ABSTRACT

ZHICHANG, YANG. Absolute Quantification of Enzymes in the Monolignol Biosynthetic Pathway under Transgenic Effect by Mass Spectrometry. (Under the direction of Dr. David C. Muddiman).

Mass spectrometry has been viewed as a very highly specific and sensitive technique. The combination of RP-HPLC separation technology with mass spectrometry has been a very big breakthrough in the analytical technology. Mass spectrometry based proteomics possesses the trait of fast acquisition rate, high mass resolving power, low detection limit, and high mass measurement accuracy. Together with high performance separations, mass spectrometry can quantify thousands of proteins over a wide dynamic range within complex sample matrix. Mass spectrometry can also provide targeted protein absolute quantification based on the signal intensity and the quantity of internal standard. The work presented in the dissertation focuses on the target protein quantification with well-developed protein cleavage isotope dilution mass spectrometry method.

Prior to carrying out quantification of the enzymes in the monolignol biosynthetic pathway, the two optimized protein digestion and fractionation methods were first compared to see which way is more efficient and beneficial to identify surrogate peptide. In gel digestion and filter aided sample preparation (FASP) stage tip digestion are the most common digestion and fractionation methods. The comparison of these two methods was made under each optimized condition on the same quantity of *P trichocarpa* stem differentiating xylem (SDX) tissue extract. At the whole proteome level and more importantly, target protein level, the comparison result concluded that FASP method was more effective and practically easier to process and achieved higher protein recovery, allowing for the identification of unique peptides for the downstream absolute quantification.

After concluding that FASP was the method of choice for all the sample preparation for enzyme absolute quantification, the absolute quantification was processed on 4 bathes of transgenic *P trichocarpa* SDX extraction. Together with the transcript abundance and the protein abundance, a comprehensive understanding of the monolignol biosynthetic

pathway, the prediction and control of the structure and quantity of lignin under transgenic stress is one of the major goals of the project. The correspondence of the gene transcript level and the protein abundance is demonstrated for each gene construct. Gene family compensation or a loop effect was also discovered at not only gene transcript level but also at protein abundance level.

© Copyright 2013 by Zhichang Yang

All Rights Reserved

Absolute Quantification of Enzymes in the Monolignol Biosynthetic Pathway under Transgenic Effect by Mass Spectrometry

by
Zhichang Yang

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Master of Science

Chemistry

Raleigh, North Carolina

2013

APPROVED BY

_______________________________ _______________________________

David C. Muddiman
Professor, Chemistry
Committee Chair

Vincent L. Chiang
Professor, Forest Biotechnology

____

Gufeng Wang
Assistant Professor, Chemistry

BIOGRAPHY

Zhichang Yang was born in Dali, Yunnan, China on August 11th, 1988 to Yun Yang and Jianmei Lou and was the second of two brothers. Zhichang kept excellent academic record during high school and attended one of the best universities in China, Nankai University. Zhichang found himself very interested in analytical chemistry in Nankai and decided to pursue a further research career in this area. So Zhichang chose to study in United States as an graduate student and joined the biological mass spectrometry laboratory under the leadership of Dr. David C. Muddiman.

ACKNOWLEDGEMENTS

There are so many people I would love to acknowledge through my whole graduate school life. The first and the most important person I will sincerely express my gratitude is my advisor Dr. David C. Muddiman. You can see in him the marks of passion, motivation, superior erudition and concentration in science. Dave is also a person with amazing kindness, compatibility and lenience. Not only has he showed his great support helping me learn how to be a conscientious scientist, he also taught me how to become a good person caring about people and the world around us by his words and deeds.

I will also thank all of my colleagues in W.M. Keck FT-ICR laboratory. As an international student first coming to United States to study, I have met so many "Maladaptation" during my graduate school life. My colleagues helped me overcome all of the cultural differences and language barrier and supported me in research with their great patience.

TABLE OF CONTENTS

LIST OF FIGURES vi

LIST OF FIGURES vii

CHAPTER 1 Introduction 1

1.1 Lignin and Monolignol Biosynthetic Pathway 1

1.2 Mass Spectrometry and Proteomics 3

1.2.1 Mass Spectrometry Preface 3

1.2.2 Proteomics 3

1.2.3 Fractionation 6

1.2.4 Reversed Phase liquid chromatography (RP-LC) 11

1.2.5 Electrospray Ionization (ESI) 11

1.2.6 Linear Trap Quadrupole-Orbitrap (LTQ- Orbitrap) 13

1.2.7 Protein Cleavage Isotope Dilution Mass Spectrometry (PCIDMS) 14

1.2.8 Synopsis of Completed Research 18

1.3 References 20

CHAPTER 2 Comparison of In-gel digestion and FASP Stage-tip digestion using LTQ-Orbitrap Global Proteomic Study on Stem Differentiating Xylem Tissue Extraction from *Populus Trichocarpa* 26

2.1 Introduction 26

2.2 Materials and Methods 27

2.2.1 Xylem protein extraction 28

2.2.2 Optimized in Gel digestion 28

2.2.3 Optimized FASP Stage Tip Digestion 29

2.2.4 NanoLC-MS/MS 30
2.3 Results and Discussion 32
2.3.1 Digestion Work Flow 32
2.3.2 Total Proteome 33
2.3.3 Target Protein 38
2.4 Conclusion 40
2.5 References 41
CHAPTER 3 Absolute Quantification of Enzyme Component in the Monolignol Biosynthetic Pathway of the Transgenic Gene Perturbed Wood *Populus Trichocarpa* 43
3.1 Introduction 43
3.2 Materials and Methods 44
3.2.1 Xylem Protein Extraction 45
3.2.2 AmiRNA and RNAi Gene Knock Down 45
3.2.3 mRNA Abundance (transcript level) Measurement Using qRT-PCR 46
3.2.4 Stable Isotope-labeled Peptide Standard 46
3.2.5 Filter Aided Sample Preparation 46
3.2.6 LC-SRM Quantitative Analysis 47
3.2.7 Data Analysis 48
3.3 Results and discussion 48
3.3.1 Surrogate Peptide Selection 48
3.3.2 Transgenic Gene Constructs 51
3.3.3 Absolute Quantification on the Target Protein of the Transgenic Tree 52
3.4 Conclusion 60
3.5 References 61

LIST OF TABLES

Table 3.1 Surrogate Peptide List 50

Table 3.2 Student-t Test Result 54

LIST OF FIGURES

Figure 1.1 Monolignol Biosynthetic Pathway 2

Figure 1.2 Selected Reaction Monitor (SRM) 6

Figure 1.3 1D SDS-PAGE Stained with Coomassie Blue 8

Figure 1.4 Strong Cation Exchange and Strong Anion Exchange 9

Figure 1.5 LTQ-Orbitrap 14

Figure 1.6 Protein Cleavage Isotope Dilution Mass Spectrometry 16

Figure 2.1 In-gel and FASP-ST Digestion-Fractionation Scheme 32

Figure 2.2 Protein Identified From In-gel and FASP Fractions 33

Figure 2.3 Protein Distribution Based on Spectral Count 34

Figure 2.4 Spectral Count Plot of In-gel against FASP-ST 35

Figure 2.5 Sequence Coverage and Peptide Scatter Plot 36

Figure 2.6 Comparison of In-gel and FASP on the Target Proteins 39

Figure 3.1 CAD1 Gene Transcript Level and Protein Abundance 53

Figure 3.2 Protein Abundance Percentage Difference at Level-2 56

Figure 3.3 Loop Effect within PAL Family 59

CHAPTER 1 Introduction

1.1 Lignin and Monolignol Biosynthetic Pathway

Lignin is one of the most abundant natural substances in the world, only second to polysaccharides.[1] As a main component to the vascular wood, lignin plays a significant role in internal water, nutrient and metabolite transportation. Lignin's wide distribution[2] within the cell and between cell walls makes it able to protect the wood from bending, impact, and compression. Practically and commercially, the study of lignin could allow for technological progression and increased energy efficiency in the area of papermaking, biofuel[3] and forage production.[4]

Lignin, as polymer, does not have a defined structure or composition as an amorphous polyphenolic material. Lignin is polymerized from 3 phenylpropanoid monomers which all derives from a pathway called monolignol biosynthetic pathway[5-6] (**Figure 1.1**). These 3 phenylpropanoid monomers are: p-coumaryl alcohol (**H**), coniferyl alcohol (**G**), and sinapyl alcohol (**S**) (**Figure 1.1**). Generally during lignin deposition, monolignol will form in the cytoplasm and then translocate to apoplast and is eventually polymerized into lignin[1]. The monolignol biosynthesis and polymerization has been well elucidated at least in angiosperms[7]. Phenylalanine, as the common starting material in the pathway will go through a series of enzymatic reactions and eventually form these 3 monomers. In grasses, the softwood lignin is mainly composed of coniferyl alcohol and less but still significant amount of p-coumaryl alcohol[5]. In hardwood plants, like *Populus trichocarpa*, lignin will be polymerized mainly from G and S monomers with a slight amount of H monolignols[8]. Since lignin will limit the access to the cell wall polysaccharide, it will negatively affect the efficiency of converting biomass to biofuel, forage digestion by livestock and paper producing productivity[9]. So far lignin has been one of the most intensively studied subjects, and it is important to understand how lignin is formed and to determine if there is a method to manipulate the content and composition of lignin. A comprehensive understanding of

the monolignol biosynthetic pathway requires quantitative estimates at each level of the system, from the transcripts to the metabolites.

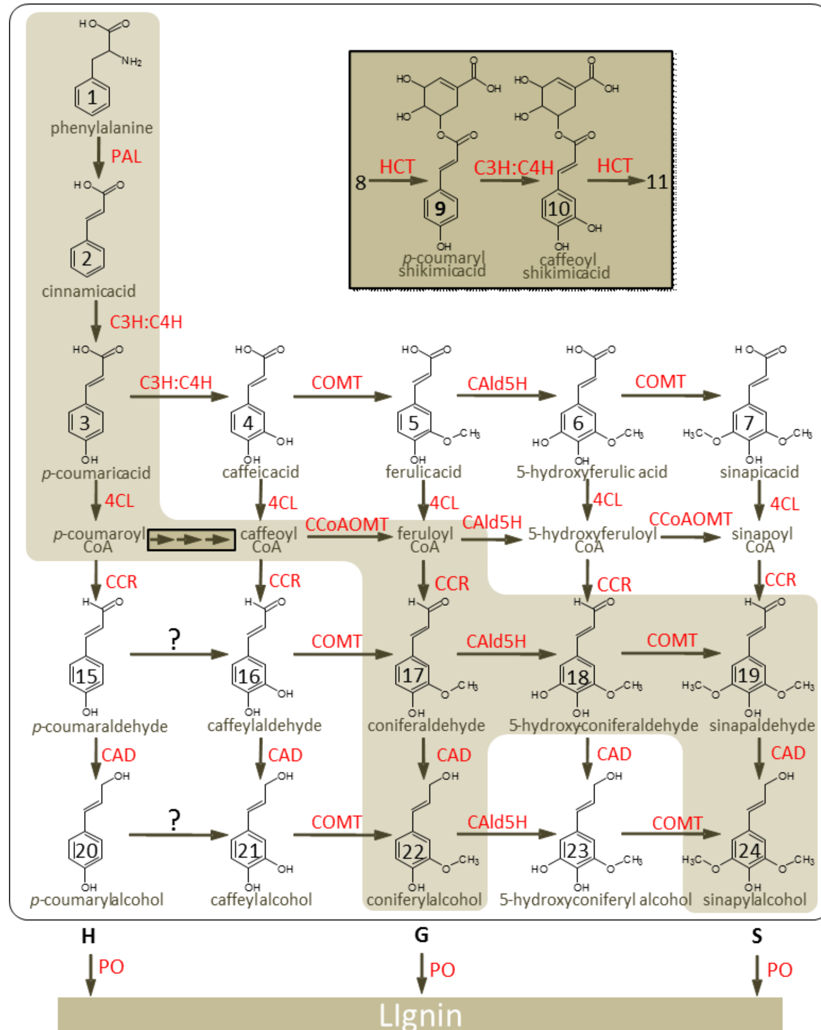


Figure 1.1 - Monolignol biosynthetic pathway leading to the synthesis of the 3 phenylpropanoid monomers are: p-coumaryl alcohol (**H**), coniferyl alcohol (**G**), and sinapyl alcohol (**S**). The highlighted pathway is the pathway known as the pathway of monolignol synthesis of *Populus trichocarpa.*

1.2 Mass Spectrometry and Proteomics

1.2.1 Mass Spectrometry Preface

Mass spectrometry (MS) is a method measuring the mass to charge ratio (*m/z*) of a specific ion in the gas phase[10]. Spectra derived from mass spectrometry analysis can provide us rich information including the mass, structure, and chemical properties of the compound which can help identify, quantify and characterize the target of interest[11].

Mass spectrometry is mainly composed of an ionization source, a mass analyzer and a detector. Generally, a compound or a complex biological sample can be analyzed by MS through the following process: First, the compound will be ionized in the gas phase; during ionization, the compound will be transferred into electrically charged particle. All the ions will be directed to the mass analyzer through ion optics. In most of the cases, the ion will collide with collision gas such as N_2 or He to form product ions[12]. Different ions will have different *m/z* or different fragmentation pattern. All these characteristics including mass, charge, and fragmentation pattern will be able to determine the behavior of the compound in the mass analyzer. And hence, the detector will be able to acquire signal which could define or quantify the compound of interest.

