02	2.35	30.31	14.98	3 (3 0 0 0 0)
9	ATP-binding cell division protein, septation process,	1.24E+02	0.96	10.27	7.14	1 (1 0 0 0 0)
10	30S ribosomal subunit protein S2 [Escherichia coli	1.23E+02	3.87	40.31	28.22	4 (4 0 0 0 0)
11	16S pseudouridylate 516 synthase [Escherichia coli	1.21E+02	1.89	20.34	12.99	2 (2 0 0 0 0)
12	galactitol-specific enzyme IIA of phosphotransferase	1.19E+02	0.96	10.26	12.67	1 (1 0 0 0 0)
13	Hypothetical protein ybjS	1.18E+02	0.91	10.21	5.64	1 (1 0 0 0 0)
14	50S ribosomal subunit protein L33 [Escherichia coli	1.18E+02	0.97	10.23	27.27	1 (1 0 0 0 0)
15	hypothetical protein [Escherichia coli K12]	1.02E+02	2.83	30.3	15.99	3 (3 0 0 0 0)
16	XerC [Escherichia coli K12]	9.91E+01	0.97	10.23	4.36	1 (1 0 0 0 0)
17	30S ribosomal subunit protein S15 [Escherichia coli K12]	9.47E+01	0.94	10.26	33.71	1 (1 0 0 0 0)
18	30S ribosomal subunit protein S10 [Escherichia coli	8.80E+01	2.68	30.2	24.27	3 (3 0 0 0 0)
19	FucR [Escherichia coli K12]	8.79E+01	1.75	20.2	16.05	2 (2 0 0 0 0)
20	putative transformylase [Escherichia coli K12]	8.27E+01	0.98	10.3	3.79	1 (1 0 0 0 0)
21	negative regulator [Escherichia coli O157:H7 EDL933]	8.08E+01	0.94	10.2	14.19	1 (1 0 0 0 0)
22	spermidine N1-acetyltransferase [Escherichia coli	7.83E+01	1.81	20.21	22.04	2 (2 0 0 0 0)
23	hypothetical protein [Escherichia coli K12]	7.31E+01	2.77	30.2	21.76	3 (3 0 0 0 0)
24	hypothetical protein [Escherichia coli K12]	6.63E+01	0.92	10.21	14.75	1 (1 0 0 0 0)
25	cell division; forms circumferential ring;	5.18E+01	0.97	10.25	4.18	1 (1 0 0 0 0)

Table 3.2: Peptide report for CHI. CHI was identified by 18 unique peptides with p<0.001 (p-values are expressed as -10 x log p).

	Reference		P (pro)	Sf	Score	Coverage	Peptide (Hits)
	Scan(s)	Peptide	P (pep)	Sf	XC	DeltaCn	Ions
	gi|15233190|ref|NP_191072.1| chalcone-flavanone isomerase		3.00E+02	17.81	200.32	70.5	20 (20 0 0 0 0)
1	13372	K.FVIFTVIGVYLEGNAVPSLSVK.W	3.00E+02	0.97	6.44	0.65	43/126
2	7681	K.TTEELTESIPFFREIVTGAFEK.F	1.26E+02	0.9	4.03	0.53	34/126
3	4821	K.VTMKLPLTGQQYSEKVTENCVAIWK.Q	1.22E+02	0.94	4.71	0.59	44/144
4	2480	K.SPASSNPLFLGGAGVR.G	1.21E+02	0.98	5.53	0.67	29/45
5	6713	K.LLAEAVLESIIGKNGVSPGTR.L	1.17E+02	0.97	5.83	0.67	42/120
6	13502	R.GLDIQGKFVIFTVIGVYLEGNAVPSLSVK.W	1.14E+02	0.97	5.36	0.69	26/84
7	3815	K.GKTTEELTESIPFFR.E	1.12E+02	0.98	5.96	0.69	24/42
8	12591	K.LLAEAVLESIIGK.N	1.03E+02	0.98	5.69	0.65	28/36
9	8352	K.EETFPPGSSILFALSPTGSLTVAFSK.D	1.00E+02	0.94	4.3	0.72	23/75
10	2073	K.LHVDSVTFVPSVK.S	9.71E+01	0.97	4.71	0.67	23/36
11	4744	K.TTEELTESIPFFR.E	9.13E+01	0.94	3.86	0.52	23/36
12	3006	K.DDSIPETGIAVIENK.L	8.67E+01	0.95	4.78	0.61	24/42
13	1356	K.QLGLYTDCEAK.A	7.87E+01	0.78	2.79	0.37	19/30
14	634	K.LPLTGQQYSEK.V	7.62E+01	0.96	3.73	0.61	21/30
15	13601	K.FLEIFKEETFPPGSSILFALSPTGSLTVAFSK.D	6.32E+01	0.94	4.85	0.67	38/186
16	6255	K.VTENCVAIWK.Q	6.15E+01	0.95	3.62	0.54	19/27
17	621	R.EIVTGAFEK.F	4.42E+01	0.72	3.08	0.32	15/24
18	3171	R.EIVTGAFEKFIK.V	4.28E+01	0.72	2.34	0.32	19/33
19	2995	K.FLEIFK.E	2.10E+01	0.92	3.01	0.41	12/15
20	352	R.GLDIQGK.F	8.99E+00	0.32	2.07	0.17	10/18

MSSSNACASPSPFPAVTK**LHVDSVTFVPSVKSPASSNPLFLGGAGVRGLDIQGKFVIFTVIGVYLEGNAV**
PSLSVKWKGKTTEELTESIPFFREIVTGAFEKFIKVTMKLPLTGQQYSEKVTENCVAIWKQLGLYTDC
EAKAVEKFLEIFKEETFPPGSSILFALSPTGSLTVAFSKDDSIPETGIAVIENKLLAEAVLESIIGKNGVSP
GTRLSVAERLSQLMMKNKDEKEVSDHSVEEKLAKEN

Figure 3.1: Sequence coverage report for CHI. The underlined amino acid sequences are part of the 18 unique peptides that identified CHI. The amino acid sequence coverage was 70.52% (see also Table 3.2).

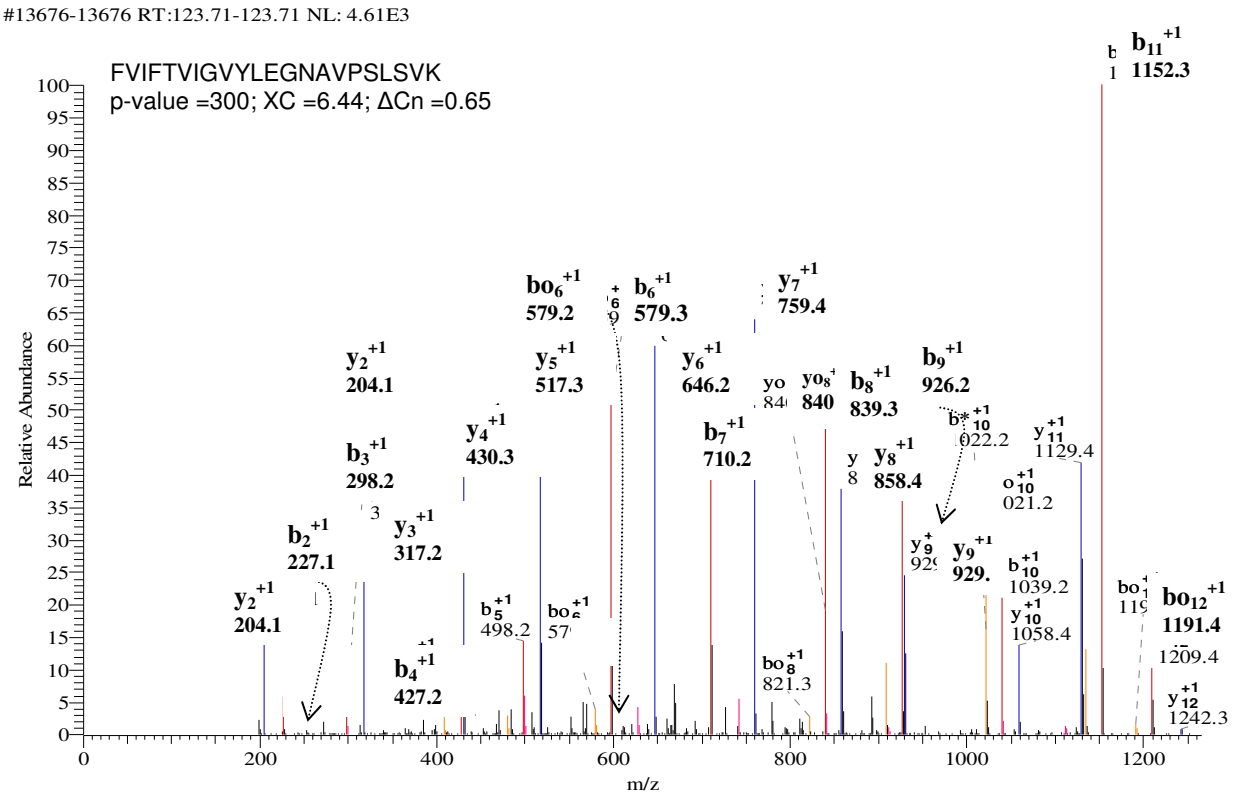


Figure 3.2: Tandem mass spectrum of a representative peptide from recombinant CHI.

Table 3.3: Peptide report for TRX. TRX was identified by 9 unique peptides with p<0.001 (p-values are expressed as -10 x log p).

Reference		P (pro)	Sf	Score	Coverage	Peptide (Hits)
Scan(s)	Peptide	P (pep)	Sf	XC	DeltaCn	Ions
gi|16762214|ref|NP_457831.1| thioredoxin		1.44E+02	9.31	110.34	81.7	11 (11 0 0 0 0)
6605	K.MIAPILDEIADEYQGKLTVAK.L	1.44E+02	0.97	5.38	0.63	26/60
3853	R.GIPTLLLFKNGEVAATK.V	1.08E+02	0.94	3.71	0.52	24/48
3591	K.IIHLTDDSFDTDVLK.A	1.04E+02	0.98	5.37	0.68	25/42
6684	K.MIAPILDEIADEYQGK.L	1.01E+02	0.98	5.98	0.65	28/45
13766	K.ADGAILVDFWAEWCGPCK.M	9.02E+01	0.99	6.62	0.68	41/102
4024	R.GIPTLLLFKNGEVAATKVGALSK.G	6.54E+01	0.9	4.11	0.68	35/132
535	K.NGEVAATKVGALSK.G	6.20E+01	0.95	3.64	0.49	25/39
289	K.LNIDQNPGTAPK.Y	4.66E+01	0.97	4.12	0.5	22/33
4704	R.GIPTLLLFK.N	3.64E+01	0.82	3.11	0.42	14/24
229	K.NGEVAATK.V	1.27E+01	0.45	2.04	0.42	12/21
172	K.VGALSK.G	4.64E+00	0.36	2.02	0.13	15/10

MSDK**<u>IIHLTDDSFDTDVLKADGAILVDFWAEWCGPCKMIAPILDEIADEYQGKLTVAKLNIDQNPGTA</u>**

<u>PKYGIR</u><u>**GIPTLLLFKNGEVAATKVGALSK**</u>GQLKEFLDANLA

Figure 3.3: Sequence coverage report for TRX. The underlined amino acid sequences are part of the 11 unique peptides that identified TRX. The amino acid sequence coverage was 81.65% (see also Table 3.3).

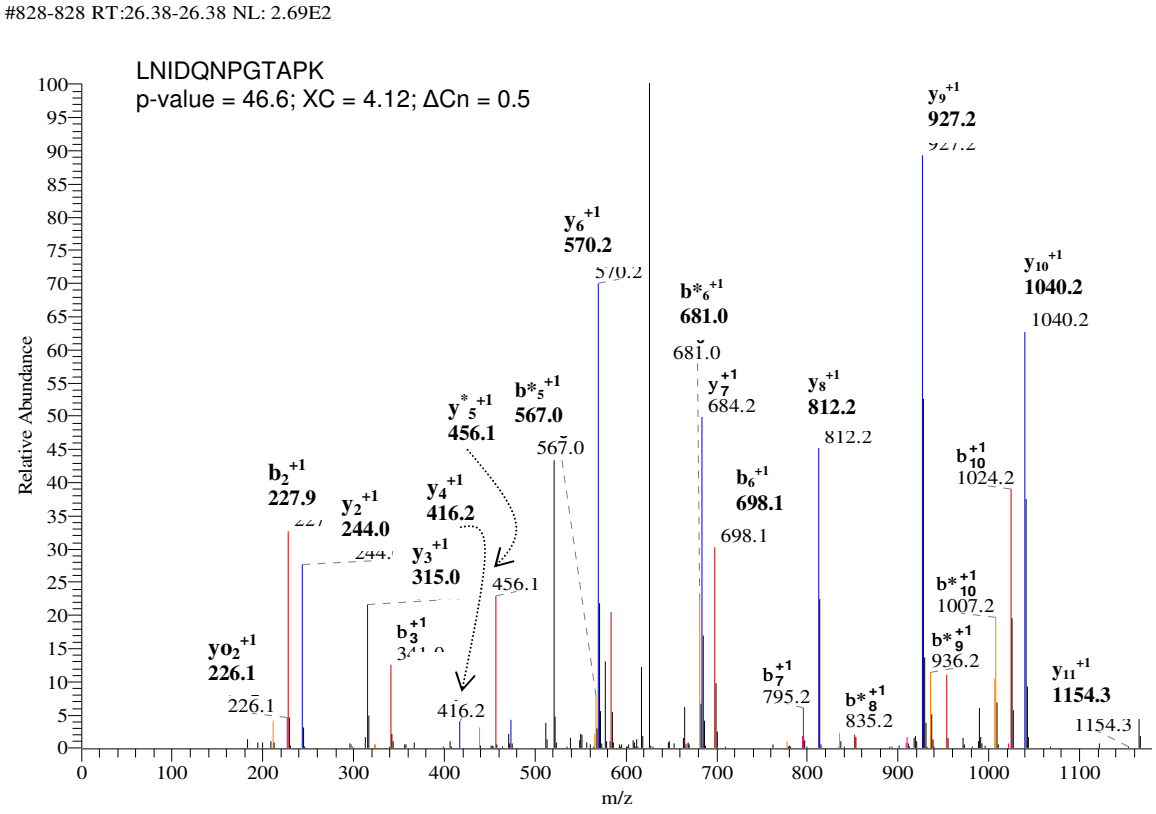


Figure 3.4: Tandem mass spectrum of a representative peptide from recombinant TRX.

Analysis of Affinity Chromatography Purified Proteins

As described in Chapter 2, affinity chromatography was performed in an effort to detect proteins interacting with the flavonoid enzyme, CHI, using extracts from wild type Landsburg ecotype and mutant *tt5*(86). Recombinant TRX-CHI was coupled to the affinity resin and the soluble protein extract prepared from wild type Landsburg ecotype and *tt5*(86) was passed over the column. The protein extract from *tt5*(86) does not contain CHI (Shirley et al., 1992). The absence of endogenous CHI protein has previously been shown to enhance interactions between the soluble proteins present in the extract and immobilized TRX-CHI (Burbulis and Winkel-Shirley, 1999). As a control experiment, recombinant TRX was attached to the affinity resin and wild type Landsburg and mutant *tt5*(86) were passed over the column, respectively.

A major challenge for doing mass spectrometric analysis of samples eluted from affinity chromatography resin is the removal of detergents and salts from the sample. Various sample preparation methods were therefore explored. Clean-up strategies that rely on Ziptip C18 cartridges are frequently used to remove salt contaminants and ionic detergents (Pluskal, 2000), and thus were initially investigated. However, some of the detergents could not be completely removed from the sample, and continued to contaminate the mass spectra, rendering the detection of peptides difficult. The lysis buffer that was used to solubilize proteins in *Arabidopsis* seedlings contained an ionic detergent (SDS), a non-ionic detergent (NP-40), and a zwitterionic detergent (CHAPS) that cannot all be removed in a single sample clean-up step. As a result, trichloroacetic acid (TCA) and acetone precipitation methods were used as an alternative sample clean-up approach. Although the method resulted in successful protein precipitation, redissolving the protein without the addition of any detergent was problematic. Ultimately, the use of a combination of clean-up cartridges proved to be most effective: the SDS detergent was removed with the aid of a solid phase extraction (SPE) C18 cartridge, and the NP-40 and CHAPS detergents were removed with a strong cation exchange (SCX) cartridge.

Preliminary data were obtained after performing several optimization studies. The results of control experiments performed using TRX bound resin and Landsburg protein extracts indicate that there is little nonspecific binding to the resin containing the TRX fusion partner.

All of the proteins had a p value lower than what is considered significant ($p<0.001$). The only observable proteins involved the high abundance components from the plant extract such as RuBisCO (Table 3.4). This result is consistent with the outcome of previous experiments suggesting that endogenous protein complexes present in the Landsburg protein extract do not interact with TRX (Burbulis and Winkel-Shirley, 1999). Similarly, when the CHI-deficient *tt5*(86) protein extract was passed over the TRX bound resin, only 13 proteins, mainly highly expressed components such as RuBisCO, ribosomal proteins, and enolase, were identified (Table 3.5).

No interacting proteins were identified in experiments using resin coupled with TRX-CHI and Landsburg protein extracts. These findings are consistent with the results of the previous experiments (Burbulis and Winkel-Shirley, 1999). Seven proteins were identified in five independent affinity experiments conducted using TRX-CHI coupled resin and *tt5*(86) protein extracts (Table 3.6). These included highly abundant proteins such as RuBisCO and enolase that were also present in the control experiments with TRX and therefore represent non-specific binding to the resin and/or the fusion partner. The other proteins were a chaperonin (at3g03960), a vacuolar ATP synthase subunit B (at1g20260), a Zinc finger family protein (at1g55255), a transcription initiation factor (at1g54270), transketolase (at2g45290), and a member of the mitochondrial carrier protein family (at5g15640). The database search results identified these proteins as top matching entries for the experimental mass spectra. Even though some of these proteins were matched by a single peptide, their probability score was sufficiently good to exclude the possibility of a random match. The data filtering parameters used in this study typically resulted in ~2.5% false positive identification rates (Sarvaiya et al., 2006). These proteins were found in individual experiments. The results demonstrate lack of reproducibility of the results, which indicates that these candidates might not be actual interacting partners. The lack of reproducibility is observed in TRX, negative control experiment, as well. There are only two datasets for TRX experiments available to compare with TRX-CHI experiments. It is likely that some of the proteins might be indentified in TRX experiments. The results of these experiments show variability and do not give exact list of proteins for each experiment. Variability is expected in such experiments and it is important to consider this fact while interpreting the results. Thus, to increase the confidence in the results it is necessary to identify

candidate proteins in multiple independent experiments, and these should not be detected in multiple independent negative control experiments.

One interesting observation is that these proteins were detected with low probability match and with only one peptide in whole-protein analysis. The fact that they are coming through affinity chromatography purification idicates a strong interaction instead of them being low abundance proteins in whole protein extract. There is no previous evidence suggesting that the identified proteins interact with CHI or any other flavonoid enzyme, therefore these represent intriguing new targets for further study. Analysis of the gene expression patterns for the candidate interacting proteins and flavonoid enzymes was performed using Genevestigator in order to determine whether these proteins generally exist together in plant cells. However, little coordinate expression was observed, except between CHI and the vacuolar ATP synthase (Figure 3.7). It was surprising not to find any flavonoid enzymes in the list of interacting proteins. Some of the possible explanations include: (1) the flavonoid enzymes may be present at a very low level in the extract; (2) the interactions are very weak and transient, thus are easily missed if less than optimal conditions are used for protein adsorption and elution; and (3) the proteins are lost during the clean up process.

Table 3.4: Result of the affinity chromatography experiment performed with TRX coupled resin and Landsburg soluble protein extract. The eluted sample was analyzed using a 2 h LC-MS/MS experiment; 8 µl of sample were injected for analysis.

	Reference		P (pro)	Sf	Score	Coverage	Peptide (Hits)
	Scan(s)	Peptide	P (pep)	Sf	XC	DeltaCn	Ions
1	gi|15222729|ref|NP_175945.1| RuBisCO subunit binding-protein beta subunit, chl		1.25E+02	0.98	10.31	3.60	1 (1 0 0 0 0)
	5459	R.GGYPILIIAEDIEQEALATLVVNK.L	1.25E+02	0.98	6.11	0.54	48/138
2	gi|30696056|ref|NP_849818.1| elongation factor 2, putative / EF-2, putative [A		9.13E+01	0.96	10.23	2.70	1 (1 0 0 0 0)
	5263	K.STLTDSLVAAAGIIAQEVAGDVR.M	9.13E+01	0.96	4.64	0.53	43/132
3	gi|15226314|ref|NP_180367.1| RuBisCO subunit binding-protein alpha subunit, ch		8.77E+01	0.94	10.25	3.40	1 (1 0 0 0 0)
	5426	R.APLLIIAEDVTGEALATLVVNK.L	8.77E+01	0.94	4.91	0.43	39/126

Table 3.5: Result of the affinity chromatography experiment performed with TRX coupled resin and *tt5*(86) soluble protein extract. The eluted sample was analyzed using a 2 h LC-MS/MS experiment; 8 µl of sample were injected for analysis.

	Reference		P (pro)	Sf	Score	Coverage	Peptide (Hits)
	Scan(s)	Peptide	P (pep)	Sf	XC	DeltaCn	Ions
1	gi|18405145|ref|NP_565913.1| ribulose bisphosphate carboxylase/oxygenase activ		1.34E+02	0.98	10.37		1 (1 0 0 0 0)
	4818	R.TDKIKDEDIVTLVDQFPGQSIDFFGALR.A	1.34E+02	0.98	7.33	0.57	51/162
2	gi|15226230|ref|NP_180339.1| 60S acidic ribosomal protein P2 (RPP2B) [Arabidop		1.14E+02	0.97	10.30		1 (1 0 0 0 0)
	3988	K.TILGSVGAETEDSQIELLLK.E	1.14E+02	0.97	6.02	0.53	27/57
3	gi|15238686|ref|NP_197294.1| 5-methyltetrahydropteroyltriglutamate--homocystei		1.09E+02	0.97	10.27		1 (1 0 0 0 0)
	5124	R.NIWANDFAASLSTLQALEGIVGK.D	1.09E+02	0.97	5.44	0.58	27/66
4	gi|15231092|ref|NP_191420.1| ketol-acid reductoisomerase [Arabidopsis thaliana		1.07E+02	0.98	10.31		1 (1 0 0 0 0)
	5460	R.GILLGAVHGIVESLFR.R	1.07E+02	0.98	6.26	0.61	31/45
5	gi|15223501|ref|NP_174060.1| 60S ribosomal protein L17 (RPL17A) [Arabidopsis t		1.04E+02	0.83	10.18		1 (1 0 0 0 0)
	6098	K.GLDVDALFISHIQVNQAAK.Q	1.04E+02	0.83	3.54	0.48	18/54
6	gi|15220876|ref|NP_173230.1| nascent polypeptide-associated complex (NAC) doma		9.25E+01	0.88	10.20		1 (1 0 0 0 0)
	4983	R.IGVNSIPAIEEVNIFKDDVVIQFINPK.V	9.25E+01	0.88	4.05	0.53	38/156
7	gi|18391006|ref|NP_563841.1| expressed protein [Arabidopsis thaliana]		9.20E+01	0.98	10.32		1 (1 0 0 0 0)
	4455	K.FTEIDKLVSYGTEVTAIVETGK.I	9.20E+01	0.98	6.30	0.64	32/63
8	gi|15231939|ref|NP_187471.1| 2,3-biphosphoglycerate-independent phosphoglycera		9.11E+01	0.93	10.24		1 (1 0 0 0 0)
	5284	R.VHILTDGRDVLDGSSVGFVETLEADLAALR.A	9.11E+01	0.93	4.79	0.41	43/174
9	gi|15226314|ref|NP_180367.1| RuBisCO subunit binding-protein alpha subunit, ch		7.66E+01	0.97	10.30		1 (1 0 0 0 0)
	5412	R.APLLIIAEDVTGEALATLVVNK.L	7.66E+01	0.97	5.95	0.63	27/63
10	gi|15231317|ref|NP_187342.1| fatty acid multifunctional protein (MFP2) [Arabid		7.43E+01	0.97	10.24		1 (1 0 0 0 0)
	5187	K.AGYISIDIITDLLEAAR.K	7.43E+01	0.97	4.79	0.49	24/48
11	gi|15228818|ref|NP_191168.1| ferritin, putative [Arabidopsis thaliana]		6.34E+01	0.86	10.21		1 (1 0 0 0 0)
	4673	K.NDDVQLADFIESVFLNEQVEAIKK.I	6.34E+01	0.86	4.19	0.45	32/138
12	gi|15227987|ref|NP_181192.1| enolase [Arabidopsis thaliana]		6.12E+01	0.85	10.21		1 (1 0 0 0 0)
	6034	R.SGETEDTFIADLAVGLSTGQIK.T	6.12E+01	0.85	3.95	0.24	22/63

13	gi|15228498|ref|NP_186975.1| UTP--glucose-1-phosphate uridylyltransferase, put		3.25E+01	0.98	20.18		2 (2 0 0 0 0)
	5957	K.EYVFVANSDNLGAIVDLTILK.H	1.86E+00	0.18	2.21	0.22	14/60
	5073	K.ASSDLLLVQSDLYTLVDGFVTR.N	3.25E+01	0.80	3.50	0.51	17/63

Table 3.6 (A), (B),(C),(D), and (E): Results of five independent affinity chromatography experiments performed with TRX-CHI coupled resin and *tt5*(86) soluble protein extract. The eluted samples were analyzed using a 2 h LC-MS/MS experiment; 8 µl of sample was injected for analysis. (A:10/03/06, B: 08/22/06, C:12/02/06, D:11/20/06, E;10/03/06 2H)

(A)

	Reference		P (pro)	Sf	Score	Coverage	Peptide (Hits)
	Scan(s)	Peptide	P (pep)	Sf	XC	DeltaCn	Ions
1	gi|15222729|ref|NP_175945.1| RuBisCO subunit binding-protein beta subunit, chl		3.00E+02	1.96	20.36		2 (2 0 0 0 0)
	5969	R.GGYPILIIAEDIEQEALATLVVNK.L	1.11E+02	0.98	6.77	0.55	50/138
	5586	R.KSQYLDDIAILTGATVIREEVGLSLDK.A	3.00E+02	0.98	7.13	0.59	44/156
2	gi|15226314|ref|NP_180367.1| RuBisCO subunit binding-protein alpha subunit, ch		1.06E+02	0.99	10.36		1 (1 0 0 0 0)
	5954	R.APLLIIAEDVTGEALATLVVNK.L	1.06E+02	0.99	7.28	0.55	31/63
3	gi|18396719|ref|NP_566219.1| chaperonin, putative [Arabidopsis thaliana]		1.05E+02	0.86	10.2		1 (1 0 0 0 0)
	5798	K.AQQEEIGDGANLTISFAGELLQNAEELIR.M	1.05E+02	0.86	4.06	0.52	38/168
4	gi|15228498|ref|NP_186975.1| UTP--glucose-1-phosphate uridylyltransferase, put		1.02E+02	0.95	10.24		1 (1 0 0 0 0)
	5663	K.ASSDLLLVQSDLYTLVDGFVTR.N	1.02E+02	0.95	4.71	0.48	38/126
5	gi|42571561|ref|NP_973871.1| vacuolar ATP synthase subunit B, putative / V-ATP		7.86E+01	0.94	10.21		1 (1 0 0 0 0)

	5320	K.AVVGEEALSSEDLLYLEFLDKFER.K	7.86E+01	0.94	4.23	0.56	39/138
6	gi|15227987|ref|NP_181192.1| enolase [Arabidopsis thaliana]		5.47E+01	0.96	10.24		1 (1 0 0 0 0)
	5012	R.SGETEDTFIADLAVGLSTGQIK.T	5.47E+01	0.96	4.87	0.63	24/63
7	gi|15226354|ref|NP_180387.1| zinc finger (C2H2 type) family protein [Arabidops		3.14E+01	0.84	10.19		1 (1 0 0 0 0)
	5545	K.SSASEEGQNSHFKVSGSALASQASNIINKANK.V	3.14E+01	0.84	3.81	0.28	39/186

(B)

	Reference		P (pro)	Sf	Score	Coverage	Peptide (Hits)
	Scan(s)	Peptide	P (pep)	Sf	XC	DeltaCn	Ions
1	gi|15221761|ref|NP_175829.1| eukaryotic translation initiation factor 4A-2 / e		1.26E+02	0.97	10.24		1 (1 0 0 0 0)
	8336	R.GIYAYGFEKPSAIQQR.G	1.26E+02	0.97	4.77	0.64	30/45
2	gi|15233190|ref|NP_191072.1| chalcone-flavanone isomerase / chalcone isomerase		9.48E+01	1.95	20.27		2 (2 0 0 0 0)
	8407	K.LHVDSVTFVPSVK.S	7.11E+01	0.97	4.02	0.63	25/36
	7254	K.SPASSNPLFLGGAGVR.G	9.48E+01	0.98	5.38	0.56	29/45
3	gi|30689983|ref|NP_566041.2| transketolase, putative [Arabidopsis thaliana]		4.33E+01	0.78	10.12		1 (1 0 0 0 0)
	7340	R.FLAIDAVEK.A	4.33E+01	0.78	2.46	0.23	17/24
4	gi|15228196|ref|NP_191140.1| pyruvate kinase, putative [Arabidopsis thaliana]		1.72E+01	0.57	10.14		1 (1 0 0 0 0)
	7223	R.TAMNNTGILCAVMLDTK.G	1.72E+01	0.57	2.78	0.25	19/48

(C)

	Reference		P (pro)	Sf	Score	Coverage	Peptide (Hits)
	Scan(s)	Peptide	P (pep)	Sf	XC	DeltaCn	Ions
1	gi|15233190|ref|NP_191072.1| chalcone-flavanone isomerase / chalcone isomerase		9.48E+01	0.97	10.22		1 (1 0 0 0 0)
	8651	K.LHVDSVTFVPSVK.S	9.48E+01	0.97	4.31	0.63	24/36
2	gi|15226755|ref|NP_180995.1| 60S ribosomal protein L18A (RPL18aB) [Arabidopsis		4.66E+01	0.49	10.13		1 (1 0 0 0 0)
	8864	K.TTYKANKPNLFM	4.66E+01	0.49	2.54	0.26	14/33
3	gi|18417739|ref|NP_568317.1| mitochondrial carrier protein family [Arabidopsis		2.08E+01	0.16	10.12		1 (1 0 0 0 0)
	9135	R.FLGYGGDSDATAAPSKSK.I	2.08E+01	0.16	2.31	0.05	17/51

(D)

	Reference		P (pro)	Sf	Score	Coverage	Peptide (Hits)
	Scan(s)	Peptide	P (pep)	Sf	XC	DeltaCn	Ions
1	gi|15233190|ref|NP_191072.1| chalcone-flavanone isomerase / chalcone isomerase		7.03E+01	0.98	10.24	4.9	1 (1 0 0 0 0)
	9555	K.LLAEAVLESIIGK.N	7.03E+01	0.98	4.85	0.49	25/36

(E)

	Reference		P (pro)	Sf	Score	Coverage	Peptide (Hits)
	Scan(s)	Peptide	P (pep)	Sf	XC	DeltaCn	Ions
1	gi|15226314|ref|NP_180367.1| RuBisCO subunit binding-protein alpha subunit		7.34E+01	0.92	10.27		1 (1 0 0 0 0)
	6700	R.APLLIIAEDVTGEALATLVVNK.L	7.34E+01	0.92	5.31	0.34	36/126

2. Proteomic Profiling of *Arabidopsis* Protein Extracts

In the current study, global proteomic profiling of extracts prepared from 4-day old *Arabidopsis* wild type Landsburg ecotype and *tt5*(86) seedlings (Kubasek et al., 19920) was performed by ESI-LC-MS/MS. The extracts were subjected to clean up procedures using C18/SCX cartridges followed by tryptic digestion. To maximize the number of identified proteins, an optimization experiment was performed with the Landsburg extract that was analyzed with different HPLC gradients, i.e., 2 h, 4 h, and 6 h in length. Duplicate runs were performed for each condition. The results are summarized in Table 3.7. The overlap in protein profiles for two consecutive runs was ~60-65%. The 2 h gradient resulted in the identification of 320 proteins in the first run and 341 proteins in the second run, the 4 h gradient in 439 proteins in the first run and 491 proteins in the second run, and the 6 h gradient in 500 proteins in the first run and 469 proteins in the second run. As there was no major difference in the number of proteins identified using the 4 h and 6 h gradients, the 4 h separation gradient was used for further analysis. Total 633 proteins were identified with the 4 h gradient for *tt5*(86) protein extract.