1.2.2 Proteomics

Proteomics can be simply defined as a systematic study on a large-scale of proteins within complex biological matrix. Proteomics commits widely on the study of protein, protein post translational modification and protein interactions, Including identification, characterization and quantification.[13] At the total proteome level, tandem mass spectrometry is the most used instrumental technology specifically when performing bottom-up proteomics[14]. In Bottom-up proteomics, proteins extracted from tissues or cells need to be digested into peptides, fractionation either at the protein level or digested peptide level can be used to

help reduce the complexity of the sample loaded on the instrument. Digested peptides will be loaded on the nano-Flow-Liquid-Chromatography platform first for peptide separation. Then the peptides will be ionized and transferred to the mass spectrometer for further analysis. The peptide fragment spectra are recorded to help map back the sequence of the peptide and identify the protein, characterize and locate the amino acids modification in the sequence; the peptide ion signal intensity could help acquire quantification information.

1.2.2.1 Global Proteomics

The method is suited for the identification of the protein components without much prior protein knowledge of the sample. Tandem mass spectrometry[15] is always applied in the discovery based proteomics. Only precursor MS analysis is not capable provide enough detail locking down the protein identity within complex sample matrix. Contamination derived from the isobaric species can seriously affect the identification of the proteome. Tandem mass spectrometry can be easily understood as multiple mass analysis accomplished with individual mass spectrometer elements separated in time or in space. Data dependent acquisition[16] as the data collection mode during tandem MS analyzing, can be described as: peptide precursor ion (e.g., intact molecular ion) will be detected in MS^1scan; the mass of the precursor ion will be recorded. The MS^1 scan can also be called survey scan or full scan which is always relatively high resolving power (e.g. 100,000) trying to narrow down the identity of the intact precursor ions. The most intense ion detected in MS^1 scan will be isolated and fragmented, the fragment spectra will then be obtained through MS/MS scan on the product ions derived from this most intense ion. After the MS/MS spectra is available, the mass of this most intense ion will be put into an exclusion list temporarily so this mass will not be reexamined for a specific period. This is called dynamic exclusion. Then MS/MS spectra of the second most intense ion will be triggered.

The mass of the precursor ion and the fragment spectra of the product ion can compose together help to identify proteins using statistical bioinformatics platform.

1.2.2.2 Targeted Proteomics

With the predetermined information of the protein(s) of interest, targeted proteomics allows for specific and sensitive quantification proteins in wide dynamic range when the identity and sequence of a certain limited amount of protein of interest is already available. The main technology targeted proteomics relies on is selected reaction monitoring (SRM) [17] also referred to multiple reaction monitoring (MRM). Three linear quadrupoles are coupled back to back to composite a triple quadrupole instrument[18], and the SRM scan mode is mainly implemented by triple quadrupole (**Figure 1.2**). Specific precursor ion mass and specific product ion mass is selected and fixed in SRM mode[18], only precursor and product ion match the mass with a range of tolerance will be detected and obtain corresponding spectra. Since SRM strategy is very sensitive and specific, it could be applied on the absolute quantification of the pre-known limited amount of target protein within biological matrix.

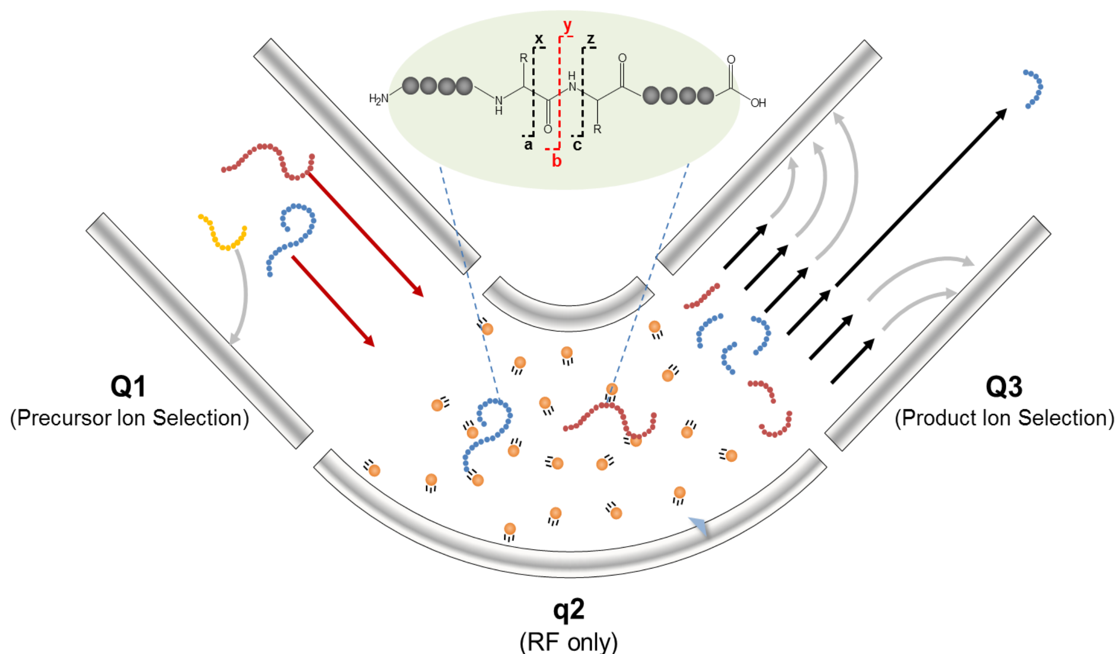


Figure 1.2 - SRM scan mode carried out on triple quadrupole. The precursor mass, product mass are already fixed on the Q1 and Q3 quadrupoles. With a certain tolerance, only precursor ions at the settled precursor mass of Q1 can pass through Q1. The passed precursor ions will fragment in non-mass selective collision cell q2 and produce product ions. Only product ion at the settled product mass of Q3 can pass through Q3. The specific precursor to product set is also called a transition.

1.2.3 Fractionation

The total protein tissue or cell extraction sample for proteomic study is complex composed of tens of thousands of proteins. There are vast array of unique proteins, including isoforms, SNPs and post translational modifications. The protein abundance also varies significantly within one single sample. For example, in human plasma, the protein abundance is spanning 10-12 orders of magnitude[19]. High abundant proteins limit the depth of the proteome that can be measured especially if no fractionation or protein enrichment is carried out. Furthermore, since bottom-up proteomic is commonly used nowadays, the sample loaded on the instrument is not intact protein but enzymatically digested peptides.

This will further increase the number of species in the mixture extensively. With the assistance of fractionation at the protein or peptide property, single complex sample will be transferred into multiple simpler samples before instrumental analysis. Even though instrumentation time will be prolonged, more information, especially for the low abundance protein of interest, will be captured. The fractionation method can be divided into two: protein level fractionation and peptide level fractionation.

1.2.3.1 Protein level Fractionation

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) is a powerful tool for protein level fractionation. SDS will help to denature the secondary and non-disulfide tertiary structure20-21. Protein extraction needs to be heated with β – Mercaptoethanol previously to reduce the disulfide bond and unfold the protein. Besides linearize the protein, SDS, as anionic detergent, also applies a negative net charge to all proteins by attaching on the surface of the protein and the heat step could promote the attachment on the protein. Mostly, the anionic ion will distribute evenly on the protein and is proportional to the mass of the protein. So the fractionation of the protein is only determined by the mass of the protein, no further factor needs to be considered. The most popular and commonly used gel electrophoresis for protein level fractionation is Laemmli buffer system. 22 In the system, the protein migration can be divided into two parts, stack the protein into one sharp band and resolve the protein. Proteins loaded on the gel will all migrate at the same velocity at the beginning in the stacking region. The gel pore size is large and the buffer pH in this region is pH 6.8. This pH will achieve the ion in the buffer the similar charge compare to the SDS coated proteins and cause an ion gradient to stack the protein in one sharp band. Then the gel pore size and buffer pH will be different in the resolving region. Gel pore size will be much smaller and cause sieving effect on the protein and separate the protein on the gel based on their molecular weight. The buffer pH in this region is pH 8.8, which allow a much

greater charge of the buffer ions than the proteins. Hence the ion will overrun the coated protein and eliminate the ion gradient and the stack effect. The Laemmli system can provide better band shape and resolution. Coomassie blue and fluorescent dye (Sypro Ruby) are usually the staining reagent chosen to visualize the protein separated on the gel (**Figure 1.3**). The SDS needs to be washed off first to allow the dye molecule bind to the protein. Further processing on the gel such as destaining, protein extraction, reduction, alkylation and digestion need to be implemented before loading sample on the MS for analysis. The shortcoming of the gel aided protein fractionation and digestion method is the protein yield/recovery is low since proteins are trapped in the gel and need to be extracted by dehydrating the gel. This method also requires a significant amount of labor and time due to the numerous manipulations that can't be readily automated.

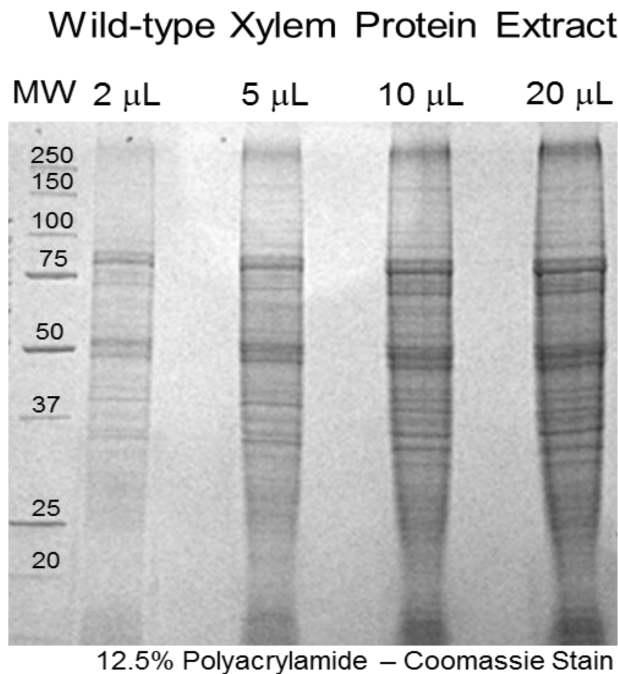


Figure 1.3 - An example of a 1D SDS-PAGE gel stained with Coomassie blue

1.2.3.2 Peptide Level Fractionation

Strong Cation/Anion Exchange on the peptide can be carried out for peptide level fractionation. The strong ion exchange system is composed with mobile phase and stationary phase. The stationary phase could be either positive charged or negative charged. For strong anion exchange (SAX), the stationary phase is made of polymer with positively charged functional group, typically quanternary ammonium ions (**Figure 1.4**). And for strong cation[23, 24, 25] exchange, stationary phase is attached with sulfonate group providing negative charge on the surface. For strong cation exchange, positive charged peptide will be able to stick on the surface of the stationary phase, mobile phase containing positively charged ion will compete with the peptide and elute the peptide out. The pH of the mobile phase and isoelectric point (pI's) of the peptide will determine the elution speed of the peptide and hence separate the peptide into different fractions.

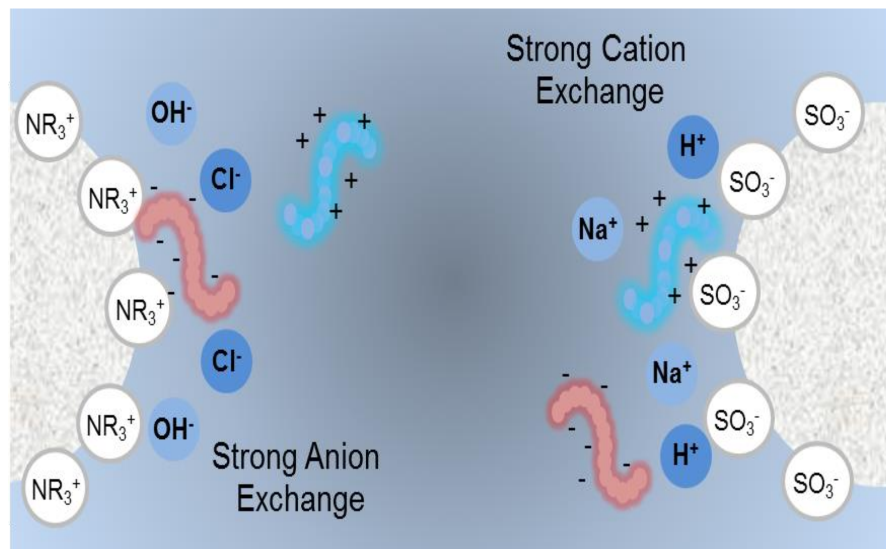


Figure 1.4 - Strong cation exchange and strong anion exchange surface interaction scheme. Ammonium ions are typically the positive charge supplier for the strong anion exchange stationery phase to attract the negatively charged peptide. And sulfonate group would be the negative charge supplier for the strong cation exchange stationery phase to attract the positively charged peptide.