Using the Bioworks software, the total number of proteins identified in the Landsburg extract was computed by performing a batch search of all raw files generated from different LC-MS runs. The MS2 data were filtered with two sets of filters. Filter set (I) included filters (1) Xcorr vs. charge state, (2) Different peptides, and (3) Top 1 protein matches only; and, filter set (II) included filter set (I) and an additional filter that passed only proteins matched by at least two unique peptides. The results of applying different filtering parameters are shown in Table 3.8. A total of 809 proteins ($p < 0.001$) was identified with filter set (I), and 475 proteins ($p<0.001$) with filter set (II). A verification of experimental false positives can be usually performed by searching the data against a reversed database of proteins (Sarvaiya et al., 2006). Using the above-described strategy for mapping proteins in cellular extracts, and filter set (I) for selecting the top matching proteins, a false positive rate of 2.4% at the protein identification level is typically obtained. A list of top proteins ($p<0.001$) from the Landsburg extract, identified by 2 unique peptides, is provided in the Appendix. This list was generated by using data filtering

parameters that typically result in false positive identification rates of less than 1%. The flavonoid enzymes, CHS and CHI, were detected in the Landsburg extract; representative tandem mass spectra of peptides corresponding to these enzymes are shown in Figures 3.5-3.6. Only FLS1 was detected in the *tt5*(86) protein extract. All interacting proteins identified in the affinity experiments were also detected in the list of proteins found in the Landsburg and *tt5(86)* extracts, respectively.

The detection limit of proteins in such complex cellular extracts was determined as well. The mass detection limit of an instrument or method is defined as the lowest amount of analyte necessary to obtain a certain quality signal, typically 3-10 times the level of the background noise. In the case of this experiment, we defined the detection limit as the minimum amount of protein injected for analysis that can be detected with the above-defined data filtering parameters. A sample containing nine bovine proteins (hemoglobin α+β, albumin, carbonic anhydrase, α-lactalbumin, fetuin, α-casein, β-casein and cytochrome C) was prepared. The protein mix was digested with trypsin and spiked into the already-digested/cleaned-up *tt5*(86) soluble protein extract at two different concentrations, 0.005 µM and 0.0005 µM. Thus, a sample injection volume of 8 µl for LC-MS/MS analysis contained 40 or 4 fmol, respectively, of each of the test protein digests. The database search was performed against a bovine protein database. The search results indicated that all nine proteins were detectable at the 40 fmol injection level, but only three at the 4 fmol level. From this experiment, it was concluded that the detection limit of the instrument for such complex extracts is in the ~ 4-40 fmol range. For clean, standard protein mixtures digests, the detection limit using MS2 data dependent acquisition conditions was approximately an order of magnitude better, i.e. 0.4-1 fmol.

Table 3.7: Proteins identified in the soluble extract prepared from 4-day-old *Arabidopsis* Landsburg ecotype seedlings.

Landsburg protein extract						
Filters applied: Xcorr vs. charge state, Different peptides, Top 1 protein matches						
No of proteins detected with p<0.001	2h		4h		6h	
	1st Run	2nd Run	1st Run	2nd Run	1st Run	2nd Run
	320	341	439	491	500	469

Table 3.8: Result of a batch database search using all six raw files generated during the 2 h, 4 h, and 6 h gradient LC-MS/MS analysis of the Landsburg protein extract.

Total number of proteins found with p<0.001	
Filters: Different peptides, Xcorr vs. charge state, and Top 1 proteins	Filters: Different peptides, Xcorr vs. charge state, Top 1 proteins, and Two unique peptides/protein
809	475

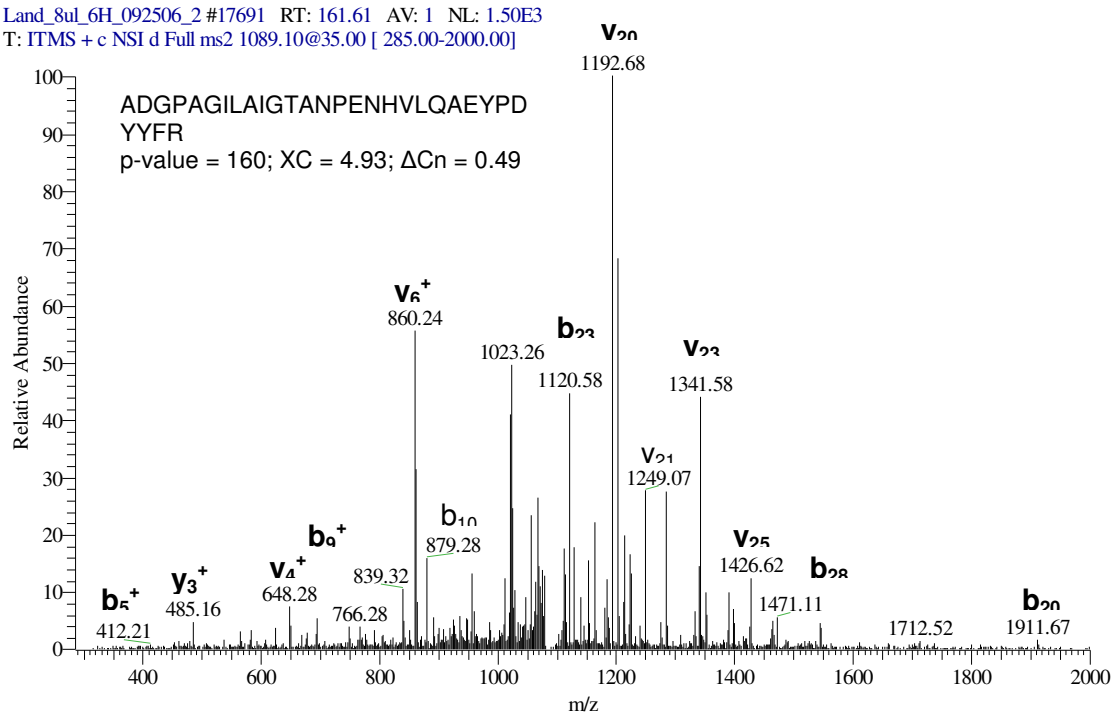


Figure 3.5: Tandem mass spectrum of a representative peptide from endogenous CHS identified in the Landsburg protein extract.

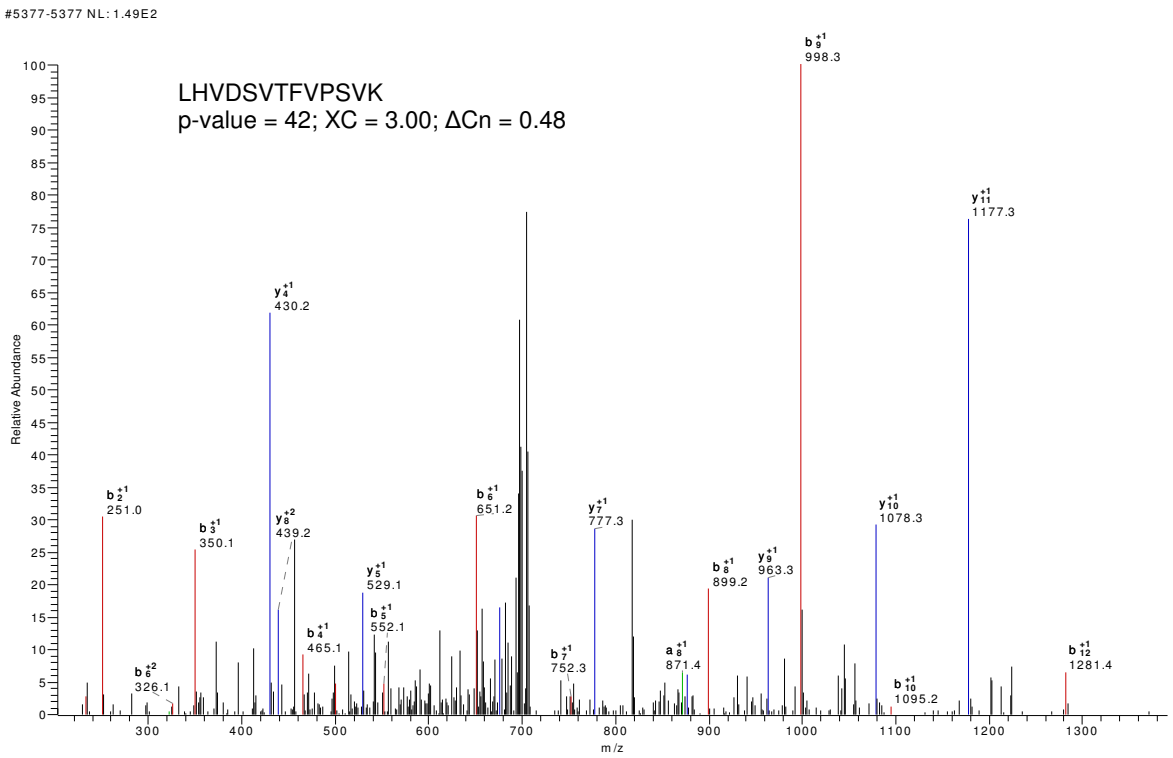


Figure 3.6: Tandem mass spectrum of a representative peptide from endogenous CHI identified in the Landsburg protein extract.

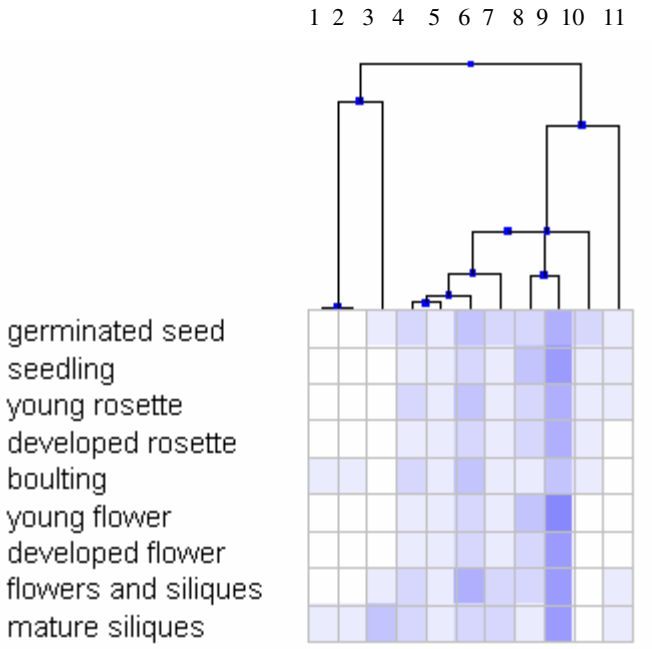


Figure 3.7 Gene expression analysis of flavonoid enzymes and possible interacting proteins at different development stages, as determined using Genevestigator. Numbers 1 to 11 represent: flavonoid 3'-hydroxylase (at5g07990), dihydroflavonol reductase (at5g42800), flavanone 3-hydroxylase (at3g51240), transcription initiation factor (at1g54270), zinc finger protein (at1g55255), chalcone synthase (at5g13930), vacuolar ATP synthase (at1g920260), chalcone isomerase (at5g66220), chaperonin (at3g03960), and flavonol synthase (at5g08640).

DISCUSSION

In the described work, a proteomics approach has been applied to analyze the flavonoid multi-enzyme complex in *Arabidopsis* using affinity chromatography and tandem mass spectrometry. The elucidation of protein interactions is a highly complex task demanding the use of different strategies to obtain convincing results. The flavonoid enzyme complex has been extensively studied in the Winkel laboratory (Burbulis and Winkel-Shirley, 1999; Saslowsky and Winkel-Shirley, 2001; Winkel-Shirley, 2001). The present research helps bridge molecular biology and mass spectrometric techniques.

This project also enabled us to better understand the challenges associated with the analysis of proteomic samples. Novel protocols have been established for protein purification, affinity chromatography, and sample clean up prior to mass spectrometry detection. A method was developed to perform LC-MS/MS analysis of samples eluted from affinity columns. This method will be applied in future experiments in the Winkel laboratory. The analysis of the soluble protein extracts prepared from wild type Landsburg and mutant *tt5*(86) enabled the detection of 491 proteins in the Landsburg extract, and 633 proteins in the *tt5*(86) extract. This result represents the first proteomic analysis of *Arabidopsis* seedling extracts, with a detection limit of 4-40 fmol per LC-MS/MS run. Further protocols are being developed in the Lazar lab for differential protein expression profiling that will be applicable to the analysis of *Arabidopsis thaliana* extracts, as well. Moreover, four possible interacting candidates, were found as result of the affinity chromatography experiments ($p<0.001$, Top 1 protein matches). These are novel candidates that will be intriguing to investigate further in the context of flavonoid metabolism. Interestingly, recent results of yeast two-hybrid experiment done in the Winkel laboratory point to interactions between chalcone synthase (CHS) and proteins involved in auxin transport (Watkinson and Winkel, *in preparation*). It is intriguing to speculate that the vacuolar ATP synthase identified in the affinity experiments may be part of this system, which includes a vacuolar ATPase complex (Li et al., 2005). These findings may provide us with new insights into the extent and nature of protein networks in cells.

A protocol has been developed to analyze complex soluble plant protein extracts with ESI LC-MS/MS. Moreover, a collaborative effort has been established between the Winkel and Lazar labs at Virginia Tech. This attempt has laid a path to perform similar experiments in the future. For example, tandem affinity purification (TAP) followed by mass spectrometry may be performed to purify the components of the flavonoid complex and identify each component of the complex as it exists *in planta.* This approach has been proven successful for the isolation and identification of various protein complexes

in several systems that include yeast, *Drosophila*, and *Arabidopsis* (Forler et al., 2003; Puig et al., 2001; Rigaut et al., 1999; Rubio et al., 2005), and is expected to yield new information for the flavonoid complex, as well. Moreover, the protein map generated from 4-day old *Arabidopsis* seedlings will be useful for future studies that focus on coordinate gene expression and the effects of specific lesions in flavonoid metabolism on the plant proteome.

Acknowledgments: The project was supported by grants from National Science Foundation (MCB-0445878) to B.S.J.W and the Virginia Tech Graduate Research Development Program to N.V.

Chapter 4

Conclusions

Conclusions

Vital cellular functions require the coordinated action of a large number of proteins that are assembled into an array of multiprotein complexes of distinct composition and structure. The analysis of protein complexes and protein-protein interaction networks are therefore of central importance in biological research. In the post-genomic era, protein identification was done by one-dimensional gels, two-dimensional gels, and gel filtration methods. Also, the availability of the genome sequences of many organisms and high-throughput analysis tools has led to a proteomics era (Pandey and Mann, 2000). There are various methods to study protein interactions including yeast two hybrid analysis, affinity chromatography, tandem affinity purification (TAP) tagging, surface plasmon resonance, isothermal titration calorimetry, and immunoprecipitation. The information obtained from these approaches can be used to decipher protein functions in complicated biological networks.

In the present study, an affinity chromatography-based proteomics approach was applied to detect interacting protein partners of flavonoid pathway enzymes in *Arabidopsis*. Several important goals were achieved during the study. First of all, a protocol was developed to analyze soluble protein extracts from *Arabidopsis* seedlings by electrospray ionization combined with liquid chromatography tandem mass spectrometry (ESI LC-MS/MS). The sample preparation protocol involved removal of the detergents and salts from the soluble extract. Also, a standard protocol was developed for tryptic digestion and LC separation of these samples. The protocol is applicable to a variety of situations and will be useful for future proteomic studies in plants.

Also in this study, soluble protein extracts prepared from 4-day old wild type Landsburg and mutant *tt5*(86) seedlings were analyzed by ESI LC-MS/MS. The study identified a large number of proteins in these seedlings, the first of its kind to compare the proteomic profiles of Landsburg and *tt5*(86) seedlings, and at a stage when they have high expression level of flavonoid enzymes. The overlap in protein profiles between two sequential runs of protein extracts was ~60-65%. The list of proteins will be useful for future studies on the interaction and localization of flavonoid pathway enzymes in the Winkel laboratory. Also, the biological samples used in this study represent a highly complex protein mixture. The LC separation

method that was developed can therefore be used in the Lazar laboratory in the future to study a variety of complex protein mixtures. Further studies will be needed to verify putative proteins interactions uncovered in this study by methods like TAP tagging and surface plasmon resonance spectroscopy.

Finally, a preliminary proteomic analysis of samples generated by affinity chromatography of plant extracts has been achieved in the study. The affinity purification approach is useful for the study of multienzyme complexes in various pathways (Turnbull et al., 2004). Moreover, the success of an affinity-based approach depends of the absence of excessive nonspecific interactions that in turn is related to the specificity of the bait-partner interaction. Additional purification steps for sample preparation may result in loss of proteins that are present at low abundance.

For future studies, more advanced affinity approaches could be applied. One of the options for isolating protein complexes on a genome-wide scale is using affinity tags that are fused to proteins of interest followed by mass spectrometric analysis. Tandem affinity purification (TAP) tags have been used in a variety of systems to increase the purity of the isolated complexes. Another new method for identifying transient protein-protein interactions is *in vivo* cross-linking and MS-based protein identification (Vasilescu et al., 2004). Recently, a quantitative proteomics technique called Proteomic Complex Detection using Sedimentation (ProCpDes) was developed to screen for novel protein complexes (Hartman et al., 2007). The method is used for profiling the sediments of a large number of proteins through a rate zonal centrifugation gradient. The approximate size of the unknown complex can be inferred from its sedimentation rate relative to know protein complexes. In the coming years, tools developed to study protein interactions coupled with mass spectrometric analysis are likely to play an increasingly important role in the characterization of protein function.

Acknowledgments: The project was supported by grants from National Science Foundation (MCB-0445878) to B.S.J.W. and the Virginia Tech Graduate Research Development Program to N.V.

References

REFERENCES

(2000). *Arabidopsis* genome initiative. Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature *408*, 796-815.

Abrahams, S., Lee, E., Walker, A. R., Tanner, G. J., Larkin, P. J., and Ashton, A. R. (2003). The Arabidopsis TDS4 gene encodes leucoanthocyanidin dioxygenase (LDOX) and is essential for proanthocyanidin synthesis and vacuole development. Plant J *35*, 624-636.

Alexandersson, E., Saalbach, G., Larsson, C., and Kjellbom, P. (2004). Arabidopsis plasma membrane proteomics identifies components of transport, signal transduction and membrane trafficking. Plant Cell Physiol *45*, 1543-1556.

Arts, I. C., and Hollman, P. C. (2005). Polyphenols and disease risk in epidemiologic studies. Am J Clin Nutr *81*, 317S-325S.

Backman, L., and Johansson, G. (1976). Enzyme-enzyme complexes between aspartate amino-transferase and malatedehydrogenase from pig heart muscle. FEBS Lett *65*, 39-43.

Baneyx, F. (1999). Recombinant protein expression in Escherichia coli. Curr Opin Biotechnol *10*, 411-421.

Beeckmans, S., Van Driessche, E., and Kanarek, L. (1989). The visualization by affinity electrophoresis of a specific association between the consecutive citric acid cycle enzymes fumarase and malate dehydrogenase. Eur J Biochem *183*, 449-454.

Borderies, G., Jamet, E., Lafitte, C., Rossignol, M., Jauneau, A., Boudart, G., Monsarrat, B., Esquerre-Tugaye, M. T., Boudet, A., and Pont-Lezica, R. (2003). Proteomics of loosely bound cell wall proteins of Arabidopsis thaliana cell suspension cultures: a critical analysis. Electrophoresis *24*, 3421-3432.

Bork, P. (2002). Comparative analysis of protein interaction networks. Bioinformatics *18 Suppl 2*, S64.

Borner, G. H., Sherrier, D. J., Weimar, T., Michaelson, L. V., Hawkins, N. D., Macaskill, A., Napier, J. A., Beale, M. H., Lilley, K. S., and Dupree, P. (2005). Analysis of detergent-resistant membranes in Arabidopsis. Evidence for plasma membrane lipid rafts. Plant Physiol *137*, 104-116.

Boudart, G., Jamet, E., Rossignol, M., Lafitte, C., Borderies, G., Jauneau, A., Esquerre-Tugaye, M. T., and Pont-Lezica, R. (2005). Cell wall proteins in apoplastic fluids of Arabidopsis thaliana rosettes: identification by mass spectrometry and bioinformatics. Proteomics *5*, 212-221.

Brugiere, S., Kowalski, S., Ferro, M., Seigneurin-Berny, D., Miras, S., Salvi, D., Ravanel, S., d'Herin, P., Garin, J., Bourguignon, J.*, et al.* (2004). The hydrophobic proteome of mitochondrial membranes from Arabidopsis cell suspensions. Phytochemistry *65*, 1693-1707.

Burbulis, I. E., Pelletier, M. K., Cain, C. C., and Shirley, B. W. (1996). Are flavonoids synthesized by a multi-enzyme complex? So Assoc Agricult Sci Bull *9*, 29-36.

Burbulis, I. E., and Winkel-Shirley, B. (1999). Interactions among enzymes of the *Arabidopsis* flavonoid biosynthetic pathway. Proc Natl Acad Sci USA *96*, 12929-12934.

Cain, C. C., Saslowsky, D. E., Walker, R. A., and Shirley, B. W. (1997). Expression of chalcone synthase and chalcone isomerase proteins in *Arabidopsis* seedlings. Plant Mol Biol *35*, 377-381.

Carter, C., Pan, S., Zouhar, J., Avila, E. L., Girke, T., and Raikhel, N. V. (2004). The vegetative vacuole proteome of Arabidopsis thaliana reveals predicted and unexpected proteins. Plant Cell *16*, 3285-3303.

Chivasa, S., Ndimba, B. K., Simon, W. J., Robertson, D., Yu, X. L., Knox, J. P., Bolwell, P., and Slabas, A. R. (2002). Proteomic analysis of the Arabidopsis thaliana cell wall. Electrophoresis *23*, 1754-1765.

Chuong, S. D. X., Good, A. G., Taylor, G. J., Freeman, M. C., Moorhead, G., and Muench, D. G. (2004). Large scale identification of tubulin-binding proteins provides insight on subcellular trafficing, metabolic channeling, and signaling in plant cells. Molecular and Cellular Proteomics *3*, 970-983.

Czubaty, A., Girstun, A., Kowalska-Loth, B., Trzcinska, A. M., Purta, E., Winczura, A., Grajkowski, W., and Staron, K. (2005). Proteomic analysis of complexes formed by human topoisomerase I. Biochim Biophys Acta *1749*, 133-141.

Dana, C. D., Watkinson, J. E., Martins, R. C. C., Kruger, S., and Winkel, B. S. J. (*submitted*). A structural model for the chalcone synthase-chalcone isomerase bienzyme complex.

Debeaujon, I., Nesi, N., Perez, P., Devic, M., Grandjean, O., Caboche, M., and Lepiniec, L. (2003). Proanthocyanidin-accumulating cells in Arabidopsis testa: regulation of differentiation and role in seed development. Plant Cell *15*, 2514-2531.

Dreger, M. (2003). Proteome analysis at the level of subcellular structures. Eur J Biochem *270*, 589-599.

Ephritikhine, G., Ferro, M., and Rolland, N. (2004). Plant membrane proteomics. Plant Physiol Biochem *42*, 943-962.

Ewing, R. M., Chu, P., Elisma, F., Li, H., Taylor, P., Climie, S., McBroom-Cerajewski, L., Robinson, M. D., O'Connor, L., Li, M.*, et al.* (2007). Large-scale mapping of human protein-protein interactions by mass spectrometry. Mol Syst Biol *3*, 89.

Feinbaum, R. L., and Ausubel, F. M. (1988). Transcriptional regulation of the *Arabidopsis thaliana* chalcone synthase gene. Mol Cell Biol *8*, 1985-1992.

Fields, S., and Sternglanz, R. (1994). The two-hybrid system: an assay for protein-protein interactions. Trends Genet *10*, 286-292.

Forler, D., Kocher, T., Rode, M., Gentzel, M., Izaurralde, E., and Wilm, M. (2003). An efficient protein complex purification method for functional proteomics in higher eukaryotes. Nat Biotechnol *21*, 89-92.

Friso, G., Giacomelli, L., Ytterberg, A. J., Peltier, J. B., Rudella, A., Sun, Q., and Wijk, K. J. (2004). In-depth analysis of the thylakoid membrane proteome of Arabidopsis thaliana chloroplasts: new proteins, new functions, and a plastid proteome database. Plant Cell *16*, 478-499.

Froehlich, J. E., Wilkerson, C. G., Ray, W. K., McAndrew, R. S., Osteryoung, K. W., Gage, D. A., and Phinney, B. S. (2003). Proteomic study of the Arabidopsis thaliana chloroplastic envelope membrane utilizing alternatives to traditional two-dimensional electrophoresis. J Proteome Res *2*, 413-425.

Fu, Q., Wang, B. C., Jin, X., Li, H. B., Han, P., Wei, K. H., Zhang, X. M., and Zhu, Y. X. (2005). Proteomic analysis and extensive protein identification from dry, germinating Arabidopsis seeds and young seedlings. J Biochem Mol Biol *38*, 650-660.

Fukao, Y., Hayashi, M., and Nishimura, M. (2002). Proteomic analysis of leaf peroxisomal proteins in greening cotyledons of Arabidopsis thaliana. Plant Cell Physiol *43*, 689-696.

Gälweiler, L., Guan, C., Müller, A., Wisman, E., Mendgen, K., Yephremov, A., and Palme, K. (1998). Regulation of polar auxin transport by AtPIN1 in *Arabidopsis* vascular tissue. Science *282*, 2226-2230.

Gavin, A. C., Aloy, P., Grandi, P., Krause, R., Boesche, M., Marzioch, M., Rau, C., Jensen, L. J., Bastuck, S., Dumpelfeld, B.*, et al.* (2006). Proteome survey reveals modularity of the yeast cell machinery. Nature *440*, 631-636.

Gavin, A. C., Bosche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J. M., Michon, A. M., Cruciat, C. M.*, et al.* (2002). Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature *415*, 141-147.

Gross, J. H. (2007). Chhabil Dass: Fundamentals of contemporary mass spectrometry. Anal Bioanal Chem.

Halper, L. A., and Srere, P. A. (1977). Interaction between cirtate synthase and mitochondrial malate dehydrogenase in the presence of polyethylene glycol. Arch Biochem Biophys *184*, 529-534.

Harborne, J. B., and Williams, C. A. (2000). Advances in flavonoid research since 1992. Phytochemistry *55*, 481-504.

Hartman, N. T., Sicilia, F., Lilley, K. S., and Dupree, P. (2007). Proteomic complex detection using sedimentation. Anal Chem *79*, 2078-2083.

Havsteen, B. H. (2002). The biochemistry and medical significance of the flavonoids. Pharmacol Ther *96*, 67-202.

Hedman, E., Widen, C., Asadi, A., Dinnetz, I., Schroder, W. P., Gustafsson, J. A., and Wikstrom, A. C. (2006). Proteomic identification of glucocorticoid receptor interacting proteins. Proteomics *6*, 3114-3126.

Hrazdina, G., and Jensen, R. A. (1992). Spatial organization of enzymes in plant metabolic pathways. Ann Rev Plant Physiol Plant Mol Biol *43*, 241-267.

Huber, C. G., Walcher, W., Timperio, A. M., Troiani, S., Porceddu, A., and Zolla, L. (2004). Multidimensional proteomic analysis of photosynthetic membrane proteins by liquid extraction-ultracentrifugation-liquid chromatography-mass spectrometry. Proteomics *4*, 3909-3920.

Jebanathirajah, J. A., and Coleman, J. R. (1998). Association of carbonic anhydrase with a Calvin cycle enzyme complex in Nicotiana tabacum. Planta *204*, 177-182.

Johnson, R. S., Martin, S. A., Biemann, K., Stults, J. T., and Watson, J. T. (1987). Novel fragmentation process of peptides by collision-induced decomposition in a tandem mass spectrometer: differentiation of leucine and isoleucine. Anal Chem *59*, 2621-2625.

Karas, M., and Hillenkamp, F. (1988). Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. Anal Chem *60*, 2299-2301.

Kempner, E. S., and Miller, J. H. (1968). The molecular biology of *Euglena gracilis*. Exp Cell Res *51*, 141-149.

Kleffmann, T., Russenberger, D., von Zychlinski, A., Christopher, W., Sjolander, K., Gruissem, W., and Baginsky, S. (2004). The Arabidopsis thaliana chloroplast proteome reveals pathway abundance and novel protein functions. Curr Biol *14*, 354-362.

Koller, A., Washburn, M. P., Lange, B. M., Andon, N. L., Deciu, C., Haynes, P. A., Hays, L., Schieltz, D., Ulaszek, R., Wei, J.*, et al.* (2002). Proteomic survey of metabolic pathways in rice. Proc Natl Acad Sci U S A *99*, 11969-11974.

Krogan, N. J., Cagney, G., Yu, H., Zhong, G., Guo, X., Ignatchenko, A., Li, J., Pu, S., Datta, N., Tikuisis, A. P.*, et al.* (2006). Global landscape of protein complexes in the yeast Saccharomyces cerevisiae. Nature *440*, 637-643.

Kubasek, W. L., Shirley, B. W., McKillop, A., Goodman, H. M., Briggs, W., and Ausubel, F. M. (1992). Regulation of flavonoid biosynthetic genes in germinating Arabidopsis seedlings. Plant Cell *4*, 1229-1236.

Larsen, M. R., and Roepstorff, P. (2000). Mass spectrometric identification of proteins and characterization of their post-translational modifications in proteome analysis. Fresenius J Anal Chem *366*, 677-690.

Li, J., Ou-Lee, T.-M., Raba, R., Amundson, R. G., and Last, R. L. (1993). Arabidopsis flavonoid mutants are hypersensitive to UV-B irradiation. Plant Cell *5*, 171-179.

Li, J., Yang, H., Peer, W. A., Richter, G., Blakeslee, J., Bandyopadhyay, A., Titapiwantakun, B., Undurraga, S., Khodakovskaya, M., Richards, E. L.*, et al.* (2005). Arabidopsis H+-PPase AVP1 regulates auxin-mediated organ development. Science *310*, 121-125.

Lister, R., Chew, O., Lee, M. N., Heazlewood, J. L., Clifton, R., Parker, K. L., Millar, A. H., and Whelan, J. (2004). A transcriptomic and proteomic characterization of the Arabidopsis mitochondrial protein import apparatus and its response to mitochondrial dysfunction. Plant Physiol *134*, 777-789.

Marmagne, A., Rouet, M. A., Ferro, M., Rolland, N., Alcon, C., Joyard, J., Garin, J., Barbier-Brygoo, H., and Ephritikhine, G. (2004). Identification of new intrinsic proteins in Arabidopsis plasma membrane proteome. Mol Cell Proteomics *3*, 675-691.