Filter aided sample preparation (FASP) combined with stage-tip SAX fractionation could help on protein digestion followed by peptide level fractionation[26]. To profile the total proteome, the cell or tissue lysed extraction needs to be completely solubilized before analysis using mass spectrometer. SDS as the detergent could help for solubilization but is detrimental to the protein digestion and not compatible with MS analysis. FASP will allow complete removal of detergent because of the key filtration device. The molecular cut-off filter could act as a reactor and help remove the low molecular components such as the SDS and retain proteins and other high molecular components. After the protein digestion takes place above the filter, the peptide can be eluted through the filter so the other high molecular weight component could be removed.

Stop and go extraction tip (Stage-tip)[27] has been demonstrated to be well performed on peptide fractionation. The separation material, such as the strong cation exchange, strong anion exchange, C18, Immobilized Metal Affinity Chromatography (IMAC) materials and poly copolymer materials can be embedded in the Telflon membrane[27]. The membrane will then be cut into small disks and multiple disks will stack into the tapered ends of pipette tips fixedly. These separation material formed as beads keeps a certain distance from each other on the membrane, and this mechanical construction could help prevent primary flow path from forming and help to facilitate the equal flow of the analyte solution as loose beads causes a primary flow channel which is the most limiting factor for micro-column separation[29]. The Stage-tip on this side can allow high loading capacity with high speed and great reproducibility.

Universal Buffer[28] can make wide (pH2-pH12) pH gradient elution possible when implementing strong ion exchange. In SAX, peptide will be eluted when its pI's is reached by the adjustment of the buffer. The first elution buffer is pH 11, and a decreasing trend of the buffer pH will be applied to separate the peptide into different fractions.

1.2.4 Reversed Phase liquid chromatography (RP-LC)

Reversed Phase liquid chromatography29 is often coupled with online with mass spectrometry. Before loading sample on the MS, previous separation of the peptide based on their hydrophobicity is very necessary. The mobile phase is always composed of mobile phase A (high aqueous mobile phase) and mobile phase B (high organic mobile phase). The stationary phase is made up by hydrophobic alkyl chains. For peptide and small molecule separation, the relatively long chains are used to capture the hydrophobic moiety of the peptide. Gradient elution rather than isocratic elution is applied during elution, generally with an increasing of the composition of mobile phase B during the gradient. Hydrophilic molecules will elute earlier than hydrophobic molecules. The reason using gradient elution is because peptide and proteins always elute over a wide percentage organic modifier. Simply applying isocratic elution will induce peak broadening decreasing chromatographic resolution and MS ion abudance since ESI is a concentration sensitive ionization method. The percent of organic mobile phase will not start from 0% basically because under 100% of aqueous condition, the alkyl chain is too hydrophobic to stay in the mobile phase and will simply not be able to interact with the peptide. 0.1% formic acid will always be added into both of the mobile phases as an ion pairing reagent.

1.2.5 Electrospray Ionization (ESI)

After sample analyte have been separated through fractionation and liquid chromatography, molecules need to be loaded on the mass spectrometer waiting to be analyzed or “weighed” to get their mass information for further identification. The way transforming normal atoms or molecules into charged ions is variable. Firstly, the molecules need to be vaporized into gas phase. Various energetic encounters[30] with different component such as electron (EI), chemical species (CI), photon (PhI) and neutral species (PeI) could bring charge on the analyte molecules. Normally it will cause the analyte molecule to lose electrons and become

positive charged. However, for high molecular weight and complex biological molecules, such as protein and DNA, it is really hard to ionize the molecule without decomposing the molecule into fragment. Mass spectrometry had been limited on its application in biological study for the restriction of analyzable mass range for a long period. In the late 1980s, two soft ionization techniques, Matrix assisted Laser Desorption/Ionization (MALDI)[31] and Electrospray Ionization (ESI)[32] were discovered and accelerated significantly the progression of mass spectrometry in biological area. This is because they are “soft” ionization method, which could ionize intact high molecular weight molecules without decomposing the molecule into fragments.

ESI is a very powerful strategy ionizing molecules for its high ionization efficiency, high throughput transferring capacity, wide molecular weight distribution and the ability to keep ions under non-excited state during ionization. As how the molecules are ionized by the electrospray, there are two models available nowadays to explain the mechanism: ion evaporation model[33] and charge residue model[34]. So far there is no conclusion of which model is correct. It is just a matter of how the experimental data can be interpreted better by either one of the two models. And for some cases, even a combination of these two models is going on in the ESI process. The ion evaporation model[33] describes that the droplet shrinks during the evaporation to a certain size limit until the field strength is large enough to expel the solvated molecules in the droplet. The expelled molecules will carry the droplet surface charge with it. During the evaporation, the size of the droplet is decreasing but the charge remains constant, hence the density of the charge is increasing and so is the electric field. Charge residue model[34] describes that the droplet shrinks by solvent evaporation until the field strength is so high and coulombic repulsion occurs to form Taylor cone. Smaller droplet will be emitted from the parent droplet and the Taylor cone will repeated forming until each droplet only contains one analyte molecule. Eventually the solvent evaporates and charge remains on the molecule to ionize the analyte.

1.2.6 Linear Trap Quadrupole-Orbitrap (LTQ- Orbitrap)

LTQ-Orbitrap as one type of Hybrid Mass Spectrometer and can fulfill from routine protein or small molecule identification to very low abundant compound analysis in very complex sample mixture. The major parts of LTQ-Orbitrap include linear ion trap (LTQ), C trap and Orbitrap (**Figure 1.5**). Linear ion trap is composed of four hyperbolic cross section rods[35]. The 2-D construction feature of linear ion trap can help reduce the space charge effect[36], which would have negative effect on the mass spectrometer's performance on mass resolution, mass accuracy, sensitivity, and dynamic range since space charge effect will restrict the number of ions stored by the ion trap. It has been demonstrated that linear ion trap can effectively improve ion capacity and trapping efficiency, and thereby increasing the sensitivity[36]. Precursor survey scan (MS), ion fragmentation, tandem MS^n scan will take place in the Linear ion trap. Then the ion package, with an increase of ion entity will be driven into RF-only C trap. The name of C-trap comes from the shape of the trap system, which is like letter C. Ion mobility will be damped by interaction with low pressure nitrogen existing in the C trap when ions are storing in the C trap. This could help decrease the velocity (kinetic energy) of the ion and temporarily store the ions. Ions will then be injected into Orbitrap, the mass analyzer, by short high electric pulse, and begin to circle around a spindle like electrode in Orbitrap. While the ion package circles around the electrode, it will also oscillate axially along the electrode. The oscillation amplitude is determined by the offset of the ions initially injected into Orbitrap from the middle of the electrode[37]. Oscillation frequency is inversely proportional to the mass to charge ratio of the ion. Two amplifiers surrounding the two halves of the electrode are in charge of recording the ion current image of the motion of ion. The ion current will then be Fourier transformed into mass spectra for further analysis[37].

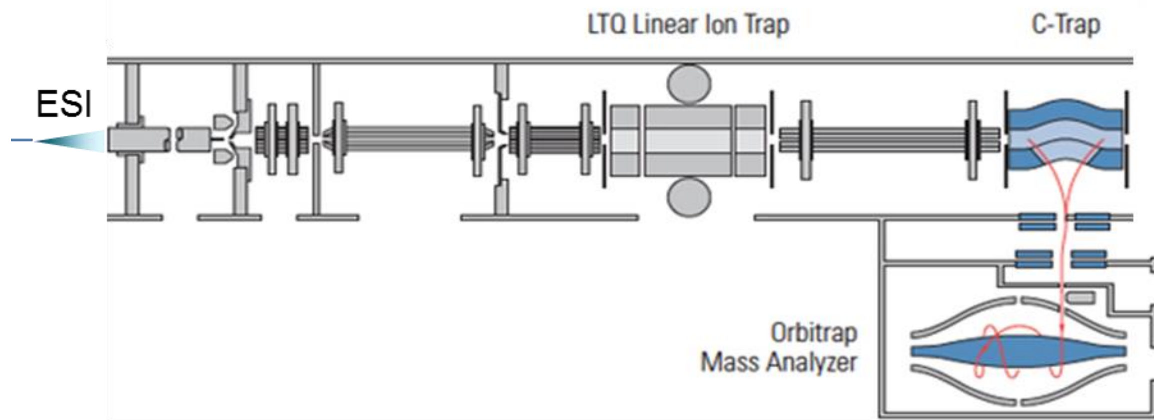


Figure 1.5 - ESI LTQ-Orbitrap: ions will be transferred into the LTQ linear ion trap after being ionized by ESI source. Linear ion trap is coupled with C trap for ion store and Orbitrap for MS analyzing.

1.2.7 Protein Cleavage Isotope Dilution Mass Spectrometry (PCIDMS)

Quantifying the protein abundance is important for us to have better understanding of the proteome of a biological system. SILAC, ICAT, iTRAQ and label free quantification using spectral count are all divisions of quantification to get relative abundance information between samples[38, 39, 40]. SILAC, ICAT, and ITARAQ are isotope labeling relative quantifying methods. Generally, two protein samples to be compared are chemically labeled, one with natural isotope reagent (light tag) and one with isotopically enriched reagent (heavy tag). Two samples are then combined in equal amounts. The light to heavy relative intensity then can be used to indicate the relative abundance of the peptide/protein between the samples. Spectral counts are widely used as label free relative quantifying strategy. The count of peptide fragment spectra is integrated to the spectral count of protein. The general idea is if the protein is more abundant, more spectral counts are measured for that given protein. Spectral counts provide the relative abundance of the same protein between different biological samples. However, to get the exact protein concentration of a sample, absolute quantification is necessary. Isotope dilution[41] as an MS-approach for small molecule

absolute quantification was used for decades and its application of proteins has been studied lately. The general principle (**Figure 1.6**) of isotope dilution is: a certain amount of isotope labeled standard was added into the real native sample as internal standard. The standard and the native target have exactly same chromatographic property but could be distinguished by mass spectrometry since they have different mass. The abundance comparison is able to be obtained from the extracted ion chromatogram by comparison of the peak area or peak height of the quantifiable or total transitions from MRM experiment[42]. Since the amount of the identical standard is known, the concentration of the target can be acquired through peak area ratio and isotopic labeled standard. In the beginning, isotope dilution was applied for polypeptide quantification in top down experiment[43]. The isotope labeled polypeptide will be added in the sample and both native and isotopic forms of intact polypeptide will be ionized and injected into the mass spectrometry for analysis. However, since top town proteomics has limit on the molecular mass, it can't be used for protein has molecular weight higher than 15kDa[43]. Bottom-up proteomics, on the other hand, will digest the proteins into small peptides, and significantly improve the mass range. The use of isotope dilution on digested proteins can help do absolute quantification to large proteins. The ideal case using peptide concentration to stand for the concentration of protein where the peptide originated from is the peptide being uniquely specific to the protein. If the peptide is shared by multiple proteins, the quantity measured of the peptide would be a measurement of all of protein sharing this peptide. Therefore it is necessary to screen out the peptide which could uniquely stand for the protein of interest. Moreover, the peptide needs to be easily synthesized and detected by the mass spectrometer. The specific peptide could uniquely stand for the protein is also called surrogate peptide.[44,45] Identical synthesized stable isotope labeled (SIL) peptide will then be added into the sample, usually before the proteolysis digestion or after the digestion, with a certain amount. The amount of stable isotopic labeled peptide should be comparable with the target, which would

guarantee accurate measurement of the target native peptide. Standards and target peptide will go through the exact same complex matrix effect, HPLC separation, ionization efficiency and instrument response[44]. This approach reduces the overall analytical variability to an acceptable level.