Mathews, C. K. (1993). The cell - bag of enzymes or network of channels? J Bact *175*, 6377-6381.

Meyer, P., Heidmann, I., Forkmann, G., and Saedler, H. (1987). A new petunia flower color generated by transformation of a mutant with a maize gene. Nature *330*, 677-678.

Millar, A. H., Heazlewood, J. L., Kristensen, B. K., Braun, H. P., and Moller, I. M. (2005). The plant mitochondrial proteome. Trends Plant Sci *10*, 36-43.

Morita, Y., Hoshino, A., Kikuchi, Y., Okuhara, H., Ono, E., Tanaka, Y., Fukui, Y., Saito, N., Nitasaka, E., Noguchi, H., and Iida, S. (2005). Japanese morning glory dusky mutants displaying reddish-brown or purplish-gray flowers are deficient in a novel glycosylation enzyme for anthocyanin biosynthesis, UDP-glucose:anthocyanidin 3-O-glucoside-2"-O-glucosyltransferase, due to 4-bp insertions in the gene. Plant J *42*, 353-363.

Muir, S. R., Collins, G. J., Robinson, S., Hughes, S., Bovy, A., De Vos, C. H. R., van Tunen, A. J., and Verhoeyen, M. E. (2001). Overexpression of petunia chalcone isomerase in tomato results in fruit containing increased levels of flavonols. Nature Biotechnol *19*, 470-474.

Orosz, F., Wágner, G., Liliom, K., Kovács, J., Baróti, K., Horányi, M., Farkas, T., Hollán, S., and Ovádi, J. (2000). Enhanced association of mutant triosephosphate isomerase to red cell membranes and to brain microtubules. Proc Natl Acad Sci USA *97*, 1026-1031.

Ovádi, J., and Srere, P. A. (1996). Metabolic consequences of enzyme interactions. Cell Biochem Funct *14*, 249-258.

Owens, D. O., Alerding, A. W., Crosby, K. C., and Winkel, B. S. J. (*in preparation*). Characterization of an apparent multigene family encoding flavonol synthase in *Arabidopsis.*

Pandey, A., and Mann, M. (2000). Proteomics to study genes and genomes. Nature *405*, 837-846.

Parrish, J. R., Gulyas, K. D., and Finley, R. L., Jr. (2006). Yeast two-hybrid contributions to interactome mapping. Curr Opin Biotechnol *17*, 387-393.

Peck, S. C. (2005). Update on proteomics in Arabidopsis. Where do we go from here? Plant Physiol *138*, 591-599.

Pelletier, M. K., Burbulis, I. E., and Shirley, B. W. (1999). Disruption of specific flavonoid genes enhances the accumulation of flavonoid enzymes and endproducts in *Arabidopsis* seedlings. Plant Mol Biol *40*, 45-54.

Pelletier, M. K., Murrell, J., and Shirley, B. W. (1997). Arabidopsis flavonol synthase and leucoanthocyanidin dioxygenase: further evidence for distinct regulation of "early" and "late" flavonoid biosynthetic genes. Plant Physiol *113*, 1437-1445.

Pelletier, M. K., and Shirley, B. W. (1996). Analysis of flavanone 3-hydroxylase in Arabidopsis seedlings: Coordinate regulation with chalcone synthase and chalcone isomerase. Plant Physiol *111*, 339-345.

Peltier, J. B., Emanuelsson, O., Kalume, D. E., Ytterberg, J., Friso, G., Rudella, A., Liberles, D. A., Soderberg, L., Roepstorff, P., von Heijne, G., and van Wijk, K. J. (2002). Central functions of the lumenal and peripheral thylakoid proteome of Arabidopsis determined by experimentation and genome-wide prediction. Plant Cell *14*, 211-236.

Peltier, J. B., Ytterberg, A. J., Sun, Q., and van Wijk, K. J. (2004). New functions of the thylakoid membrane proteome of Arabidopsis thaliana revealed by a simple, fast, and versatile fractionation strategy. J Biol Chem *279*, 49367-49383.

Pendle, A. F., Clark, G. P., Boon, R., Lewandowska, D., Lam, Y. W., Andersen, J., Mann, M., Lamond, A. I., Brown, J. W., and Shaw, P. J. (2005). Proteomic analysis of the Arabidopsis nucleolus suggests novel nucleolar functions. Mol Biol Cell *16*, 260-269.

Persson, L., and Srere, P. A. (1992). Purification of the mitochondrial citrate transporter in yeast. Biochem Biophys Res C *183*, 70-76.

Pluskal, M. G. (2000). Microscale sample preparation. Nat Biotechnol *18*, 104-105.

Pocsfalvi, G., Cuccurullo, M., Schlosser, G., Cacace, G., Siciliano, R. A., Mazzeo, M. F., Scacco, S., Cocco, T., Gnoni, A., Malorni, A., and Papa, S. (2006). Shotgun proteomics for the characterization of subunit composition of mitochondrial complex I. Biochim Biophys Acta *1757*, 1438-1450.

Prokisch, H., Scharfe, C., Camp, D. G., 2nd, Xiao, W., David, L., Andreoli, C., Monroe, M. E., Moore, R. J., Gritsenko, M. A., Kozany, C.*, et al.* (2004). Integrative analysis of the mitochondrial proteome in yeast. PLoS Biol *2*, e160.

Puig, O., Caspary, F., Rigaut, G., Rutz, B., Bouveret, E., Bragado-Nilsson, E., Wilm, M., and Seraphin, B. (2001). The tandem affinity purification (TAP) method: a general procedure of protein complex purification. Methods *24*, 218-229.

Rice-Evans, C. (2001). Flavonoid antioxidants. Curr Med Chem *8*, 797-807.

Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M., and Seraphin, B. (1999). A generic protein purification method for protein complex characterization and proteome exploration. Nat Biotechnol *17*, 1030-1032.

Rosner, B. A. (2006). Fundamentals of Biostatistics, Vol 6: Thomson Brooks/Cole).

Rubio, V., Shen, Y., Saijo, Y., Liu, Y., Gusmaroli, G., Dinesh-Kumar, S. P., and Deng, X. W. (2005). An alternative tandem affinity purification strategy applied to Arabidopsis protein complex isolation. Plant J *41*, 767-778.

Sadygov, R. G., Cociorva, D., and Yates, J. R., 3rd (2004). Large-scale database searching using tandem mass spectra: looking up the answer in the back of the book. Nat Methods *1*, 195-202.

Santos, M. O., Crosby, W. L., and Winkel-Shirley, B. (2001). Use of a phage-derived antibody to modulate flavonoid metabolism in Arabidopsis. *submitted for publication.*

Sarvaiya, H. A., Yoon, J. H., and Lazar, I. M. (2006). Proteome profile of the MCF7 cancer cell line: a mass spectrometric evaluation. Rapid Commun Mass Spectrom *20*, 3039-3055.

Saslowsky, D., and Winkel-Shirley, B. (2001). Localization of flavonoid enzymes in *Arabidopsis* roots. Plant J *27*, 37-48.

Saslowsky, D. E., Warek, U., and Winkel, B. S. (2005). Nuclear localization of flavonoid enzymes in Arabidopsis. J Biol Chem *280*, 23735-23740.

Schoenbohm, C., Martens, S., Eder, C., Forkmann, G., and Weisshaar, B. (2000). Identification of the *Arabidopsis thaliana* flavonoid 3'-hydroxylase gene and functional expression of the encoded P450 enzyme. Biol Chem *381*, 749-753.

Shimaoka, T., Ohnishi, M., Sazuka, T., Mitsuhashi, N., Hara-Nishimura, I., Shimazaki, K., Maeshima, M., Yokota, A., Tomizawa, K., and Mimura, T. (2004). Isolation of intact vacuoles and proteomic analysis of tonoplast from suspension-cultured cells of Arabidopsis thaliana. Plant Cell Physiol *45*, 672-683.

Shirley, B. W., Hanley, S., and Goodman, H. M. (1992). Effects of ionizing radiation on a plant genome: analysis of two Arabidopsis *transparent testa* mutations. Plant Cell *4*, 333-347.

Shirley, B. W., Kubasek, W. L., Storz, G., Bruggemann, E., Koornneef, M., Ausubel, F. M., and Goodman, H. M. (1995). Analysis of *Arabidopsis* mutants deficient in flavonoid biosynthesis. Plant J *8*, 659-671.

Spivey, H. O., and Merz, J. M. (1989). Metabolic compartmentation. Bioessays *10*, 127-130.

Stafford, H. A. (1981). Compartmentation in natural product biosynthesis by multienzyme complexes, In The Biochemistry of Plants, E. E. Conn, ed. (New York: Academic Press), pp. 117-137.

Stafford, H. A. (1990). Flavonoid Metabolism (Boca Raton: CRC Press).

Takebe, S., Witola, W. H., Schimanski, B., Gunzl, A., and Ben Mamoun, C. (2007). Purification of components of the translation elongation factor complex of Plasmodium falciparum by tandem affinity purification. Eukaryot Cell *6*, 584-591.

Tompa, P., Batke, J., Ovadi, J., Welch, G. R., and Srere, P. A. (1987). Quantitation of the interaction between citrate synthase and malate dehydrogenase. J Biol Chem *262*, 6089-6092.

Toueille, M., Saint-Jean, B., Castroviejo, M., and Benedetto, J. P. (2007). The elongation factor 1A: a novel regulator in the DNA replication/repair protein network in wheat cells? Plant Physiol Biochem *45*, 113-118.

Turnbull, J. J., Nakajima, J., Welford, R. W., Yamazaki, M., Saito, K., and Schofield, C. J. (2004). Mechanistic studies on three 2-oxoglutarate-dependent oxygenases of flavonoid biosynthesis: anthocyanidin synthase, flavonol synthase, and flavanone 3beta-hydroxylase. J Biol Chem *279*, 1206-1216.

Uetz, P., and Hughes, R. E. (2000). Systematic and large-scale two-hybrid screens. Curr Opin Microbiol *3*, 303-308.

Van Leene, J., Stals, H., Eeckhout, D., Persiau, G., Van De Slijke, E., Van Isterdael, G., De Clercq, A., Bonnet, E., Laukens, K., Remmerie, N.*, et al.* (2007). A tandem affinity purification-based technology platform to study the cell cycle interactome in Arabidopsis thaliana. Mol Cell Proteomics *6*, 1226-1238.

Vasilescu, J., Guo, X., and Kast, J. (2004). Identification of protein-protein interactions using in vivo cross-linking and mass spectrometry. Proteomics *4*, 3845-3854.

Vom Endt, D., Kijne, J. W., and Memelink, J. (2002). Transcription factors controlling plant secondary metabolism: what regulates the regulators? Phytochemistry *61*, 107-114.

Watkinson, J. I., and Winkel, B. S. J. (*in preparation*). Evidence for a protein network that connects flavonoid metabolism with auxin transport.

Weiss, M. R. (1991). Floral colour changes as cues for pollinators. Nature *354*, 227-229.

Welch, G. R. (1977). On the role of organized multienzyme systems in cellular metabolism: a general synthesis. Prog Biophys Mol Biol *32*, 103-191.

Werhahn, W., Niemeyer, A., Jansch, L., Kruft, V., Schmitz, U. K., and Braun, H. (2001). Purification and characterization of the preprotein translocase of the outer mitochondrial membrane from Arabidopsis. Identification of multiple forms of TOM20. Plant Physiol *125*, 943-954.

Winkel-Shirley, B. (2001). Flavonoid biosynthesis: a colorful model for genetics, biochemistry, cell biology and biotechnology. Plant Physiol *126*, 485-493.

Winkel, B. S. J. (2004). Metabolic channeling in plants. Annu Rev Plant Biol *55*, 85-107.

Yasukawa, T., Kanei-Ishii, C., Maekawa, T., Fujimoto, J., Yamamoto, T., and Ishii, S. (1995). Increase of solubility of foreign proteins in Escherichia coli by coproduction of the bacterial thioredoxin. J Biol Chem *270*, 25328-25331.

Yu, O., Shi, J., Hession, A. O., Maxwell, C. A., McGonigle, B., and Odell, J. T. (2003). Metabolic engineering to increase isoflavone biosynthesis in soybean seed. Phytochemistry *63*, 753-763.

Zalokar, M. (1960). Cytochemistry of centrifuged hyphae of *Neurospora*. Exp Cell Res *19*, 114-132.

Appendix A

Appendix A Results from proteomic analysis of Landsburg seedlings

A list of top proteins ($p<0.001$) from the Landsburg extract, identified by two unique peptides, is provided in Appendix A. The protein list was generated by using data filtering parameters that typically result in <1% false positive identification rates.

Reference	P (pro)	Coverage	MW	Peptide (Hits)
RCA_ARATH (P10896) Ribulose bisphosphate carboxylase/oxygenase activase, chloroplast precursor (RuBisCO activase) (RA)	171	20.00	51947.95	7 (7 0 0 0 0)
BIP1_ARATH (Q9LKR3) Luminal-binding protein 1 precursor (BiP1) (AtBP1)	300	8.70	73584.01	7 (7 0 0 0 0)
RBS2B_ARATH (P10797) Ribulose bisphosphate carboxylase small chain 2B, chloroplast precursor (EC 4.1.1.39) (RuBisCO small subunit 2B)	211	14.40	20337.10	7 (7 0 0 0 0)
RBS1A_ARATH (P10795) Ribulose bisphosphate carboxylase small chain 1A, chloroplast precursor (EC 4.1.1.39) (RuBisCO small subunit 1A)	178	14.40	20202.95	7 (7 0 0 0 0)
MDHM1_ARATH (Q9ZP06) Malate dehydrogenase 1, mitochondrial precursor (EC 1.1.1.37) (mNAD-MDH 1)	300	20.20	35781.88	6 (6 0 0 0 0)
PSBQ2_ARATH (Q41932) Oxygen-evolving enhancer protein 3-2, chloroplast precursor (OEE3) (16 kDa subunit of oxygen evolving system of photosystem II) (OEC 16 kDa subunit)	300	26.10	24628.06	4 (4 0 0 0 0)
RBS1B_ARATH (P10796) Ribulose bisphosphate carboxylase small chain 1B, chloroplast precursor (EC 4.1.1.39) (RuBisCO small subunit 1B)	300	10.50	20273.11	5 (5 0 0 0 0)
RLA21_ARATH (P51407) 60S acidic ribosomal protein P2-1	194	39.10	11444.86	4 (4 0 0 0 0)
BAS1_ARATH (Q96291) 2-cys peroxiredoxin BAS1, chloroplast precursor (EC 1.11.1.15)	168	13.20	29073.96	2 (2 0 0 0 0)
CAPP2_ARATH (Q5GM68) Phosphoenolpyruvate carboxylase 2 (EC 4.1.1.31) (PEPCase 2) (PEPC 2) (AtPPC2)	300	3.30	109684.50	2 (2 0 0 0 0)
PSBQ1_ARATH (Q9XFT3) Oxygen-evolving enhancer protein 3-1, chloroplast precursor (OEE3) (16 kDa subunit of oxygen evolving system of photosystem II) (OEC 16 kDa subunit)	162	40.80	23780.68	9 (9 0 0 0 0)
O80852 (O80852) Glutathione S-transferase (At2g30860/F7F1.7)	154	33.50	24130.52	5 (5 0 0 0 0)
METE_ARATH (O50008) 5-methyltetrahydropteroyltriglutamate--homocysteine methyltransferase (EC 2.1.1.14) (Vitamin-B12-independent methionine synthase isozyme) (Cobalamin-independent methionine synthase isozyme)	150	22.90	84303.66	16 (16 0 0 0 0)
Q3EC72 (Q3EC72) Protein At2g04030	150	8.80	88201.58	5 (5 0 0 0 0)
COR47_ARATH (P31168) Dehydrin COR47 (Cold-induced COR47 protein)	150	19.60	29878.95	3 (3 0 0 0 0)
SUCB_ARATH (O82662) Succinyl-CoA ligase [GDP-forming] beta-chain, mitochondrial precursor (EC 6.2.1.4) (Succinyl-CoA synthetase, beta chain) (SCS-beta)	147	12.80	45317.00	4 (4 0 0 0 0)
ENO_ARATH (P25696) Enolase (EC 4.2.1.11) (2-phosphoglycerate dehydratase) (2-phospho-D-glycerate hydro-lyase)	145	31.80	47689.37	8 (8 0 0 0 0)
G3PB_ARATH (P25857) Glyceraldehyde-3-phosphate dehydrogenase B, chloroplast precursor (EC 1.2.1.13) (NADP-dependent glyceraldehydephosphate dehydrogenase subunit B)	145	17.70	47629.61	8 (8 0 0 0 0)
TPIS_ARATH (P48491) Triosephosphate isomerase, cytosolic (EC 5.3.1.1) (TIM) (Triose-phosphate isomerase)	145	15.40	27152.06	5 (5 0 0 0 0)
TBA2_ARATH (P29510) Tubulin alpha-2/alpha-4 chain	144	10.90	49509.32	3 (3 0 0 0 0)
O24433 (O24433) Beta-glucosidase	144	6.50	59949.11	2 (2 0 0 0 0)