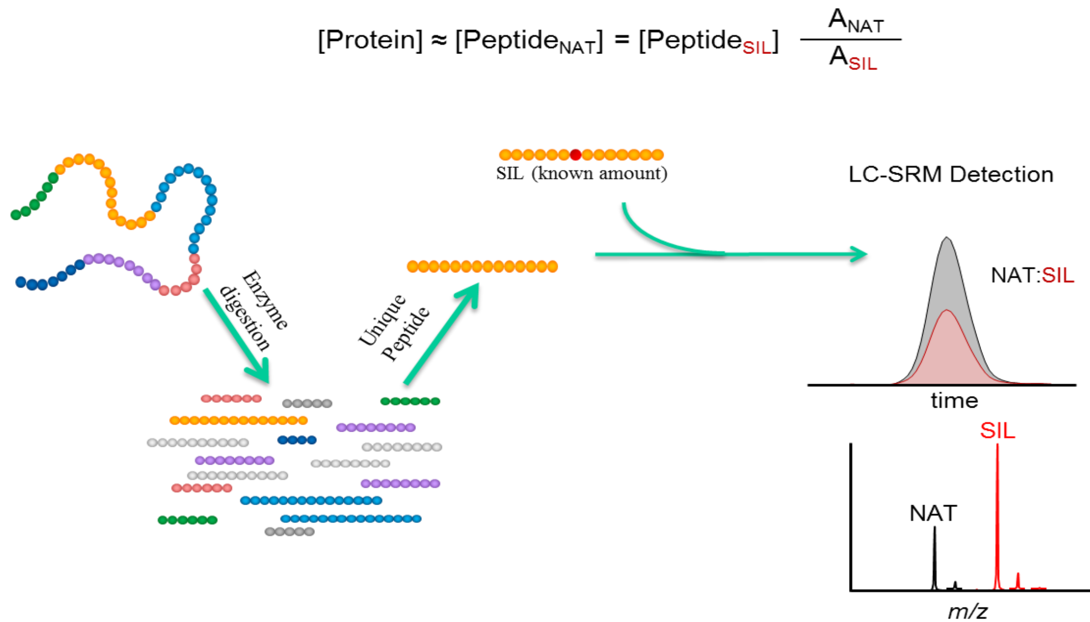


Figure 1.6 - Protein Cleavage Isotope Dilution Mass Spectrometry can be used to quantify protein by using an isotopic labeled analogue of the target peptide as an internal standard2. The general idea is based on the premise that the protein quantity is equivalent to concentration of the unique peptide produced via enzyme digestion. The peptide concentration is determined by spiking a known amount of its stable isotopic labeled analogue as internal standard4. The analogue and the unique peptide have exactly same chemical and physical property but are different in mass. By this feature, the SIL analogue and the unique peptide will elute at the same time point on the RP-HPLC and can be differentiating by the mass spectrometry. The relative abundance can be determined by the ratio of LC-chromatographic peak area, and since the amount of the SIL analogue is known, the concentration of the peptide and the corresponding protein can be determined

Currently in bottom-up experiments, the PCIDMS strategy can be classified into three types according to the origin of the internal standard: AQUA,[46, 47] QconCAT,[48, 49] and PSAQ.[50]

AQUA was first described by Barr[46] et al. when they use SIL synthesized peptides mix to quantify target protein. QcanCAT is abbreviation of Quantitative Peptide Concatemer and was applied by Beynon[48] et al, when they found the surrogate peptides of multiple target proteins could be recombinantly synthesized into peptide concatemer. With enzyme digestion, the concatemer would release surrogate peptide. PSAQ stands for Protein Standard for Absolute Quantification. The internal standards are biological full length proteins which are exactly the same with the target protein, and the surrogate peptide within the protein is isotopic labeled.

1.2.7.1 AQUA Strategy

AQUA is the most widely used method for protein absolute quantification for its commercial availability. When screening the surrogate peptide, some criteria besides uniqueness needs to follow: the sequence length need to be within 7-25 amino acids, chemical reactive residue (Tryptophan, Methionine, Cysteine) needs to be avoided, some amino acids patterns (Aspartate-Glycine, N-terminal Glutamine)[44] which make the peptide synthesis impossible needs to be avoided. These surrogate peptide constraints can lead no surrogate peptide available due to the high sequence similarity between proteins. The protein recovery from pre-fraction and pre-process on the protein needs to be guaranteed. Protein needs to be digested completely but not degraded by over digestion. The time point to add the internal standards into the sample needs to be concurrent with the addition of proteolysis enzyme to compensate for the effect of peptide decay during digestion[51].

1.2.7.2 Quantitative Peptide Concatemer (QconCAT) Strategy

QcanCAT was introduced by Beynon[48] et al. in 2005 when he tried to do large scale protein absolute quantification. Fifty surrogate peptides corresponding to different proteins could be incorporated into one concatemer which significantly increase the quantification scale.

The concatemer is biologically produced once the gene is cloned and expressed; the downstream labeling, quantifying and calibrating could be done on demand. Several surrogate peptides corresponding to one protein could be incorporated into the concatemer to improve the robustness of the quantification. The protein stoichiometry within the sample could be determined if the isotopic labeled peptides within the concatemer is 1 to 1 ratio.[52]

1.2.7.3 Protein Standard for Absolute Quantification (PSAQ)

PSAQ is the most accurate way among these three methods[44,50] since the internal standard protein could be added in the sample at the very beginning protein extraction step. The target protein and the internal standard go through every sample preparation and instrument analyzing step together. All the protein loss during the sample process could be compensated. However, the production for high purity internal standard protein is cost prohibitive for large scale studies.

1.2.8 Synopsis of Completed Research

This dissertation is focused on the mass spectrometry based proteomic study of the *Populus trichocarpa* stem differentiating xylem (SDX) extraction for the purpose of comprehensive understanding of monolignol biosynthetic pathway. **Chapter 2** describes an overall comparison between two common protein digestion-fractionation methods: 1) in gel and 2) FASP-Stage-Tip. The comparison was made at the total proteome level and at the targeted level. Sequence coverage, relative protein abundance, peptide identification and potential chosen as surrogate peptides were all considered as parameters for the comparison. FASP-Stage-Tip was demonstrated a better method for the absolute quantification of the SDX extraction for its high recovery of peptide after series of sample preparation and capability

providing more unique peptide as surrogate candidate for protein absolute quantification. **Chapter 3** describes the absolute quantification on the target proteins of a large set of transgenic *P. trichocarpa* SDX extraction. Genes specific perturbation were applied on the *P. trichocarpa* at 3 transcript levels. All of the target protein related to the perturbed gene was quantified using protein cleavage isotope dilution mass spectrometry. The correspondence of protein abundance and gene transcript level of each transgenic construct was examined. A significance test was applied to compare the protein abundance in the transgenic and native wild type to have a better understanding of the effect of perturbed genes to the other pathway genes and proteins.

1.3 References

1. Boerjan, W.; Ralph, J.; Baucher, M., Lignin biosynthesis. *Annu Rev Plant Biol* **2003,** *54*, 519-46.

2. Fergus, B.; Goring, D., The distribution of lignin in birch wood as determined by ultraviolet microscopy. *Holzforschung-International Journal of the Biology, Chemistry, Physics and Technology of Wood* **1970,** *24* (4), 118-124.

3. Sarkanen, K. V., Renewable resources for the production of fuels and chemicals. *Science* **1976,** *191* (4228), 773-6.

4. Jung, H. G.; Deetz, D. A., Cell-Wall Lignification and Degradability. *Forage Cell Wall Structure and Digestibility* **1993**, 315-346.

5. Sarkanen, K., Precursors and their polymerization. *Lignins, occurrence, formation, structure and reactions* **1971**.

6. Ralph, J.; Brunow, G.; Harris, P. J.; Dixon, R. A.; Schatz, P. F.; Boerjan, W., Lignification: are Lignins Biosynthesized via simple Combinatorial Chemistry or via Proteinaceous Control and Template Replication? *Rec Adv Polyphen Res* **2008,** *1*, 36-66.

7. Li, X.; Weng, J. K.; Chapple, C., Improvement of biomass through lignin modification. *Plant J* **2008,** *54* (4), 569-581.

8. Lewis, N. G.; Paice, M. G., *Plant cell wall polymers. Biogenesis and biodegradation*. American Chemical Society: 1989.

9. Chen, F.; Dixon, R. A., Lignin modification improves fermentable sugar yields for biofuel production. *Nat Biotechnol* **2007,** *25* (7), 759-61.

10. Gross, J. H., *Mass spectrometry*. Wiley Online Library: 2004.

11. Siuzdak, G., The Emergence of Mass-Spectrometry in Biochemical-Research. *P Natl Acad Sci USA* **1994,** *91* (24), 11290-11297.

12. Mclafferty, F. W., Tandem Mass-Spectrometry (Ms-Ms) - Promising New Analytical Technique for Specific Component Determination in Complex-Mixtures. *Accounts Chem Res* **1980,** *13* (2), 33-39.

13. Kenyon, G. L.; DeMarini, D. M.; Fuchs, E.; Galas, D. J.; Kirsch, J. F.; Leyh, T. S.; Moos, W. H.; Petsko, G. A.; Ringe, D.; Rubin, G. M.; Sheahan, L. C.; Comm, N. R. C. S., Defining the mandate of proteomics in the post-genomics era: Workshop report. *Mol Cell Proteomics* **2002,** *1* (10), 763-780.

14. Chait, B. T., Mass spectrometry: Bottom-up or top-down? *Science* **2006,** *314* (5796), 65-66.

15. Mclafferty, F. W., Tandem Mass-Spectrometry. *Science* **1981,** *214* (4518), 280-287.

16. Stahl, D. C.; Swiderek, K. M.; Davis, M. T.; Lee, T. D., Data-controlled automation of liquid chromatography tandem mass spectrometry analysis of peptide mixtures. *J Am Soc Mass Spectrom* **1996,** *7* (6), 532-540.

17. Lange, V.; Picotti, P.; Domon, B.; Aebersold, R., Selected reaction monitoring for quantitative proteomics: a tutorial. *Mol Syst Biol* **2008,** *4*.

18. Lawson, G.; Todd, J., Radio-Frequency Quadrupole Mass Spectrometers. *Chem. Brit* **1972,** *8*, 373-380.

19. Anderson, N. L.; Anderson, N. G., The human plasma proteome - History, character, and diagnostic prospects. *Mol Cell Proteomics* **2002,** *1* (11), 845-867.

20. Shapiro, A. L.; Vinuela, E.; Maizel, J. V., Molecular Weight Estimation of Polypeptide Chains by Electrophoresis in Sds-Polyacrylamide Gels. *Biochem Biophys Res Commun* **1967,** *28* (5), 815-&.

21. Weber, K.; Osborn, M., The reliability of molecular weight determinations by dodecyl sulfate-polyacrylamide gel electrophoresis. *J Biol Chem* **1969,** *244* (16), 4406-12.

22. Laemmli, U. K., Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **1970,** *227* (5259), 680-5.

23. Peng, J.; Elias, J. E.; Thoreen, C. C.; Licklider, L. J.; Gygi, S. P., Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: the yeast proteome. *J Proteome Res* **2003,** *2* (1), 43-50.

24. Vollmer, M.; Horth, P.; Nagele, E., Optimization of two-dimensional off-line LC/MS separations to improve resolution of complex proteomic samples. *Anal Chem* **2004,** *76* (17), 5180-5.

25. Shen, Y.; Jacobs, J. M.; Camp, D. G., 2nd; Fang, R.; Moore, R. J.; Smith, R. D.; Xiao, W.; Davis, R. W.; Tompkins, R. G., Ultra-high-efficiency strong cation exchange LC/RPLC/MS/MS for high dynamic range characterization of the human plasma proteome. *Anal Chem* **2004,** *76* (4), 1134-44.

26. Wisniewski, J. R.; Zougman, A.; Nagaraj, N.; Mann, M., Universal sample preparation method for proteome analysis. *Nat Methods* **2009,** *6* (5), 359-62.

27. Ishihama, Y.; Rappsilber, J.; Mann, M., Modular stop and go extraction tips with stacked disks for parallel and multidimensional Peptide fractionation in proteomics. *J Proteome Res* **2006,** *5* (4), 988-94.

28. Rappsilber, J.; Ishihama, Y.; Mann, M., Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal Chem* **2003,** *75* (3), 663-70.

29. Snyder, L. R.; Kirkland, J. J.; Glajch, J. L., *Practical HPLC method development*. Wiley-Interscience: 2012.

30. Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M., Electrospray Ionization-Principles and Practice. *Mass Spectrom Rev* **1990,** *9* (1), 37-70.

31. Karas, M.; Hillenkamp, F., Laser Desorption Ionization of Proteins with Molecular Masses Exceeding 10000 Daltons. *Anal Chem* **1988,** *60* (20), 2299-2301.

32. Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M., Electrospray Ionization for Mass-Spectrometry of Large Biomolecules. *Science* **1989,** *246* (4926), 64-71.

33. Iribarne, J. V.; Thomson, B. A., Evaporation of Small Ions from Charged Droplets. *J Chem Phys* **1976,** *64* (6), 2287-2294.

34. Dole, M.; Mack, L. L.; Hines, R. L., Molecular Beams of Macroions. *J Chem Phys* **1968,** *49* (5), 2240-&.

35. Mayya, V.; Rezaul, K.; Cong, Y. S.; Han, D., Systematic comparison of a two-dimensional ion trap and a three-dimensional ion trap mass spectrometer in proteomics. *Mol Cell Proteomics* **2005,** *4* (2), 214-223.

36. Schwartz, J. C.; Senko, M. W.; Syka, J. E. P., A two-dimensional quadrupole ion trap mass spectrometer. *J Am Soc Mass Spectrom* **2002,** *13* (6), 659-669.

37. Olsen, J. V.; de Godoy, L. M. F.; Li, G. Q.; Macek, B.; Mortensen, P.; Pesch, R.; Makarov, A.; Lange, O.; Horning, S.; Mann, M., Parts per million mass accuracy on an orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol Cell Proteomics* **2005,** *4* (12), 2010-2021.

38. Ong, S. E.; Mann, M., Mass spectrometry-based proteomics turns quantitative. *Nat Chem Biol* **2005,** *1* (5), 252-262.

39. Bantscheff, M.; Schirle, M.; Sweetman, G.; Rick, J.; Kuster, B., Quantitative mass spectrometry in proteomics: a critical review. *Anal Bioanal Chem* **2007,** *389* (4), 1017-31.