PSBP1_ARATH (Q42029) Oxygen-evolving enhancer protein 2-1, chloroplast precursor (OEE2) (23 kDa subunit of oxygen evolving system of photosystem II) (OEC 23 kDa subunit) (23 kDa thylakoid membrane protein)	143	13.70	28077.94	3 (3 0 0 0 0)
12S1_ARATH (P15455) 12S seed storage protein CRA1 precursor [Contains: 12S seed storage protein CRA1 alpha chain (12S seed storage protein CRA1 acidic chain); 12S seed storage protein CRA1 beta chain (12S seed storage protein CRA1 basic chain)]	142	29.70	52562.77	12 (12 0 0 0 0)
ATPBM_ARATH (P83483) ATP synthase subunit beta-1, mitochondrial precursor (EC 3.6.3.14)	140	25.40	59634.18	11 (11 0 0 0 0)
ROC1_ARATH (Q9ZUU4) Putative ribonucleoprotein At2g37220, chloroplast precursor	140	6.90	30699.18	2 (2 0 0 0 0)
EF1G1_ARATH (O04487) Probable elongation factor 1-gamma 1 (EF-1-gamma 1) (eEF-1B gamma 1)	139	7.70	46630.64	2 (2 0 0 0 0)
O04313 (O04313) Jasmonate inducible protein isolog (Jasmonate inducible protein; myrosinase binding protein-like)	139	10.80	32213.46	4 (4 0 0 0 0)
CH10C_ARATH (O65282) 20 kDa chaperonin, chloroplast precursor (Protein Cpn21) (Chloroplast protein Cpn10) (Chloroplast chaperonin 10) (Ch-CPN10) (Chaperonin 20)	138	50.20	26785.46	8 (8 0 0 0 0)
Q9LMQ2 (Q9LMQ2) F7H2.16 protein (Chlorophyll A-B binding protein) (Lhcb6 protein) (Putative chlorophyll binding protein)	138	15.90	27505.07	3 (3 0 0 0 0)
Q9LPW0 (Q9LPW0) F13K23.15 protein (At1g12900) (Putative calcium-binding protein, calreticulin)	138	10.50	42820.17	2 (2 0 0 0 0)
EF1G2_ARATH (Q9FVT2) Probable elongation factor 1-gamma 2 (EF-1-gamma 2) (eEF-1B gamma 2)	137	9.40	46370.41	2 (2 0 0 0 0)
Q9SGT4 (Q9SGT4) Elongation factor EF-2	137	11.60	94185.52	10 (10 0 0 0 0)
GLGS_ARATH (P55228) Glucose-1-phosphate adenylyltransferase small subunit, chloroplast precursor (EC 2.7.7.27) (ADP-glucose synthase) (ADP-glucose pyrophosphorylase) (AGPase B) (Alpha-D-glucose-1-phosphate adenyl transferase)	135	6.70	56615.16	2 (2 0 0 0 0)
Q9SRZ6 (Q9SRZ6) F12P19.10 protein (Isocitrate dehydrogenase, putative) (At1g65930/F12P19_10)	134	11.20	45717.18	4 (4 0 0 0 0)
Q94KE3 (Q94KE3) AT3g52990/F8J2_160 (Putative pyruvate kinase)	134	7.00	57459.25	3 (3 0 0 0 0)
Q9ZRW8 (Q9ZRW8) Glutathione transferase (EC 2.5.1.18) (At1g78380/F3F9_11)	133	14.60	25634.34	4 (4 0 0 0 0)
S17P_ARATH (P46283) Sedoheptulose-1,7-bisphosphatase, chloroplast precursor (EC 3.1.3.37) (Sedoheptulose-bisphosphatase) (SBPase) (SED(1,7)P2ase)	132	11.70	42387.59	3 (3 0 0 0 0)
Q9SAJ4 (Q9SAJ4) T8K14.3 protein (At1g79550) (Hypothetical protein At1g79550; T8K14.3) (Putative phosphoglycerate kinase) (Cytosolic phosphoglycerate kinase)	132	20.00	42105.43	8 (8 0 0 0 0)
G3PC_ARATH (P25858) Glyceraldehyde-3-phosphate dehydrogenase, cytosolic (EC 1.2.1.12)	132	28.40	36891.12	7 (7 0 0 0 0)
ACT2_ARATH (Q96292) Actin-2	132	9.00	41849.94	2 (2 0 0 0 0)
Q9SJQ9 (Q9SJQ9) Putative fructose bisphosphate aldolase	131	11.20	38362.89	2 (2 0 0 0 0)
Q8H0S9 (Q8H0S9) Putative aminopeptidase (M1 aminopeptidase) (At1g63770)	131	3.70	99096.17	2 (2 0 0 0 0)
Q9LD57 (Q9LD57) Phosphoglycerate kinase (AT3g12780/MBK21_14)	131	6.00	50080.81	2 (2 0 0 0 0)
Y2766_ARATH (O80934) Protein At2g37660, chloroplast precursor	130	14.20	34858.19	3 (3 0 0 0 0)
RBL_ARATH (O03042) Ribulose bisphosphate carboxylase large chain precursor (EC 4.1.1.39) (RuBisCO large subunit)	130	24.40	52921.67	13 (13 0 0 0 0)
12S2_ARATH (P15456) 12S seed storage protein CRB precursor [Contains: 12S seed storage protein CRB alpha chain (12S seed storage protein CRB acidic chain); 12S seed storage protein CRB beta chain (12S seed storage protein CRB basic chain)]	130	20.40	50526.65	8 (8 0 0 0 0)
Q9FWR4 (Q9FWR4) Putative GSH-dependent dehydroascorbate reductase (GSH-dependent dehydroascorbate reductase 1, putative) (F14P1.45/F14P1.45) (At1g19570/F14P1.45) (Putative	130	16.40	23626.18	2 (2 0 0 0 0)

GSH-dependent dehydroascorbate reductase 1)				
Y5224_ARATH (Q94EG6) Protein At5g02240	129	19.40	27086.19	3 (3 0 0 0 0)
Q8GUM2 (Q8GUM2) Heat shock protein 70 like protein (At4g37910)	128	4.40	73029.71	3 (3 0 0 0 0)
MPPB_ARATH (Q42290) Probable mitochondrial-processing peptidase subunit beta, mitochondrial precursor (EC 3.4.24.64) (Beta-MPP)	127	14.10	59123.35	5 (5 0 0 0 0)
TBB2_ARATH (P29512) Tubulin beta-2/beta-3 chain	126	4.20	50700.97	4 (4 0 0 0 0)
Q56YY3 (Q56YY3) 12S cruciferin seed storage protein	125	57.90	14840.70	7 (7 0 0 0 0)
Q9LV03 (Q9LV03) NADH-dependent glutamate synthase	124	2.50	242723.20	3 (3 0 0 0 0)
PDI1_ARATH (Q9XI01) Probable protein disulfide-isomerase 1 precursor (EC 5.3.4.1) (PDI 1)	124	15.60	55567.13	6 (6 0 0 0 0)
NDK1_ARATH (P39207) Nucleoside diphosphate kinase 1 (EC 2.7.4.6) (Nucleoside diphosphate kinase I) (NDK I) (NDP kinase I) (NDPK I)	123	11.40	16489.53	2 (2 0 0 0 0)
Q9SI75 (Q9SI75) F23N19.11 (Elongation factor G) (Hypothetical protein At1g62750)	123	9.20	86003.36	5 (5 0 0 0 0)
O23254 (O23254) Hydroxymethyltransferase (AT4g13930/dl3005c)	122	16.80	51685.24	8 (8 0 0 0 0)
ACP1_ARATH (P11829) Acyl carrier protein 1, chloroplast precursor (ACP)	122	13.90	15045.94	3 (3 0 0 0 0)
APX1_ARATH (Q05431) L-ascorbate peroxidase, cytosolic (EC 1.11.1.11) (AP)	122	20.50	27412.73	4 (4 0 0 0 0)
SYS_ARATH (Q39230) Seryl-tRNA synthetase (EC 6.1.1.11) (Serine--tRNA ligase) (SerRS)	122	7.50	51596.39	3 (3 0 0 0 0)
Q9LZC3 (Q9LZC3) Malate synthase-like protein	122	5.30	63846.49	2 (2 0 0 0 0)
O23211 (O23211) Globulin-like protein	121	9.00	56412.86	3 (3 0 0 0 0)
CATA2_ARATH (P25819) Catalase-2 (EC 1.11.1.6)	121	14.80	56895.45	6 (6 0 0 0 0)
Q9STW6 (Q9STW6) Hsp 70-like protein	121	3.90	76461.50	3 (3 0 0 0 0)
Q9LF98 (Q9LF98) Fructose-bisphosphate aldolase	120	10.60	38515.88	4 (4 0 0 0 0)
Q9C7J4 (Q9C7J4) Phosphoglycerate kinase, putative	120	4.80	49907.88	2 (2 0 0 0 0)
Q712H0 (Q712H0) Hypothetical protein (Fragment)	120	9.20	53416.90	4 (4 0 0 0 0)
SAHH1_ARATH (O23255) Adenosylhomocysteinase 1 (EC 3.3.1.1) (S-adenosyl-L-homocysteine hydrolase 1) (SAH hydrolase 1) (AdoHcyase 1) (HOMOLOGY-DEPENDENT GENE SILENCING 1 protein)	120	13.60	53344.16	5 (5 0 0 0 0)
UGPA1_ARATH (P57751) UTP--glucose-1-phosphate uridylyltransferase 1 (EC 2.7.7.9) (UDP-glucose pyrophosphorylase 1) (UDPGP 1) (UGPase 1)	120	10.60	51887.28	3 (3 0 0 0 0)
GOX1_ARATH (Q9LRS0) Probable peroxisomal (S)-2-hydroxy-acid oxidase 1 (EC 1.1.3.15) (Glycolate oxidase 1) (GOX 1) (Short chain alpha-hydroxy acid oxidase 1)	119	7.40	40281.20	2 (2 0 0 0 0)
RS3A2_ARATH (Q42262) 40S ribosomal protein S3a-2	118	7.30	29653.74	2 (2 0 0 0 0)
IF4A1_ARATH (P41376) Eukaryotic initiation factor 4A-1 (EC 3.6.1.-) (ATP-dependent RNA helicase eIF4A-1) (eIF-4A-1) (DEAD-box ATP-dependent RNA helicase 4)	118	16.30	46675.03	5 (5 0 0 0 0)
ATPAM_ARATH (P92549) ATP synthase subunit alpha, mitochondrial (EC 3.6.3.14)	118	17.20	55010.95	7 (7 0 0 0 0)
NACA1_ARATH (Q9LHG9) Nascent polypeptide-associated complex subunit alpha-like protein 1 (NAC-alpha-like protein 1) (Alpha-NAC-like protein 1)	118	13.30	21968.88	2 (2 0 0 0 0)
Q96318 (Q96318) 12S cruciferin seed storage protein	117	22.70	58199.39	7 (7 0 0 0 0)
RS71_ARATH (Q9C514) 40S ribosomal protein S7-1	117	18.30	21909.00	3 (3 0 0 0 0)
RS102_ARATH (Q9FFS8) 40S ribosomal protein S10-2	116	18.30	19721.04	3 (3 0 0 0 0)

PGKH_ARATH (P50318) Phosphoglycerate kinase, chloroplast precursor (EC 2.7.2.3)	116	21.50	49893.77	9 (9 0 0 0 0)
Q9S726 (Q9S726) Putative ribose 5-phosphate isomerase (Putative ribose 5-phosphate isomerase; 91580-90750)	115	11.20	29287.41	2 (2 0 0 0 0)
CALR1_ARATH (O04151) Calreticulin-1 precursor	115	8.00	48497.41	3 (3 0 0 0 0)
RS21_ARATH (Q8L8Y0) 40S ribosomal protein S2-1	114	9.50	30722.26	4 (4 0 0 0 0)
RS72_ARATH (Q9M885) 40S ribosomal protein S7-2	113	18.30	22181.92	2 (2 0 0 0 0)
HSP82_ARATH (P55737) Heat shock protein 81-2 (HSP81-2)	112	15.30	80013.91	11 (11 0 0 0 0)
MDHC1_ARATH (P93819) Malate dehydrogenase, cytoplasmic 1 (EC 1.1.1.37)	112	16.00	35548.30	3 (3 0 0 0 0)
ILV5_ARATH (Q05758) Ketol-acid reductoisomerase, chloroplast precursor (EC 1.1.1.86) (Acetohydroxy-acid reductoisomerase) (Alpha-keto-beta-hydroxylacil reductoisomerase)	112	7.60	63772.49	3 (3 0 0 0 0)
IDI2_ARATH (Q42553) Isopentenyl-diphosphate delta-isomerase II (EC 5.3.3.2) (IPP isomerase II) (Isopentenyl pyrophosphate isomerase II)	111	12.70	32586.76	3 (3 0 0 0 0)
RL4B_ARATH (P49691) 60S ribosomal protein L4-2 (L1)	111	7.90	44693.92	3 (3 0 0 0 0)
THIK2_ARATH (Q56WD9) 3-ketoacyl-CoA thiolase 2, peroxisomal precursor (EC 2.3.1.16) (Beta-ketothiolase 2) (Acetyl-CoA acyltransferase 2) (Peroxisomal 3-oxoacyl-CoA thiolase 2) (Peroxisome defective protein 1)	111	10.60	48548.05	3 (3 0 0 0 0)
PSBO1_ARATH (P23321) Oxygen-evolving enhancer protein 1-1, chloroplast precursor (OEE1) (33 kDa subunit of oxygen evolving system of photosystem II) (OEC 33 kDa subunit) (33 kDa thylakoid membrane protein)	111	19.90	35120.64	4 (4 0 0 0 0)
RL4A_ARATH (Q9SF40) 60S ribosomal protein L4-1 (L1)	111	8.60	44674.06	5 (5 0 0 0 0)
HSP71_ARATH (P22953) Heat shock cognate 70 kDa protein 1 (Hsc70.1)	110	10.40	71313.20	4 (4 0 0 0 0)
Q9SYT0 (Q9SYT0) Annexin (Ca2+-dependent membrane-binding protein annexin) (F14D7.2 protein)	110	8.80	36181.52	3 (3 0 0 0 0)
Q8LPR9 (Q8LPR9) At1g06940/F4H5_1	110	4.70	112051.50	3 (3 0 0 0 0)
METK_ARATH (P23686) S-adenosylmethionine synthetase 1 (EC 2.5.1.6) (Methionine adenosyltransferase 1) (AdoMet synthetase 1)	110	6.40	43130.73	2 (2 0 0 0 0)
RL151_ARATH (O23515) 60S ribosomal protein L15-1	109	15.20	24224.17	2 (2 0 0 0 0)
RRFC_ARATH (Q9M1X0) Ribosome recycling factor, chloroplast precursor (Ribosome-releasing factor, chloroplast) (RRF) (CpFrr) (RRFHCP)	108	14.20	30403.38	2 (2 0 0 0 0)
Q0WLF5 (Q0WLF5) Cytosolic O-acetylserine(Thiol)lyase	108	6.60	52980.07	2 (2 0 0 0 0)
MYRO_ARATH (P37702) Myrosinase precursor (EC 3.2.1.147) (Sinigrinase) (Thioglucosidase)	108	5.70	61093.96	2 (2 0 0 0 0)
Q9STM9 (Q9STM9) Stromal ascorbate peroxidase (EC 1.11.1.11)	108	10.80	40382.59	2 (2 0 0 0 0)
Q9S7C0 (Q9S7C0) Putative heat-shock protein; 37113-40399 (At1g79930/F19K16_11) (F18B13.1 protein)	108	4.10	91692.30	3 (3 0 0 0 0)
Q9LZY8 (Q9LZY8) Transketolase-like protein	107	9.50	81424.05	5 (5 0 0 0 0)
Q9LTX9 (Q9LTX9) Heat shock protein 70	107	5.40	76949.79	3 (3 0 0 0 0)
ACEA_ARATH (P28297) Isocitrate lyase (EC 4.1.3.1) (Isocitrase) (Isocitratase) (ICL)	106	12.50	64204.08	5 (5 0 0 0 0)
CAH2_ARATH (P42737) Carbonic anhydrase 2 (EC 4.2.1.1) (Carbonate dehydratase 2)	106	21.20	28326.37	4 (4 0 0 0 0)
MB33_ARATH (Q9SDM9) Myrosinase-binding protein-like At3g16400	106	6.20	51637.59	2 (2 0 0 0 0)
RS3A1_ARATH (Q9CAV0) 40S ribosomal protein S3a-1	106	12.60	29700.81	2 (2 0 0 0 0)
Q8L611 (Q8L611) Hypothetical protein At3g63460	105	2.50	119799.20	2 (2 0 0 0 0)

PMG2_ARATH (Q9M9K1) Probable 2,3-bisphosphoglycerate-independent phosphoglycerate mutase 2 (EC 5.4.2.1) (Phosphoglyceromutase 2) (BPG-independent PGAM 2) (PGAM-I 2)	105	10.70	60726.08	3 (3 0 0 0 0)
ATPB_ARATH (P19366) ATP synthase subunit beta (EC 3.6.3.14) (ATPase subunit beta) (ATP synthase F1 sector subunit beta)	104	17.90	53900.17	8 (8 0 0 0 0)
PSBB_ARATH (P56777) Photosystem II P680 chlorophyll A apoprotein (CP-47 protein)	104	7.70	56001.50	4 (4 0 0 0 0)
O80860 (O80860) FtsH protease (VAR2) (Zinc dependent protease)	104	4.50	74111.48	2 (2 0 0 0 0)
Q3E716 (Q3E716) Protein At4g38970	103	3.70	41318.48	2 (2 0 0 0 0)
NIR_ARATH (Q39161) Ferredoxin--nitrite reductase, chloroplast precursor (EC 1.7.7.1) (NiR)	103	8.40	65463.73	4 (4 0 0 0 0)
RL91_ARATH (P49209) 60S ribosomal protein L9-1	103	30.40	22003.87	5 (5 0 0 0 0)
CALM1_ARATH (P25854) Calmodulin-1/4 (CaM-1/4)	103	19.60	16719.90	2 (2 0 0 0 0)
RS81_ARATH (Q93VG5) 40S ribosomal protein S8-1	103	25.20	24978.31	4 (4 0 0 0 0)
RUBB_ARATH (P21240) RuBisCO large subunit-binding protein subunit beta, chloroplast precursor (60 kDa chaperonin subunit beta) (CPN-60 beta)	101	18.50	63769.78	11 (11 0 0 0 0)
Q9SZD6 (Q9SZD6) Hypothetical protein F19B15.90 (AT4g29060/F19B15_90) (Hypothetical protein AT4g29060)	101	3.60	103718.20	3 (3 0 0 0 0)
MLP34_ARATH (Q9SSK7) MLP-like protein 34	101	8.90	35547.78	2 (2 0 0 0 0)
RL72_ARATH (P60040) 60S ribosomal protein L7-2	100	20.70	28153.48	3 (3 0 0 0 0)
RL131_ARATH (P41127) 60S ribosomal protein L13-1 (Protein BBC1 homolog)	100	23.30	23752.20	4 (4 0 0 0 0)
RUBA_ARATH (P21238) RuBisCO large subunit-binding protein subunit alpha, chloroplast precursor (60 kDa chaperonin subunit alpha) (CPN-60 alpha)	100	21.70	62033.95	14 (14 0 0 0 0)
Q9LN13 (Q9LN13) T6D22.2	100	6.80	107106.60	4 (4 0 0 0 0)
RS131_ARATH (P59223) 40S ribosomal protein S13-1	99	27.20	17084.51	5 (5 0 0 0 0)
AAT2_ARATH (P46645) Aspartate aminotransferase, cytoplasmic isozyme 1 (EC 2.6.1.1) (Transaminase A)	99	6.90	44238.59	2 (2 0 0 0 0)
AVP1_ARATH (P31414) Pyrophosphate-energized vacuolar membrane proton pump 1 (EC 3.6.1.1) (Pyrophosphate-energized inorganic pyrophosphatase 1) (H(+)-PPase 1) (Vacuolar proton pyrophosphatase 1) (Vacuolar proton pyrophosphatase 3)	98	4.90	80767.95	2 (2 0 0 0 0)
EFTU_ARATH (P17745) Elongation factor Tu, chloroplast precursor (EF-Tu)	98	5.70	51597.68	2 (2 0 0 0 0)
Q9LDZ0 (Q9LDZ0) Heat shock protein 70 (Hsc70-5)	97	5.70	72945.84	3 (3 0 0 0 0)
Q9FM01 (Q9FM01) UDP-glucose dehydrogenase	97	8.50	53063.19	2 (2 0 0 0 0)
AAT5_ARATH (P46248) Aspartate aminotransferase, chloroplast precursor (EC 2.6.1.1) (Transaminase A)	96	13.00	49799.54	4 (4 0 0 0 0)
RS241_ARATH (Q9SS17) 40S ribosomal protein S24-1	96	27.10	15362.62	3 (3 0 0 0 0)
GCST_ARATH (O65396) Aminomethyltransferase, mitochondrial precursor (EC 2.1.2.10) (Glycine cleavage system T protein) (GCVT)	96	15.00	44416.79	4 (4 0 0 0 0)
RK121_ARATH (P36210) 50S ribosomal protein L12-1, chloroplast precursor (CL12-A)	96	12.00	20062.69	2 (2 0 0 0 0)
CATA1_ARATH (Q96528) Catalase-1 (EC 1.11.1.6)	96	5.10	56726.42	2 (2 0 0 0 0)
RS31_ARATH (Q9SIP7) 40S ribosomal protein S3-1	96	14.80	27501.74	4 (4 0 0 0 0)
GLNA2_ARATH (Q43127) Glutamine synthetase, chloroplast/mitochondrial precursor (EC 6.3.1.2) (Glutamate--ammonia ligase) (GS2)	96	6.70	47381.04	2 (2 0 0 0 0)
EF1B2_ARATH (Q9SCX3) Elongation factor 1-beta 2 (EF-1-beta 2) (Elongation factor 1B-alpha	95	12.90	24185.94	2 (2 0 0 0 0)

2) (eEF-1B alpha 2) (Elongation factor 1-beta' 2) (EF-1-beta' 2)				
Q9MAM6 (Q9MAM6) T25K16.8	95	4.90	75231.63	2 (2 0 0 0 0)
GRP7_ARATH (Q03250) Glycine-rich RNA-binding protein 7	94	15.30	16879.63	3 (3 0 0 0 0)
GSA1_ARATH (P42799) Glutamate-1-semialdehyde 2,1-aminomutase 1, chloroplast precursor (EC 5.4.3.8) (GSA 1) (Glutamate-1-semialdehyde aminotransferase 1) (GSA-AT 1)	94	13.50	50337.49	4 (4 0 0 0 0)
Q9ZWA9 (Q9ZWA9) F21M11.18 protein (Putative cruciferin 12S seed storage protein)	93	6.90	49643.90	2 (2 0 0 0 0)
Q9LU86 (Q9LU86) Peroxiredoxin Q-like protein (AT3g26060/MPE11_21)	93	7.40	23663.27	2 (2 0 0 0 0)
FL3H_ARATH (Q9S818) Naringenin,2-oxoglutarate 3-dioxygenase (EC 1.14.11.9) (Flavanone 3-hydroxylase) (Naringenin 3-dioxygenase) (FH3) (TRANSPARENT TESTA 6 protein)	93	12.00	40250.44	2 (2 0 0 0 0)
VDAC1_ARATH (Q9SRH5) Outer mitochondrial membrane protein porin 1 (Voltage-dependent anion-selective channel protein 1) (VDAC 1)	93	12.00	29276.28	2 (2 0 0 0 0)
RK1_ARATH (Q9LY66) 50S ribosomal protein L1, chloroplast precursor (CL1)	93	13.90	37608.85	3 (3 0 0 0 0)
VATA_ARATH (O23654) Vacuolar ATP synthase catalytic subunit A (EC 3.6.3.14) (V-ATPase subunit A) (Vacuolar proton pump subunit alpha) (V-ATPase 69 kDa subunit)	92	15.70	68768.93	7 (7 0 0 0 0)
RL121_ARATH (P50883) 60S ribosomal protein L12-1	92	31.30	17930.74	5 (5 0 0 0 0)
TRXF1_ARATH (Q9XFH8) Thioredoxin F-type 1, chloroplast precursor (TRX-F1)	92	14.00	19313.12	2 (2 0 0 0 0)
CAPP1_ARATH (Q9MAH0) Phosphoenolpyruvate carboxylase 1 (EC 4.1.1.31) (PEPCase 1) (PEPC 1) (AtPPC1)	92	9.60	110217.00	6 (6 0 0 0 0)
ATPA_ARATH (P56757) ATP synthase subunit alpha (EC 3.6.3.14) (ATPase subunit alpha) (ATP synthase F1 sector subunit alpha)	92	9.50	55294.12	4 (4 0 0 0 0)
Q9XI93 (Q9XI93) F7A19.2 protein (At1g13930/F16A14.27) (F16A14.14)	92	11.00	16153.81	2 (2 0 0 0 0)
Q9LZ66 (Q9LZ66) Sulphite reductase (AT5g04590)	92	5.30	71904.71	2 (2 0 0 0 0)
RL73_ARATH (P60039) 60S ribosomal protein L7-3	92	12.00	27922.40	2 (2 0 0 0 0)
PURA_ARATH (Q96529) Adenylosuccinate synthetase, chloroplast precursor (EC 6.3.4.4) (IMP--aspartate ligase) (AdSS) (AMPSase)	91	7.60	52931.41	2 (2 0 0 0 0)
TBB1_ARATH (P12411) Tubulin beta-1 chain (Beta-1 tubulin)	91	12.30	50184.89	6 (6 0 0 0 0)
CATA3_ARATH (Q42547) Catalase-3 (EC 1.11.1.6)	90	13.80	56659.66	5 (5 0 0 0 0)
KAD1_ARATH (O82514) Adenylate kinase 1 (EC 2.7.4.3) (ATP-AMP transphosphorylase 1)	90	9.80	26915.01	2 (2 0 0 0 0)
O81288 (O81288) T14P8.5 (Hypothetical protein AT4g02450)	90	9.50	27925.15	2 (2 0 0 0 0)
NACA4_ARATH (Q9SZY1) Nascent polypeptide-associated complex subunit alpha-like protein 4 (NAC-alpha-like protein 4) (Alpha-NAC-like protein 4)	89	13.70	23157.29	3 (3 0 0 0 0)
CB4A_ARATH (Q07473) Chlorophyll a-b binding protein CP29.1, chloroplast precursor (LHCII protein 4.1) (LHCB4.1)	89	7.90	31119.98	2 (2 0 0 0 0)
GLTB1_ARATH (Q9ZNZ7) Ferredoxin-dependent glutamate synthase 1, chloroplast precursor (EC 1.4.7.1) (Fd-GOGAT 1)	88	3.20	179805.80	4 (4 0 0 0 0)
RS23_ARATH (P49688) 40S ribosomal protein S2-3	88	9.80	30859.34	2 (2 0 0 0 0)
AAT1_ARATH (P46643) Aspartate aminotransferase, mitochondrial precursor (EC 2.6.1.1) (Transaminase A)	87	8.40	47727.02	2 (2 0 0 0 0)
RL81_ARATH (P46286) 60S ribosomal protein L8-1 (60S ribosomal protein L2) (Protein EMBRYO DEFECTIVE 229)	87	9.70	27841.81	3 (3 0 0 0 0)
MTHR1_ARATH (Q9SE60) Methylenetetrahydrofolate reductase 1 (EC 1.5.1.20) (AtMTHFR1)	86	9.00	66246.48	4 (4 0 0 0 0)
AMPL1_ARATH (P30184) Leucine aminopeptidase 1 (EC 3.4.11.1) (LAP 1) (Leucyl	86	6.30	54475.07	2 (2 0 0 0 0)