40. Iliuk, A.; Galan, J.; Tao, W. A., Playing tag with quantitative proteomics. *Anal Bioanal Chem* **2009,** *393* (2), 503-13.

41. Viswanathan, C. T.; Bansal, S.; Booth, B.; DeStefano, A. J.; Rose, M. J.; Sailstad, J.; Shah, V. P.; Skelly, J. P.; Swann, P. G.; Weiner, R., Quantitative bioanalytical methods validation and implementation: best practices for chromatographic and ligand binding assays. *Pharm Res* **2007,** *24* (10), 1962-73.

42. Zhang, G.; Neubert, T. A., Automated comparative proteomics based on multiplex tandem mass spectrometry and stable isotope labeling. *Molecular & cellular proteomics : MCP* **2006,** *5* (2), 401-11.

43. Stocklin, R.; Vu, L.; Vadas, L.; Cerini, F.; Kippen, A. D.; Offord, R. E.; Rose, K., A stable isotope dilution assay for the in vivo determination of insulin levels in humans by mass spectrometry. *Diabetes* **1997,** *46* (1), 44-50.

44. Brun, V.; Masselon, C.; Garin, J.; Dupuis, A., Isotope dilution strategies for absolute quantitative proteomics. *J Proteomics* **2009,** *72* (5), 740-9.

45. Pan, S.; Aebersold, R.; Chen, R.; Rush, J.; Goodlett, D. R.; McIntosh, M. W.; Zhang, J.; Brentnall, T. A., Mass spectrometry based targeted protein quantification: methods and applications. *J Proteome Res* **2009,** *8* (2), 787-97.

46. Barr, J. R.; Maggio, V. L.; Patterson, D. G., Jr.; Cooper, G. R.; Henderson, L. O.; Turner, W. E.; Smith, S. J.; Hannon, W. H.; Needham, L. L.; Sampson, E. J., Isotope dilution--mass spectrometric quantification of specific proteins: model application with apolipoprotein A-I. *Clin Chem* **1996,** *42* (10), 1676-82.

47. Gerber, S. A.; Rush, J.; Stemman, O.; Kirschner, M. W.; Gygi, S. P., Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *P Natl Acad Sci USA* **2003,** *100* (12), 6940-5.

48. Beynon, R. J.; Doherty, M. K.; Pratt, J. M.; Gaskell, S. J., Multiplexed absolute quantification in proteomics using artificial QCAT proteins of concatenated signature peptides. *Nat Methods* **2005,** *2* (8), 587-9.

49. Pratt, J. M.; Simpson, D. M.; Doherty, M. K.; Rivers, J.; Gaskell, S. J.; Beynon, R. J., Multiplexed absolute quantification for proteomics using concatenated signature peptides encoded by QconCAT genes. *Nat Protoc* **2006,** *1* (2), 1029-43.

50. Brun, V.; Dupuis, A.; Adrait, A.; Marcellin, M.; Thomas, D.; Court, M.; Vandenesch, F.; Garin, J., Isotope-labeled protein standards: toward absolute quantitative proteomics. *Molecular & cellular proteomics : MCP* **2007,** *6* (12), 2139-49.

51. Shuford, C. M.; Sederoff, R. R.; Chiang, V. L.; Muddiman, D. C., Peptide production and decay rates affect the quantitative accuracy of protein cleavage isotope dilution mass spectrometry (PC-IDMS). *Molecular & cellular proteomics : MCP* **2012,** *11* (9), 814-23.

52. Schmidt, C.; Lenz, C.; Grote, M.; Luhrmann, R.; Urlaub, H., Determination of Protein Stoichiometry within Protein Complexes Using Absolute Quantification and Multiple Reaction Monitoring. *Anal Chem* **2010,** *82* (7), 2784-2796.

53. Kito, K.; Ota, K.; Fujita, T.; Ito, T., A synthetic protein approach toward accurate mass spectrometric quantification of component stoichiometry of multiprotein complexes. *J Proteome Res* **2007,** *6* (2), 792-800.

CHAPTER 2 Comparison of In-gel digestion and FASP Stage-tip digestion using LTQ-Orbitrap Global Proteomic Study on Stem Differentiating Xylem Tissue Extraction from *Populus Trichocarpa*

2.1 Introduction

The ability to comprehensively study all of the biological entities in a complex matrix at the system level is one of the big trends in biological research. Shotgun proteomics is one of the most successful methods capable of achieving extensive proteome coverage[1] in complex biological matrix. Proteins are extracted from tissue cells and digested enzymatically into peptides. These peptides are loaded onto a mass spectrometer allowing for confident identification and quantification. Intact mass and fragmented product ion masses can specifically designate the identification of the peptide. Software use algorithms to match MS and MS/MS spectra to identify these peptides[2, 3] with their corresponding proteins from a user defined protein database. Mass spectrometry, has progressed rapidly in recent decades and is known for its characteristic of high resolving power, parts-per-million mass accuracy, high sensitivity and high speed duty cycle[4]. However, it is still very difficult to comprehensively identify all the proteins within a complex sample such as human plasma[5] due to the wide dynamic range of proteins in the sample and high complexity of tissue or cell extraction. Normally there are two strategies applied to increase the proteome coverage: fractionate the complex sample into multiple less complex ensembles using the physical-chemical property of the proteins (size, isoelectric point) or extensively repeat the experiment given the stochastic nature of the sequencing process.

Lignin is one of the most abundant natural substances in the world and its structure and content determine significantly the efficiency of converting biomass to biofuel, paper and forage. To better understand lignin and ultimately, manipulate lignin artificially, research on the origin of lignin is very necessary. In our study, we choose *Populus trichocarpa* as the

source of lignin since *P. trichocarpa* is the only wood plant with full sequence genome[6, 7] . *P. trichocarpa* is also a very fast growing tree, which makes the study of lignin transcriptomic, proteomic, enzyme activity,, and metabolic response effective. The formation of *P. trichocarpa* lignin monomer: monolignol has been widely studied. A metabolic grid describing the pathway of how monolignol is formed was reported[7, 8, 9]. 11 enzyme families including 25 enzymes and 24 metabolites was already identified as enzyme and metabolite components in the pathway[10].

In-gel[11] and FASP Stage-Tip[12] digestion methods are the most common and robust protein digestion-fractionation methods in proteomic study. During in-gel digestion, protein are loaded onto a SDS-PAGE gel and separate according to the size of the protein. Further protein reduction, alkylation, and digestion are processed on the gel as gel is the carrier of the protein. As for FASP-Stage tip, protein reduction, alkylation and digestion occur prior to fractionation. Peptides are then fractionated according to the net charge and the isoelectric point of each peptide and the fractionation is carried out by strong ion exchange. In this study, the comparison between the two optimized digestion-fractionation methods was made on *P. trichocarpa* stem differentiating xylem extraction. The spectral counts[13], sequence coverage and the peptide identification result were chosen as parameters for the comparison at both total proteome and the target protein (e.g., enzyme component of monolignol biosynthetic pathway) level.

2.2 Materials and Methods

All solvents utilized here were of HPLC-grade and were purchased from Honeywell Burdick & Jackson (Muskegon, MI). All other chemicals and reagents were from Sigma-Aldrich (St. Louis, MO) unless otherwise stated.

2.2.1 Xylem protein extraction

The stem differentiating xylem tissue was harvested from 6 months old tree[10]. 100 cm segment of the tree was cut at 30 cm height from the soil; bark was removed and the stem differentiating xylem tissue was scraped directly into liquid nitrogen. 3 g of tissue would then be grinded within the liquid nitrogen and was followed by a 2 min homogenization on ice in 15 ml extraction buffer (50 mM Bis Tris pH 8.0, 20 mM sodium ascorbate, 0.4 M sucrose, 100 mM NaCl, 5 mM DTT, and 10% (w/w) polyvinylpolypyrrolidone). After Homogenization, sample was centrifuged (3000 g*15 min* 4°C) twice to remove the cellular debris. Final protein concentration was measured by Coomassie Plus Bradford assay (Thermo Scientific, Rockford, IL) and stored in -80°C for further process.

2.2.2 Optimized in Gel digestion

Wild type tree stem differentiating xylem (SDX) tissue extraction containing 50 µg total proteins was loaded on the 1D SDS-PAGE gel for in gel digestion. One whole lane containing proteins was cut and divided into 12 fractions. Each fraction was cut approximately into 1 mm cubes. The gel cubes were destained with 300 µL 50:50 100 mM ammonium bicarbonate (pH 8.0): acetonitrile for 30 min. Gel pieces were then dehydrated by leaving the cubes in 300 µL acetonitrile for 20 min until the cube pieces turned white, the liquid was discarded after dehydration. Subsequently, the protein was reduced (10 mM DTT, 56 °C, 30 min), alkylated (55 mM Iodoacetamide, room temperature, 20 min, dark) and digested (bovine trypsin, 1:5 enzyme: substrate, 37 °C, 16 hours). Following each reduction and alkylation steps, solvents were discarded and the gel pieces were then dehydrated using acetonitrile. All solutions were prepared by 100 mM ammonium bicarbonate (pH 8.0). Digestion was then quenched by 33/66/2 H_2O/acetonitrile/Formic acid (v/v/v) (37 °C, 30 min). Supernatant needs to be transferred into a new vial. Two step peptide extractions on the gel pieces were processed (1. acetonitrile, 20 min; 2. 100 µL 2:1 ammonium bicarbonate:

acetonitrile, 15 min). Each extraction supernatant will combine into the new vial containing the after-quench supernatant. The combination was then frozen, dried and stored in -20°C for further LC-MS/MS study. The dried material is dissolved in 60 µL 0.001% 3-16 Zwittergent for analysis on LC-MS/MS.

2.2.3 Optimized FASP Stage Tip Digestion

Protein extraction sample was diluted two fold in 100 mM DTT denaturing solution (Thermo Fischer Scientific, Rockford, IL), and incubate at 56 °C for 30 min to denature the sample. Proteins were alkylated with 1M Iodoacetamide to give a final concentration of Iodoacetamide of 200 mM and incubate at 37 °C for 60 min. The denaturing and alkylation solution were made by 8M Urea, 50mM Tris-HCl (pH8.0). Pipette an appropriate volume to make final amount of the protein is 50 µg into 10kDa 0.5ml - Amicon MWCO-filter unit ((Millipore, Billerica, MA). Samples were concentrated by centrifugation for 15 min at 14,000 × g at 20°C. After the concentration step, the retentate is rinsed with digestion buffer (2M Urea, 10 mM $CaCl_2$ 50 mM Tris-HCl) 3 times for 15 min, 14,000 × g and 20°C centrifugation condition for buffer exchange. An appropriate volume of enzyme solution (100 µg/ml Bovine Trypsin, 2M Urea, 10 mM $CaCl_2$) was then added into the unit to give a 1:20 enzyme: protein for digestion. The unit is incubated at 37°C for 16 hours.

After digestion, add 100 µL pH 11 Britton & Robinson Universal buffer and 3 µL 1 M NaOH (pH will end up at 11) into the unit and then spin (15min, 14,000g, 20°C) to elute the peptides through the filter. The total protein concentration is measured by UV-Vis (A280) before Stage tip processing on the eluent. The samples were then stored in -20°C before Stage tip processing. A Kel-F Hub (KF), point style 3, gauge 16(1.19 mm) needle, a plunger assembly: N,RN, LT, LTN, for model 1702 (25 µL), were used to extract 6 anion exchange disks from Empore Extraction Disks (3M, St. Paul) and the 6 disks were then stacked into a 200 µL pipette tip[12].

The stage tip assembly was then fixed onto a 2 mL microcentrifuge tube for eluent collection. 1800 × g, 2 min, 20°C centrifugation condition is applied to make the condition solution, anion exchange buffer go through the stage tip. Stage tip is first washed with 100 µL Methanol, 100 µL 1 M NaOH, 200 µL pH 11 (Britton & Robinson Universal Buffer) sequentially to condition the stage tip. The eluent was discarded after each step. After conditioning, 200 µL of the sample (100 µL once) was loaded onto the stage tip and spun. A 100 µL pH 11 buffer was added to the tip-column to elute the peptides. 100 µL Britton & Robinson Universal Buffer pH8, pH6, pH5, pH4, pH3 are used sequentially and the eluent is collected into separate tubes as fractions. Each fraction was frozen and then concentrated to dryness under speed vacuum and stored in -20°C until further analysis in LC-MS/MS. The dried material is dissolved in 60 µL 0.001% 3-16 Zwittergent for analysis on LC-MS/MS.

2.2.4 NanoLC-MS/MS

All the FASP-Stage Tip and In Gel digestion fractions were analyzed on a hybrid LTQ-Orbitrap XL MS (Thermo Fisher Scientific, Bremen, Germany) and the MS result is combined and searched against the database with MASCOT distiller and ProteoIQ.

LC separation system is run by a NanoLC-AS1 autosampler (Eksigent) and the cHiPLC nanoflex system. The cHiPLC trap ((200 µm*0.5 mm, Chrom XP C18-CL 3 µm 120 Å)) is placed in line with cHiPLC analytical column (75 µm*15 cm, Chrom XP C18-CL 3 µm 120 Å). Injection volume for FASP-ST is 10 µL and for 5 µL In-gel digestion. After sample loading on the trap, 9000 nL 100% Mobile Phase A (2 µL/min) metered injection is used to flush the trap to wash out the salts. 135 min gradient is applied on In-gel digestion fractions with a flow rate of 400 nL/min. The gradient is composed of 5-40%B (0-120 min), 40-95%B (120-120.5 min), 95%B (120.5-128.5 min), 95-5%B (128.5-129 min), 5%B (129-135 min). 240 min gradient with 400 nL/min flow rate gradient is applied on FASP-ST fractions. The gradient is composed of 5-35%B (0-225 min), 35-95%B (225-225.5 min), 95%B (225.5-233.5 min), 95-

5%B (233.5-234 min), 5%B (234-240 min). The Mobile Phase A and Mobile Phase B were composed of water/Acetonitrile/formic acid (98/2/0.2% and 2/98/0.2%).

For data-dependent acquisition by LTQ-Orbitrap, the column eluent was ionized by ESI source with 2 keV ion source voltage. Capillary temperature was set at 200°C. The other parameters for the instrument were set as the same of one of our group's publication[14] which could optimize the performance of the Orbitrap to achieve maximum proteome coverage. The data-dependent acquisition result was further searched by Mascot Daemon V2.3.01 (Matrix Science Inc., Boston, MA, USA), the searching database is *P. trichocarpa* JGI V2.2, a modified database including concatenated forward and reverse database for estimate the estimated discovery rate. For the Mascot parameter setting, peptide tolerance is set as ±5 ppm, allowable peptide charge is set as 2+, 3+, and 4+. MS/MS tolerance is set at ± 0.6 Da and the data format is set as Mascot genetic. Carbamidomethyl was set as the fixed modification; Deamination and Oxidation were set as the variable modification. The '.RAW' files derived from MS/MS analyzer are imported in the Mascot and the Mascot distiller was used as data import filter to create the peak-list .MFG file. The peak picking result, peak-list file was searched against database to create the .dat file as the search result. To apply 1% False discovery rate to the search result to make the search result more certain, the '.dat' file is searched by ProteoIQ version 2.1.01_SILAC_beta08 (Bioinquire, Athens, GA, USA). The sequence coverage (%), spectral count information could also be achieved by searching with ProteoIQ.

2.3 Results and Discussion

2.3.1 Digestion Work Flow

The work flow of the two digestion-fractionation methods is shown below (**Figure 2.1**): sample 1 and sample 2 were from the exact same biological sample (*P. trichocarpa* wild type tissue extraction.

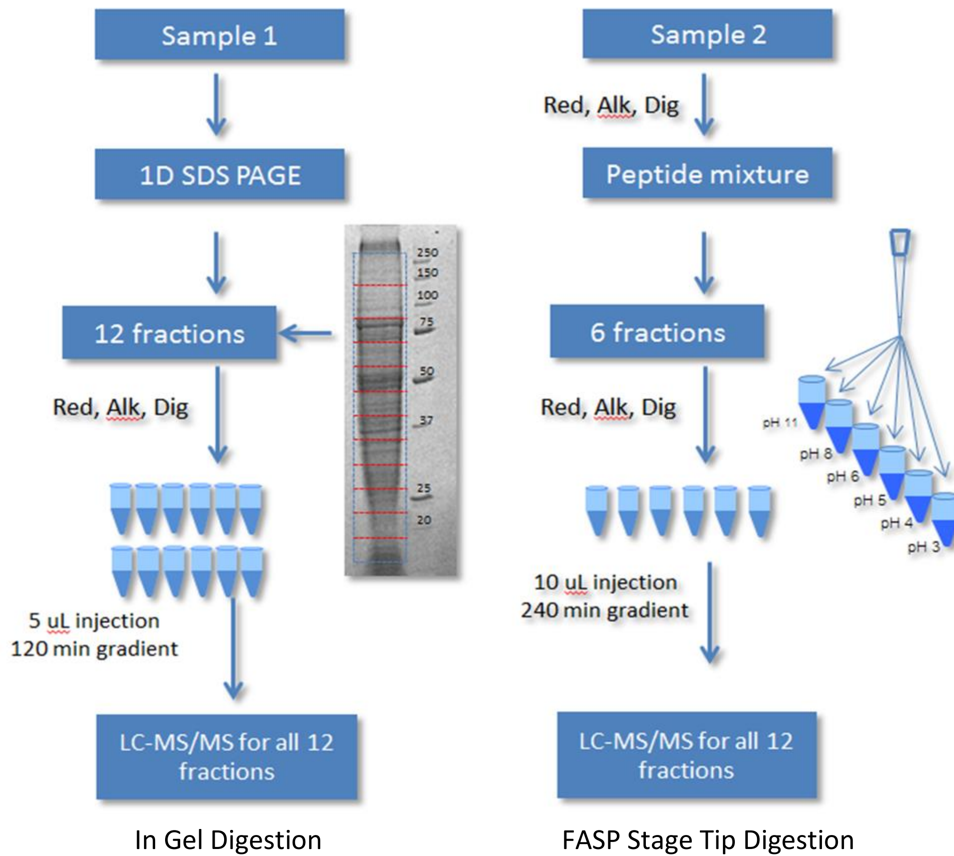


Figure 2.1 - experimental workflow: both methods are optimized protein digestion-fractionation methods. Tissue extraction containing 50 µg of total protein was subjected to each method. Enzyme to protein ratio is 1:5 (w/w) for in gel digestion and 1:20 for FASP Stage tip digestion.

2.3.2 Total Proteome

In-gel digestion contained 12 fractions while FASP Stage tip provided 6 fractions. Each single in-gel processed fraction was subjected to LC-MS/MS analysis with 5 µL injection and 120 min RP-HPLC gradient and each single FASP-ST processed fraction was injected at 10 µL and separated by RP-HPLC with 240 min gradient. Two sets data was combined and analyzed by mascot distiller and proteoIQ. 6030 total proteins were identified and 4352 out of the 6030 proteins were commonly identified in in-gel digestion processed fractions and FASP processed fractions (**Figure 2.2**). In-gel fractions have 556 uniquely identified proteins while FASP fractions have 1122 uniquely identified proteins.

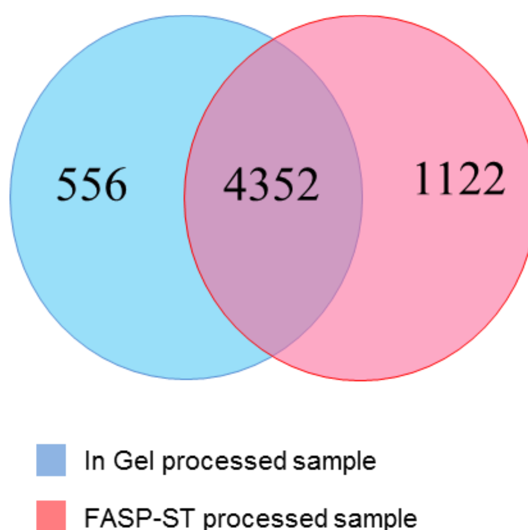


Figure 2.2 - protein identified from in gel and FASP-ST processed fractions

Spectral count[13] can be defined simply as the number of fragment spectra. The peptide spectral count can be summed and integrated to the spectral count of a protein. The change of the spectral count of protein can simply represent the change of the relative abundance

of the protein between different biological samples. A spectral count distribution of the 4352 commonly identified proteins is shown in **Figure 2.3**.

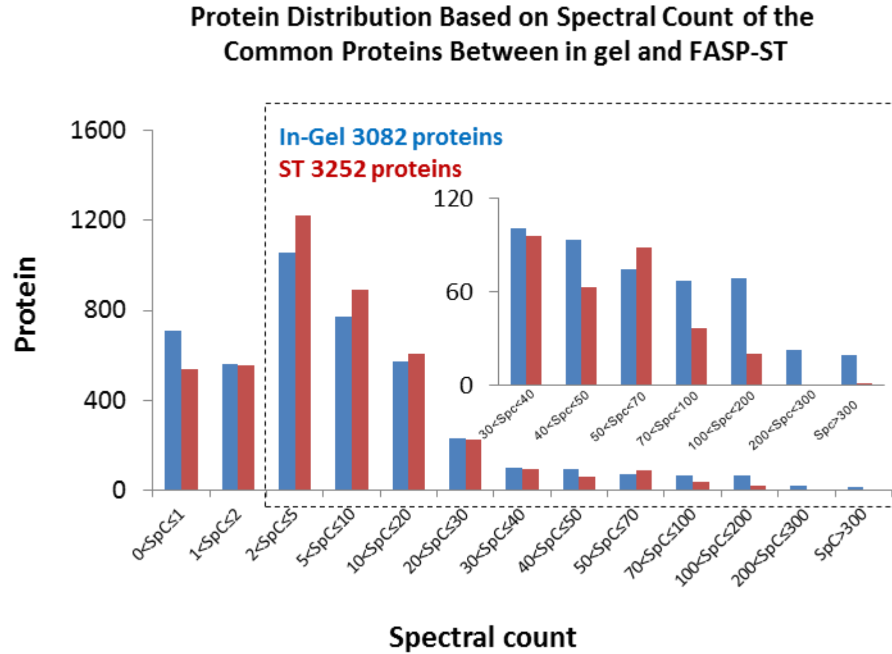


Figure 2.3 - The protein distribution histogram on spectral count. 4352 protein (2716 protein groups, data not shown) were commonly identified from the two digestion-fractionation processed fractions. Each protein has an integrated spectral count designate its relative abundance. Among these 4352 proteins, 3252 proteins from FASP-ST fractions have spectral count greater than 2 while 3082 proteins for in gel fractions respectively. The insect is a clear view of protein distribution when spectral count is greater than 30

Spectral count greater than 2 is more reliable for protein ID positive confirmation. If a protein only has one spectral count, it means only one fragment spectra was identified to specify the protein. The comparison of relative abundance of proteins that only have 1 or 2 spectral count is poor. Among these 4352 protein identified both in the two digestion fractions, 3252 and 3082 protein have spectral count greater than 2 for FASP and in-gel

digestion, respectively. Meanwhile according to the spectral count distribution histogram shown in **Figure 2.3**, most protein's spectral counts lie within the range of 2<SpC<30. (Counts higher than 85% of the SpC > 2 proteins for both methods fractions) and more protein identifications belong to FASP-ST fractions than in gel in this range. FASP-ST provided more "high spectral counts" protein identification than in-gel did. We then plotted the spectral counts scatter of individual protein that were both identified in the two digestion-fractionation methods fractions in condition of spectral count>2 (**Figure 2.4**).

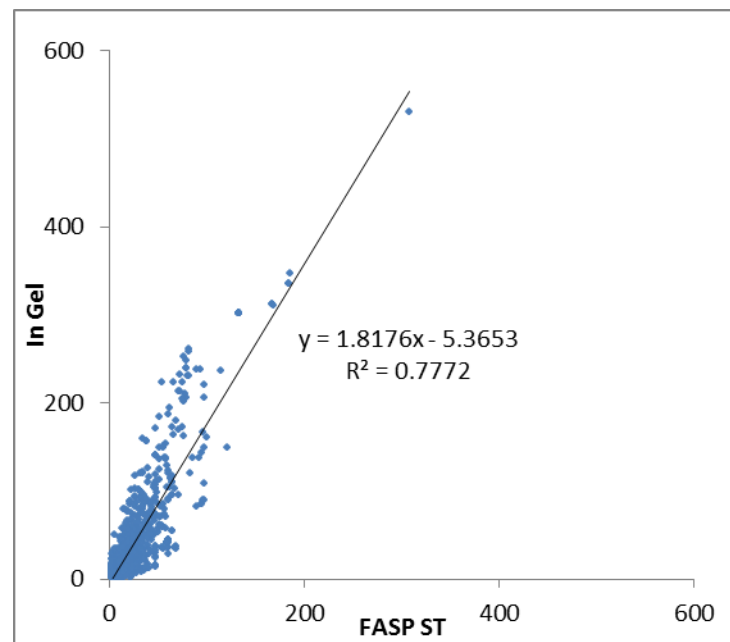


Figure 2.4 - Spectral count plot of in gel against FASP-ST: each spot represent an individual protein and the value of the spot x coordinate is the spectral count of the protein identified in FASP fractions while the value of the spot y coordinate is the spectral count of the same protein identified in in gel fractions. Every protein's spectral count is greater than 2.

Figure 2.4 shows that the slope of the spectral count plot of in-gel against FASP is 1.8176 with R^2=0.7772. We can say that: among the commonly identified "high spectral counts" (SpC>2) proteins, most proteins identified in in-gel have higher spectral counts than the corresponding protein identified in FASP-ST. For example, phenylalanine ammonia-lyase 4 (PAL4) was identified in both fractions. From the in-gel fraction result, 138 spectral counts were integrated to PAL4 while in the FASP-ST fraction result, 92 spectral counts were integrated to PAL4.