aminopeptidase 1) (Proline aminopeptidase 1) (EC 3.4.11.5) (Prolyl aminopeptidase 1)				
RL51_ARATH (Q8LBI1) 60S ribosomal protein L5-1	84	9.60	34336.86	4 (4 0 0 0 0)
PMG1_ARATH (O04499) 2,3-bisphosphoglycerate-independent phosphoglycerate mutase 1 (EC 5.4.2.1) (Phosphoglyceromutase 1) (BPG-independent PGAM 1) (PGAM-I 1)	84	9.30	60541.82	4 (4 0 0 0 0)
SPD1_ARATH (Q9ZUB3) Spermidine synthase 1 (EC 2.5.1.16) (Putrescine aminopropyltransferase 1) (SPDSY 1)	83	12.30	36530.06	4 (4 0 0 0 0)
Q9LSB4 (Q9LSB4) Arabidopsis thaliana genomic DNA, chromosome 3, P1 clone: MVC8 (Hypothetical protein At3g15950)	82	2.80	84965.04	2 (2 0 0 0 0)
MDAR3_ARATH (Q9LFA3) Probable monodehydroascorbate reductase, cytoplasmic isoform 3 (EC 1.6.5.4) (MDAR 3)	82	11.10	46458.19	3 (3 0 0 0 0)
RL7A1_ARATH (P49692) 60S ribosomal protein L7a-1	82	8.90	29111.27	2 (2 0 0 0 0)
O65390 (O65390) F12F1.24 protein (Putative aspartic proteinase) (At1g11910/F12F1_24)	81	6.50	54578.59	2 (2 0 0 0 0)
Q9SUC7 (Q9SUC7) Hypothetical protein T13K14.10 (Hypothetical protein AT4g20850)	81	3.50	154078.90	4 (4 0 0 0 0)
PEPCK_ARATH (Q9T074) Phosphoenolpyruvate carboxykinase [ATP] (EC 4.1.1.49) (PEP carboxykinase) (Phosphoenolpyruvate carboxylase) (PEPCK)	81	6.60	73358.79	3 (3 0 0 0 0)
ADT1_ARATH (P31167) ADP,ATP carrier protein 1, mitochondrial precursor (ADP/ATP translocase 1) (Adenine nucleotide translocator 1) (ANT 1)	81	9.70	41449.37	4 (4 0 0 0 0)
RLA01_ARATH (O04204) 60S acidic ribosomal protein P0-1	80	8.80	33645.72	3 (3 0 0 0 0)
TPIC_ARATH (Q9SKP6) Triosephosphate isomerase, chloroplast precursor (EC 5.3.1.1) (TIM) (Triose-phosphate isomerase)	80	12.10	33325.16	3 (3 0 0 0 0)
Q8H1E2 (Q8H1E2) NADP-dependent malate dehydrogenase	80	12.00	48285.84	3 (3 0 0 0 0)
UMP6_ARATH (Q9SX77) Unknown protein At1g47420, mitochondrial precursor	79	12.80	28088.21	2 (2 0 0 0 0)
SERC_ARATH (Q96255) Phosphoserine aminotransferase, chloroplast precursor (EC 2.6.1.52) (PSAT)	79	9.50	47329.13	3 (3 0 0 0 0)
IPYR1_ARATH (Q9LXC9) Soluble inorganic pyrophosphatase 1, chloroplast precursor (EC 3.6.1.1) (Pyrophosphate phospho-hydrolase 1) (PPase 1) (Chloroplast inorganic pyrophosphatase 1)	79	15.00	33359.09	4 (4 0 0 0 0)
PSBD_ARATH (P56761) Photosystem II D2 protein (Photosystem Q(A) protein) (PSII D2 protein)	78	9.10	39390.86	3 (3 0 0 0 0)
TRXH3_ARATH (Q42403) Thioredoxin H-type 3 (TRX-H-3)	78	28.00	13100.78	3 (3 0 0 0 0)
CB26_ARATH (Q9XF89) Chlorophyll a-b binding protein CP26, chloroplast precursor (Light-harvesting complex II protein 5) (LHCB5) (LHCIIc)	77	15.70	30137.68	2 (2 0 0 0 0)
PSA2A_ARATH (O23708) Proteasome subunit alpha type 2-A (EC 3.4.25.1) (20S proteasome alpha subunit B) (Proteasome component 3)	77	11.50	25685.28	3 (3 0 0 0 0)
FLA1_ARATH (Q9FM65) Fasciclin-like arabinogalactan protein 1 precursor	77	8.00	44821.18	2 (2 0 0 0 0)
Q8L784 (Q8L784) Cytoplasmic aconitate hydratase (At2g05710)	76	2.60	108132.80	2 (2 0 0 0 0)
Q9C522 (Q9C522) ATP citrate lyase, putative; 3734-7120 (Putative ATP citrate lyase) (ATP citrate lyase, putative; 38389-41775)	75	7.70	65772.08	3 (3 0 0 0 0)
PREP1_ARATH (Q9LJL3) Presequence protease 1, chloroplast/mitochondrial precursor (EC 3.4.24.-) (PreP 1) (AtPreP1) (Zinc metalloprotease 1) (AtZnMP1)	75	5.10	120938.80	4 (4 0 0 0 0)
SYGM1_ARATH (O23627) Glycyl-tRNA synthetase 1, mitochondrial precursor (EC 6.1.1.14) (Glycine--tRNA ligase 1) (GlyRS 1)	74	5.20	81892.14	3 (3 0 0 0 0)
RS201_ARATH (P49200) 40S ribosomal protein S20-1	74	10.50	13869.55	2 (2 0 0 0 0)
GLTB2_ARATH (Q9T0P4) Ferredoxin-dependent glutamate synthase 2, chloroplast precursor	73	1.90	177648.80	2 (2 0 0 0 0)

(EC 1.4.7.1) (Fd-GOGAT 2)				
TL29_ARATH (P82281) Putative L-ascorbate peroxidase, chloroplast precursor (EC 1.11.1.11) (Thylakoid lumenal 29 kDa protein) (TL29) (P29)	73	10.60	37910.56	2 (2 0 0 0 0)
Q9FI56 (Q9FI56) ATP-dependent Clp protease, ATP-binding subunit (ATP-dependent Clp protease) (AT5g50920/K3K7_7)	73	4.30	103388.30	3 (3 0 0 0 0)
RS191_ARATH (Q9SGA6) 40S ribosomal protein S19-1	73	28.70	15818.36	3 (3 0 0 0 0)
R27A2_ARATH (Q9LR33) 60S ribosomal protein L27a-2	72	21.20	16281.92	2 (2 0 0 0 0)
G3PA_ARATH (P25856) Glyceraldehyde-3-phosphate dehydrogenase A, chloroplast precursor (EC 1.2.1.13) (NADP-dependent glyceraldehydephosphate dehydrogenase subunit A)	72	21.00	42463.04	8 (8 0 0 0 0)
PSB6_ARATH (Q8LD27) Proteasome subunit beta type 6 precursor (EC 3.4.25.1) (Proteasome subunit beta type 1) (20S proteasome beta subunit A-1) (Proteasome component CD)	72	11.20	25135.55	2 (2 0 0 0 0)
HEM6_ARATH (Q9LR75) Coproporphyrinogen III oxidase, chloroplast precursor (EC 1.3.3.3) (Coproporphyrinogenase) (Coprogen oxidase) (LESION INITIATION 2 protein)	72	12.40	43768.70	3 (3 0 0 0 0)
MDARP_ARATH (P92947) Monodehydroascorbate reductase, chloroplast precursor (EC 1.6.5.4) (MDAR)	72	12.40	53268.64	5 (5 0 0 0 0)
Q9LUJ7 (Q9LUJ7) Arabidopsis thaliana genomic DNA, chromosome 3, P1 clone: MWI23	71	6.20	55028.96	3 (3 0 0 0 0)
GSTF3_ARATH (P42761) Glutathione S-transferase ERD13 (EC 2.5.1.18) (GST class-phi)	71	11.20	24083.72	2 (2 0 0 0 0)
AAT3_ARATH (P46644) Aspartate aminotransferase, chloroplast precursor (EC 2.6.1.1) (Transaminase A)	71	5.30	48923.47	2 (2 0 0 0 0)
Q9LVK7 (Q9LVK7) Dihydrolipoamide acetyltransferase	70	6.40	59651.92	3 (3 0 0 0 0)
GWD1_ARATH (Q9SAC6) Alpha-glucan water dikinase 1, chloroplast precursor (EC 2.7.9.4) (Protein starch-related R1) (Protein starch excess 1)	70	2.10	156484.20	2 (2 0 0 0 0)
Q9CA67 (Q9CA67) Geranylgeranyl reductase; 47568-49165 (Geranylgeranyl reductase) (At1g74470/F1M20_15)	70	4.90	51804.73	2 (2 0 0 0 0)
Q9FLP1 (Q9FLP1) Trigger factor-like protein	68	10.90	65119.73	6 (6 0 0 0 0)
RL101_ARATH (Q93VT9) 60S ribosomal protein L10-1	68	10.50	24901.16	2 (2 0 0 0 0)
MDHG1_ARATH (Q9ZP05) Malate dehydrogenase, glyoxysomal precursor (EC 1.1.1.37) (mbNAD-MDH)	67	12.10	37345.63	3 (3 0 0 0 0)
KPPR_ARATH (P25697) Phosphoribulokinase, chloroplast precursor (EC 2.7.1.19) (Phosphopentokinase) (PRKase) (PRK)	66	8.60	44435.66	2 (2 0 0 0 0)
ENPL_ARATH (Q9STX5) Endoplasmin homolog precursor (GRP94 homolog) (Protein SHEPHERD) (HSP90-like protein 7)	66	6.10	94145.79	4 (4 0 0 0 0)
ASSY_ARATH (Q9SZX3) Argininosuccinate synthase, chloroplast precursor (EC 6.3.4.5) (Citrulline--aspartate ligase)	66	6.10	53811.71	2 (2 0 0 0 0)
Q9FFR3 (Q9FFR3) 6-phosphogluconate dehydrogenase (At5g41670/MBK23_20)	66	7.40	53284.25	4 (4 0 0 0 0)
Q8VZW6 (Q8VZW6) Putative elongation factor P (EF-P)	66	5.50	26211.33	2 (2 0 0 0 0)
GSH1_ARATH (P46309) Glutamate--cysteine ligase, chloroplast precursor (EC 6.3.2.2) (Gamma-glutamylcysteine synthetase) (Gamma-ECS) (GCS)	65	5.20	58524.83	2 (2 0 0 0 0)
PDIA6_ARATH (O22263) Probable protein disulfide-isomerase A6 precursor (EC 5.3.4.1) (P5)	64	13.00	39472.40	4 (4 0 0 0 0)
HEM3_ARATH (Q43316) Porphobilinogen deaminase, chloroplast precursor (EC 2.5.1.61) (PBG) (Hydroxymethylbilane synthase) (HMBS) (Pre-uroporphyrinogen synthase)	63	11.30	41017.46	5 (5 0 0 0 0)
RSSA1_ARATH (Q08682) 40S ribosomal protein Sa-1 (p40) (Laminin receptor homolog)	63	12.10	32270.80	4 (4 0 0 0 0)
RH3_ARATH (Q8L7S8) DEAD-box ATP-dependent RNA helicase 3 (EC 3.6.1.-) (Protein EMBRYO DEFECTIVE 1138)	63	4.30	81106.97	2 (2 0 0 0 0)

RR20_ARATH (Q9ASV6) 30S ribosomal protein S20, chloroplast precursor	62	10.40	21812.58	2 (2 0 0 0 0)
RS171_ARATH (P49205) 40S ribosomal protein S17-1	62	8.60	15905.53	2 (2 0 0 0 0)
CD48A_ARATH (P54609) Cell division control protein 48 homolog A (AtCDC48a)	62	5.40	89337.55	4 (4 0 0 0 0)
Q8RY71 (Q8RY71) Putative jasmonate inducible protein	62	8.40	28507.98	2 (2 0 0 0 0)
Q0WLB5 (Q0WLB5) Hypothetical protein At3g08530	61	1.70	193147.10	2 (2 0 0 0 0)
Q9LZ23 (Q9LZ23) Hypothetical protein T1E3_100 (Hypothetical protein At5g04740) (Genomic DNA, chromosome 5, P1 clone:MUK11)	60	7.60	33221.14	2 (2 0 0 0 0)
MB22_ARATH (O80950) Myrosinase-binding protein-like At2g39310	60	6.80	50431.43	2 (2 0 0 0 0)
Q0WLE7 (Q0WLE7) Hypothetical protein At1g43170	60	6.70	63314.81	3 (3 0 0 0 0)
RS41_ARATH (Q93VH9) 40S ribosomal protein S4-1	59	10.00	29784.22	2 (2 0 0 0 0)
FRI1_ARATH (Q39101) Ferritin-1, chloroplast precursor (EC 1.16.3.1) (AtFer1)	59	9.80	28160.19	2 (2 0 0 0 0)
VDAC2_ARATH (Q9SMX3) Outer mitochondrial membrane protein porin 2 (Voltage-dependent anion-selective channel protein 2) (VDAC 2)	59	11.70	29061.97	2 (2 0 0 0 0)
Q94AR8 (Q94AR8) Hypothetical protein At4g13430	58	8.40	54978.63	3 (3 0 0 0 0)
PSBS_ARATH (Q9XF91) Photosystem II 22 kDa protein, chloroplast precursor (CP22)	58	11.70	27990.28	4 (4 0 0 0 0)
Q9LF96 (Q9LF96) Peroxiredoxin-like protein	58	12.40	24668.89	4 (4 0 0 0 0)
SBP_ARATH (O23264) Putative selenium-binding protein	58	5.50	54022.94	2 (2 0 0 0 0)
RS162_ARATH (Q9M8X9) 40S ribosomal protein S16-2	56	8.20	16575.84	3 (3 0 0 0 0)
R13A1_ARATH (Q9SFU1) 60S ribosomal protein L13a-1	56	12.60	23450.85	2 (2 0 0 0 0)
TCTP_ARATH (P31265) Translationally-controlled tumor protein homolog (TCTP)	55	11.30	18898.49	2 (2 0 0 0 0)
Q9SA73 (Q9SA73) T5I8.3 protein (At1g30580)	54	5.30	44443.23	2 (2 0 0 0 0)
MDAR4_ARATH (Q93WJ8) Probable monodehydroascorbate reductase, cytoplasmic isoform 4 (EC 1.6.5.4) (MDAR 4)	53	7.10	47450.38	2 (2 0 0 0 0)
Q93ZN9 (Q93ZN9) AT4g33680/T16L1_170 (Aminotransferase AGD2) (Hypothetical protein At4g33680)	53	9.50	50364.03	3 (3 0 0 0 0)
Q9ZUG8 (Q9ZUG8) Putative NADP-dependent glyceraldehyde-3-phosphate dehydrogenase (At2g24270/F27D4.18)	53	5.40	53026.29	2 (2 0 0 0 0)
RL241_ARATH (Q42347) 60S ribosomal protein L24-1	53	15.20	18838.44	3 (3 0 0 0 0)
CFI_ARATH (P41088) Chalcone--flavonone isomerase (EC 5.5.1.6) (Chalcone isomerase) (TRANSPARENT TESTA 5 protein)	52	11.80	26578.85	2 (2 0 0 0 0)
Q9C9W5 (Q9C9W5) Hydroxypyruvate reductase (HPR); 50972-48670 (Hydroxypyruvate reductase) (HPR) (Putative hydroxypyruvate reductase HPR)	52	7.00	42220.93	2 (2 0 0 0 0)
Q42601 (Q42601) Carbamoyl phosphate synthetase large chain (EC 6.3.4.16) (At1g29900/F1N18_6) (CarB)	52	2.00	129874.60	2 (2 0 0 0 0)
IF35_ARATH (O04202) Eukaryotic translation initiation factor 3 subunit 5 (eIF-3 epsilon) (eIF3 p32 subunit) (eIF3f)	51	7.20	31842.28	2 (2 0 0 0 0)
Q8L7C9 (Q8L7C9) 2,4-D-inducible glutathione S-transferase, putative	51	9.70	24990.59	2 (2 0 0 0 0)
EFTM_ARATH (Q9ZT91) Elongation factor Tu, mitochondrial precursor	50	9.30	49378.63	3 (3 0 0 0 0)
14333_ARATH (P42644) 14-3-3-like protein GF14 psi (General regulatory factor 3) (14-3-3-like protein RCI1)	48	16.50	28589.27	3 (3 0 0 0 0)
Q94K56 (Q94K56) Putative chloroplast GrpE protein	47	6.50	35471.93	2 (2 0 0 0 0)

CH60A_ARATH (P29197) Chaperonin CPN60, mitochondrial precursor (HSP60)	46	7.10	61242.33	3 (3 0 0 0 0)
RL171_ARATH (Q93VI3) 60S ribosomal protein L17-1	46	10.80	19884.52	2 (2 0 0 0 0)
Q3EAL7 (Q3EAL7) Protein At3g51800	45	5.50	43916.75	2 (2 0 0 0 0)
O80576 (O80576) Similar to late embryogenesis abundant proteins (At2g44060)	44	14.20	36013.78	3 (3 0 0 0 0)
RS91_ARATH (Q9LXG1) 40S ribosomal protein S9-1	44	13.60	23022.21	3 (3 0 0 0 0)
CYF_ARATH (P56771) Apocytochrome f precursor	43	9.10	35334.77	2 (2 0 0 0 0)
VATB_ARATH (P11574) Vacuolar ATP synthase subunit B (EC 3.6.3.14) (V-ATPase B subunit) (Vacuolar proton pump B subunit) (V-ATPase 57 kDa subunit)	38	7.60	54073.68	3 (3 0 0 0 0)
ATPO_ARATH (Q96251) ATP synthase delta chain, mitochondrial precursor (EC 3.6.3.14) (Oligomycin sensitivity conferral protein) (OSCP)	37	9.20	26304.98	2 (2 0 0 0 0)
CISY4_ARATH (P20115) Citrate synthase 4, mitochondrial precursor (EC 2.3.3.1)	32	4.20	52749.16	2 (2 0 0 0 0)
MSRA_ARATH (P54150) Peptide methionine sulfoxide reductase (EC 1.8.4.6) (Protein-methionine-S-oxide reductase) (Peptide Met(O) reductase)	30	8.50	28626.33	2 (2 0 0 0 0)

VITA

VITA

Nileshwari Vaghela was born on 1st July, 1980 in Baroda, Gujarat, India. She received her Bachelor of Engineering in Civil (Environmental) Engineering in May 2001 from M.S. University, Baroda, India. In August 2004, she began her study for the Master of Science in Biological Sciences at Virginia Tech. Currently she is working at Dana Farber Cancer Institute, Harvard Medical School, Boston.