Just like the spectral counts scatter plot in **Figure 2.4**, we also plotted the sequence coverage and total peptide of all the commonly identified proteins between these two digestion-fractionation methods fractions

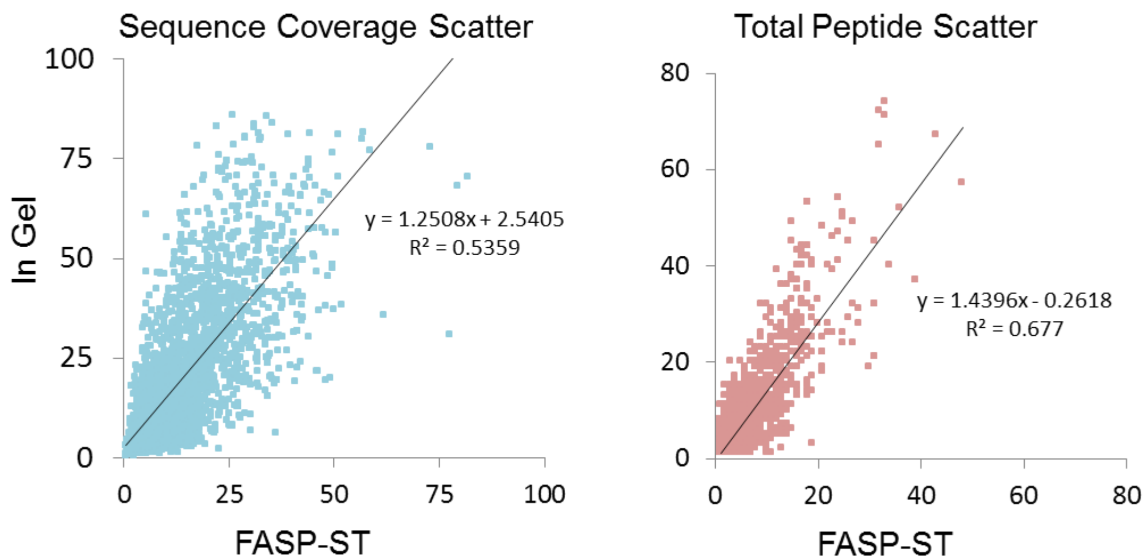


Figure 2.5 - Sequence coverage and total peptide scatter of In Gel against FASP-ST. The slope of the sequence coverage plot is 1.2508 with a coefficient of determination 0.5359, and the slope of the total peptide plot is 1.4396 with a coefficient of determination 0.677. Among all the commonly identified proteins, most proteins (at least more than half) have higher sequence coverage and have more peptide identified from in gel fractions than FASP-ST fractions.

ProteoIQ provides the sequence coverage and the number of peptides of each single identified protein. We plotted the sequence coverage and peptide amount of all the commonly identified protein as in-gel digestion fractions against FASP-ST fractions. The slope of the sequence coverage plot is 1.2508 with a coefficient of determination 0.5359, and the slope of the total peptide plot is 1.4396 with a coefficient of determination 0.677 (**Figure 2.5**). The slope of the plot shows that most commonly identified proteins have higher sequence coverage and peptide identifications from in-gel fractions than FASP-ST fractions. For example, PAL4 was identified at 47.68% sequence coverage and total 40 peptides were identified in in-gel fractions. As in FASP-ST fractions, PAL4 was identified at 41.07% sequence coverage and total 34 peptides were identified. This is consistent with the result from spectral count scatter. Hence we can conclude that comparing in-gel digestion and FASP ST digestion fractions, FASP is able to provide more protein identifications (1122 over 556) and more quantifiable protein (3252 over 3082) than in-gel digestion. However, among the commonly identified "high spectral count" proteins, protein relative abundance acquired from label free spectral count is more reliable from in-gel digestion processed fractions since in-gel digestion resulted in more spectral count on most of these proteins. And the optimized in-gel digestion can provide higher confidence on a single protein identification because it can provide higher sequence coverage and peptide identifications. This could be attributed to the higher enzyme:protein ratio during in- gel digestion. The digestion is probably more complete and more peptide is produced to identify the protein (e.g., more peptide can provide fragment spectra). FASP can provide more protein identifications since the protein loss is less significant during sample preparation and recovery of peptides is significantly greater. The repeatedly dehydration and protein extraction on the gel during in gel processing would cause protein loss at a considerable.

2.3.3 Target Protein

In addition to the whole proteome information of the *P. trichocarpa* SDX tissue extraction, we are more interested on the target protein involved in the monolignol biosynthetic pathway. 11 enzyme family including 25 proteins are identified as enzyme component in the pathway[10]. To better understanding the pathway, not only identification and characterization of these enzyme component is necessary, but also the acquirement of the quantity of each enzyme. Systematic gene specific perturbation in transgenic *P. trichocarpa* is the tool to independently perturb the gene participating in the monolignol biosynthesis[10]. After gene perturbation, each protein in the pathway needs to be absolutely quantified to provide the proteome quantity data for the construction of an integrated statistical model understanding the pathway. We are using protein cleavage isotope dilution mass spectrometry (PCIDMS)[15] to do absolute quantification on these enzyme components. PCIDMS requires availability of surrogate peptide whose concentration could uniquely stand for the concentration of the origin protein. These 25 proteins belong to 11 enzyme family, which determines that most of the proteins have high sequence identity with other protein from the same family. And this would limit largely the amount of surrogate peptide choice. Therefore, besides sequence coverage, and spectral count, we are more concerned about the unique peptide each target protein could provide as candidate of surrogate peptide. Even each protein has a unique peptide that is not shared by other similar protein from the same enzyme family, some criteria is needed to screen out the surrogate peptide. The criteria includes: peptide sequence is not only unique to the protein, but easily detected by mass spectrometry and can be synthesized as the stable isotopic labeled form. Hence we are looking forward more unique peptide could be identified from these target protein.

We performed the comparison of the capability of the two digestion-fractionation methods for providing unique peptide from each target protein together with the comparison of sequence coverage and spectral count

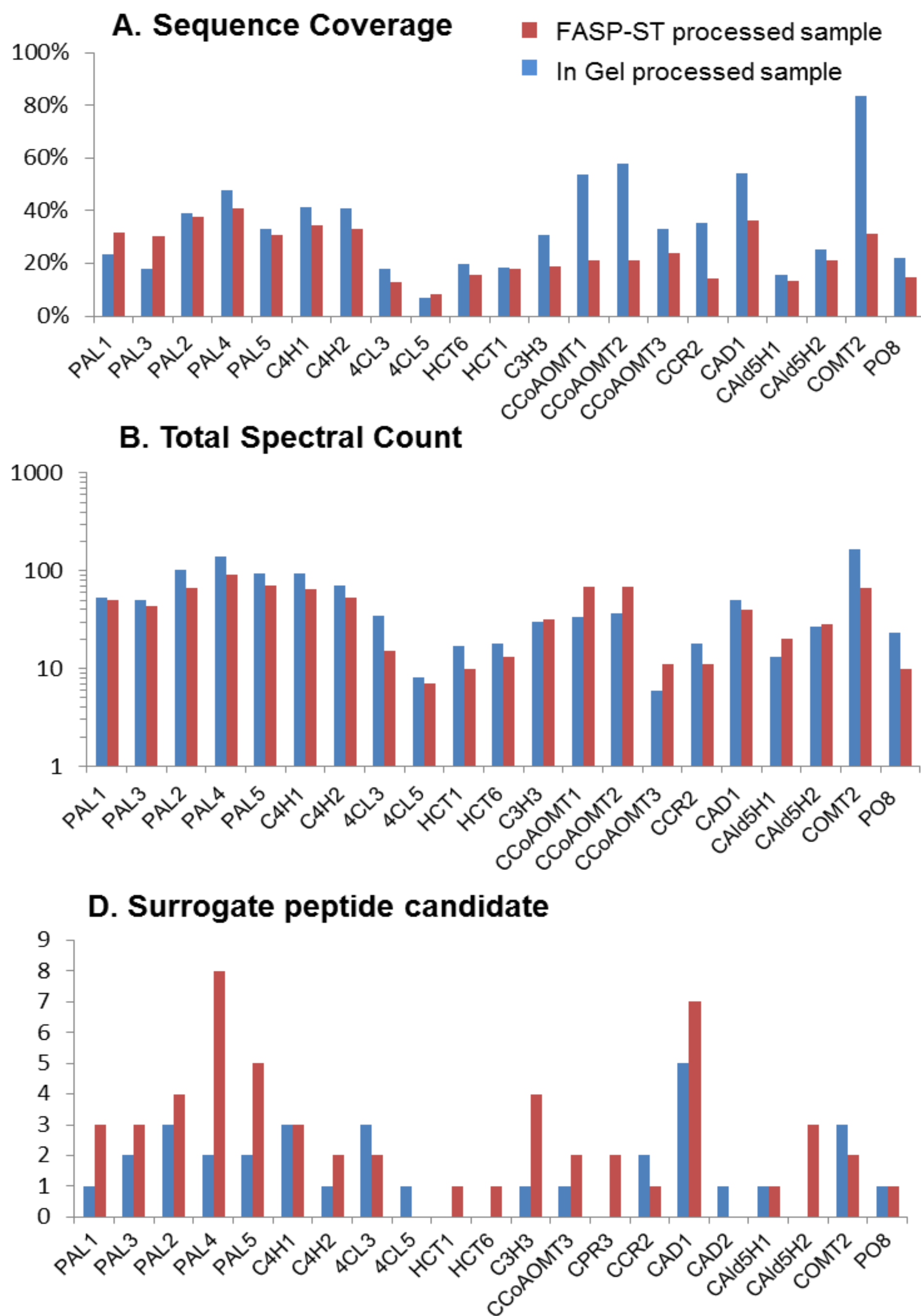


Figure 2.6 - Comparison of in gel and FASP fractions on the target proteins in terms of sequence coverage, spectral count and number of unique peptide. A: sequence coverage of in Gel and FASP-ST on target proteins; B spectral count of in Gel and FASP-ST on target proteins; C: number of unique peptide (surrogate peptide candidate) of in Gel and FASP-ST on target proteins.

Figure 2.6 shows the comparison of sequence coverage, spectral count and number of surrogate peptide candidate (unique peptide) available between in gel and FASP-ST processed samples. In gel and FASP-ST processed sample didn't show very big difference at the sequence coverage and peptide amount of the detected target proteins. But still, in-gel fraction proteins have slightly higher sequence coverage and spectral count than the FASP-ST fraction proteins, which is consistent with the consequence got from total proteome comparison. However, we can see in Figure 2.6.C, most target proteins detected in FASP-ST processed fractions provide more unique peptides than in-gel. FASP was proven more valuable methods processing the protein digestion for our absolute quantification study.

2.4 Conclusion

We compared the optimized in-gel digestion fractionation and FASP Stage Tip digestion-fractionation method performance on total proteome and target protein characterization. Optimized in-gel protein digestion-fractionation provided more peptide dues to relatively complete protein digestion from higher enzyme to protein ratio. Thus the sequence coverage, spectral count and peptide amount of a single protein is higher than the same identified protein from the FASP-ST fractions. FASP-ST method, on the other hand, provided more protein identifications and higher spectral counts for the target protein. This is probably due to the protein and peptide loss is higher during the in-gel digestion process. Protein reduction, alkylation and digestion is processed within the gel core and extracted after they are digested into peptides, the peptide recovery is compromised more significantly for in gel digestion. Besides, in-gel digestion is more time, labor consuming compared to FSAP-ST. The gel needs to be dehydrated after each protein denature step, and peptides need to be extracted after the digestion.

2.5 References

1. Claassen, M.; Aebersold, R.; Buhmann, J. M., Proteome Coverage Prediction for Integrated Proteomics Datasets. *J Comput Biol* 2011, *18* (3), 283-293.

2. Aebersold, R.; Goodlett, D. R., Mass spectrometry in proteomics. *Chem Rev* 2001, *101* (2), 269-95.

3. Aebersold, R.; Mann, M., Mass spectrometry-based proteomics. *Nature* 2003, *422* (6928), 198-207.

4. Gokce, E.; Andrews, G. L.; Dean, R. A.; Muddiman, D. C., Increasing proteome coverage with offline RP HPLC coupled to online RP nanoLC-MS. *J Chromatogr B* 2011, *879* (9-10), 610-614.

5. Anderson, N. L.; Anderson, N. G., The human plasma proteome - History, character, and diagnostic prospects. *Mol Cell Proteomics* 2002, *1* (11), 845-867.

6. Boerjan, W.; Ralph, J.; Baucher, M., Lignin biosynthesis. *Annu Rev Plant Biol* 2003, *54*, 519-46.

7. Dixon, R. A.; Chen, F.; Guo, D.; Parvathi, K., The biosynthesis of monolignols: a "metabolic grid", or independent pathways to guaiacyl and syringyl units? *Phytochemistry* 2001, *57* (7), 1069-84.

8. Higuchi, T., Pathways for monolignol biosynthesis via metabolic grids: coniferyl aldehyde 5-hydroxylase, a possible key enzyme in angiosperm syringyl lignin biosynthesis. *P Jpn Acad B-Phys* 2003, *79* (8), 227-236.

9. Ralph, J.; Brunow, G.; Harris, P. J.; Dixon, R. A.; Schatz, P. F.; Boerjan, W., Lignification: are Lignins Biosynthesized via simple Combinatorial Chemistry or via Proteinaceous Control and Template Replication? *Rec Adv Polyphen Res* 2008, *1*, 36-66.

10. Shi, R.; Sun, Y. H.; Li, Q.; Heber, S.; Sederoff, R.; Chiang, V. L., Towards a systems approach for lignin biosynthesis in Populus trichocarpa: transcript abundance and specificity of the monolignol biosynthetic genes. *Plant Cell Physiol* 2010, *51* (1), 144-63.

11. Shevchenko, A.; Tomas, H.; Havlis, J.; Olsen, J. V.; Mann, M., In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat Protoc* 2006, *1* (6), 2856-60.

12. Rappsilber, J.; Mann, M.; Ishihama, Y., Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat Protoc* 2007, *2* (8), 1896-906.

13. Wang, W.; Zhou, H.; Lin, H.; Roy, S.; Shaler, T. A.; Hill, L. R.; Norton, S.; Kumar, P.; Anderle, M.; Becker, C. H., Quantification of proteins and metabolites by mass spectrometry without isotopic labeling or spiked standards. *Anal Chem* 2003, *75* (18), 4818-26.

14. Andrews, G. L.; Dean, R. A.; Hawkridge, A. M.; Muddiman, D. C., Improving proteome coverage on a LTQ-Orbitrap using design of experiments. *J Am Soc Mass Spectrom* 2011, *22* (4), 773-83.

15. Barr, J. R.; Maggio, V. L.; Patterson, D. G., Jr.; Cooper, G. R.; Henderson, L. O.; Turner, W. E.; Smith, S. J.; Hannon, W. H.; Needham, L. L.; Sampson, E. J., Isotope dilution--mass spectrometric quantification of specific proteins: model application with apolipoprotein A-I. *Clin Chem* 1996, *42* (10), 1676-82.

CHAPTER 3 Absolute Quantification of Enzyme Component in the Monolignol Biosynthetic Pathway of the Transgenic Gene Perturbed Wood *Populus Trichocarpa*

3.1 Introduction

Lignin is one of the most abundant biopolymers in the terrestrial biosphere; it has been widely studied for its importance in nature and the biochemical industry1. Lignin is polymerized from 3 forms of phenylpropanoid monomer: p-coumaryl alcohol (H), coniferyl alcohol (G), and sinapyl alcohol (S)2, 3. The formation of these 3 phenylpropanoid monomers can be described in a monolignol biosynthetic pathway where phenylalanine initiates the beginning of the pathway. A metabolic grid describing the pathway on how monolignol is formed was reported3, 4, 5, and shown in Figure 1.1. There are 24 metabolites and 11 enzyme families that participate in the formation of monolignol and the polymerization of lignin. A comprehensive understanding of the pathway includes identifying, characterizing and quantifying each component (e.g., metabolites and enzymes) under different circumstances (e.g., gene expression knock down specifically in this project), and eventually and primarily, being able to quantitatively describe how the monolignol biosynthetic pathway flux and direction are regulated and integrated. A statistical steady-state model will be constructed to describe regulatory control of the monolignol pathway to the lignin content, composition and the linkage structure.

All the transcription factor and enzyme genes, including their corresponding proteins related within and across the pathway have been identified[6]. A systematic gene perturbation approach was applied on the transgenic *P. trichocarpa* with modified gene expression (suppression mostly). The perturbation (e.g., gene knock-down) is performed using RNA interference (RNAi) and artificial microRNAs (amiRNA)[7] approach. Gene knock-down was carried out in a xylem specific manner at 3 levels: (1) the individual gene level, (2) the phylogenetic gene-pair level, and (3) the gene family level. These gene specific and gene

family specific perturbations will allow us have a deep and comprehensive understanding of the contribution of specific genes to specific pathway flux. Besides, redundancy, feedback, feed-forward effect of one specific gene or gene family perturbation to the other pathway related enzyme and metabolite can also be studied based on the quantification result.

Protein cleavage isotope dilution mass spectrometry (PCIDMS)8 is the method we use to absolutely quantify the enzyme component of the pathway in the P. trichocarpa stem differentiating xylem (SDX) tissue extraction. The idea of PCIDMS is determining the concentration of specific target protein by using the concentration of an enzymatically digested peptide, which is unique to the target protein. The concentration of the native peptide can be obtained by spiking a certain amount of isotopically labeled peptide as internal standard. With the peak ratio of the isotopically labeled and native peptide, and the amount of the isotope, we can calculate the native peptide amount, and that is assumed to be the amount of the protein. The peptides unique to the target protein are further screened to meet a series of criteria9 including: avoiding peptide with sites of known chemical modifications (e.g., Met, Asn, Gln, Cys) or highly probable post-translational modifications and favoring peptides with higher detected ion abundances. Eventually, if peptides selected are not stable or synthesizable, it would be necessary go back to unique peptide list for substitute. Method development includes SRM transition selection, maximizing ion sensitivity, specifically determining the quantity of the peptide, transition validation using ion relative abundance assessment, collision energy optimization, and digestion condition optimization to ensure complete proteolysis of protein to produce expected target peptide[9].

3.2 Materials and Methods

Unless otherwise stated, all reagents were purchased from Sigma-Aldrich (St. Louis, MO). All solvents used were HPLC-grade from Honeywell Burdick & Jackson (Muskegon, MI).

3.2.1 Xylem Protein Extraction

The stem differentiating xylem tissue was harvested from a 6 month old tree. 100 cm segment of the tree was cut at 30 cm height from the soil. The bark was removed, and the stem differentiating xylem tissue was scraped directly into liquid nitrogen. 3 g of tissue would then be grinded within the liquid nitrogen and was followed by a 2 min homogenization on ice in 15 ml extraction buffer (50 mM Bis Tris pH 8.0, 20 mM sodium ascorbate, 0.4 M sucrose, 100 mM NaCl, 5 mM DTT, and 10% (w/w) polyvinylpolypyrrolidone). After homogenization, sample was centrifuged (3000 × g for 15 min* 4°C) twice to remove the cellular debris. Final protein concentration was measured by Coomassie Plus Bradford assay (Thermo Scientific, Rockford, IL) and stored in -80°C for further process.

3.2.2 AmiRNA and RNAi Gene Knock Down

AmiRNAs were devised by using an online program (WMD) based on *P. trichocarpa* genome v1.0 (JGI). The oligo function within the WMD program was used to generate amiRNA* sequences by using the vector option of "RS3000 (MIR319a)". To complete the aim of integrated designed amiRNA and amiRNA* sequences into Ptr-MIR408 transgenic bone, RS3000 vector sequences were replaced respectively by Ptr-MIR408. The strategy overlapping PCR in a SK plasmid was used to prepare amiRNA transgenic fragments. pBI121-based amiRNA binary vectors were constructed mainly followed the procedure described in elsewhere[7], except amiRNA transgene sequence was driven by the xylem specific promoter of 4CL gene.

RNAi transgene with an inverted repeat was prepared following Miki's protocol[10]. Transgene fragments were prepared based on pCR2.1-GL plasmid containing a 680 bp GUS linker[11]. For knockdown, multiple genes within one gene family, two fragments with high homology to target genes were PCR amplified and linked together followed procedure of Li

et. al11 and integrated as sense and antisense besides the GUS sequence to on pCR2.1-GL plasmid.

3.2.3 mRNA Abundance (transcript level) Measurement Using qRT-PCR

Total RNAs were isolated from SDX tissues extractions using the Qiagen Plant RNeasy kit (Invitrogen). Absolute abundances of mRNAs of different Ptr genes were quantitated by qRT-PCR as described in Shi's publication[7].

3.2.4 Stable Isotope-labeled Peptide Standard

The stable isotope-labeled (SIL) surrogate peptides were synthesized by Mayo Clinic Proteomics Research Center (Rochester, MN). The peptide stock solution was made by dissolving 2 mg into water and the absolute concentration was determined by spectrophotometry using Scope's method[12].

Finally, the SIL peptide standards were mixed and called SIL Cocktail. The concentration of each SIL in the cocktail was prepared for the purpose that: each SIL and corresponding NAT peptide from the real sample should produce comparable signal in the LC-SRM-MS experiment. SIL cocktail were then aliquoted and each aliquot could accommodate more than 30 samples of 100 μg total protein. Each aliquot was concentrated into dryness and stored in -80°C for further use

3.2.5 Filter Aided Sample Preparation

Samples were prepared using filter aided sample preparation (FASP)13. First, samples were diluted two fold in 100 mM DTT denaturing solution (Thermo Fischer Scientific, Rockford, IL), and incubated at 56 °C for 30 min to denature the proteins. The denatured proteins were alkylated with Iodoacetamide at a final concentration off 200 mM and incubated at 37 °C

for 60 min. The denaturing and alkylation solutions are made with 8M Urea, 50 mM Tris-HCl (pH8.0). An appropriate volume was pipetted to deliver100 µg of protein into a 10 kDa Amicon MWCO-filter unit (Millipore, Billerica, MA). The sample was concentrated by centrifugation for 15 min at 14,000 x g at 20°C. After the concentration step, the retentate is rinsed with digestion buffer (2M Urea, 10mM $CaCl_2$) 3 times. 5 µL of SIL Cocktail was added into the unit right before the trypsin was added (SIL Cocktail needed to be reconstituted with 183.4 µL 0.001% Zwittergent 3-16). 45 µL of enzyme solution (445 µg/mL bovine trypsin, 2 M Urea, 10 mM $CaCl_2$) is added into the unit to give a 1:5 enzyme: protein ratio (w/w) to digest the proteins into peptides. The unit is incubated at 37°C for 8 hours[9]. After digestion, 50 µL quench buffer (1% Formic Acid (v/v), 0.001% Zwittergent 3-16 (Calbiochem, La Jolla, CA)) was added into the filter unit to stop the digestion. Samples were centrifuged (15 min, 14000 × g, 20 °C) and eluted into the new collection tube. To ensure maximum peptide recovery, the filter unit was rinsed with 400 µL quench buffer and the eluent was collected into the same collection tube. The final total protein concentration was measured by UV absorbance (λ = 280) before analyzing by LC-SRM.

3.2.6 LC-SRM Quantitative Analysis

The FASP processed sample was loaded on a TSQ Vantage triple stage quadrupole mass spectrometer (Thermo Scientific, San Jose, CA) for absolute quantitative analysis. The TSQ Vantage was equipped with an AS1 auotsampler and cHiPLC-nanoflex system (Eksigent, Dublin, CA). The cHiPLC trap (200 µm*0.5 mm, Chrom XP C18-CL 3 µm 120 Å) was placed in line with cHiPLC analytical column (75 µm*15 cm, Chrom XP C18-CL 3 µm 120 Å). 5 µL of sample (~ 1 µg total protein) was injected and loaded on the trap and washed with 100% Mobile Phase A (1.5 µl/min). Peptides were subsequently eluted using a 36 min gradient with 400 nL/min flow rate. The gradient is composed of 5-38.5%B (0-22 min), 38.5-95%B (22-22.5 min), 95%B (22.5-30.5 min), 95-5%B (30.5-31 min), 5%B (31-36 min). The Mobile

Phase A and Mobile Phase B were composed of water/acetonitrile/formic acid (98/2/0.2% and 2/98/0.2%, respectively). The column eluent is ionized under the following conditions: 1400 v ESI potential, 200 °C capillary temperature. Scheduled SRM was performed using the EZ Method with a defined cycle time of 1.5 s, 30 s chrom filter, 1.5 mTorr collision gas, and 2.5 min retention time window.

3.2.7 Data Analysis

The LC-SRM data files (.RAW) were imported into Skyline (v/ 1.1.0.2095)[14] for automatic peak detection and integration. With the assistance of Skyline, the transitions need to be manually validated to ensure there are no co-eluting contaminants selected by Skyline. Peak areas of each transition were exported from Skyline and imported into Excel 2010 (Microsoft, Redmond, WA) for peak purity and transition selection assessment. The quantification of the natural peptide was determined using peak area summing for all non-contaminated transitions according to the following equation:

$$\text{Protein}_{\text{NAT}} \approx \text{Peptide}_{\text{NAT}} = \frac{\sum A\ \text{NAT}}{\sum A\ \text{SIL}}\,[\text{Peptide}_{\text{SIL}}]$$

After absolute quantification, the specific perturbed gene transcription level and corresponding protein amount in the transgenic *P. trichocarpa* was correlated to see if the protein abundance reflected the gene perturbation. The protein abundance in the transgenic and wild type (e.g., native tree sample) was compared using student-t test to check the effect of transgene behavior on the enzyme abundance involved in the pathway.

3.3 Results and discussion

3.3.1 Surrogate Peptide Selection

The target enzyme sequence was subjected to an in silico digestion allowing no mis-cleavages with tryptic digestion rules. Peptide length was set at 7-25 amino acids. The list of

tryptic peptides was then compared with an annotated proteome for *P. trichocarpa* (JGI v2.2) to obtain the list of unique peptide. Proteins from the same enzyme family usually have high sequence identity which will lead to a very limited amount of unique peptides available. For example, PAL 4 and PAL 5 are highly similar in sequence and can only be differentiated by 2 amino acids. Only one common shared peptide was available as the surrogate to quantify the total amount of PAL4 and PAL5.

To check if the *in silico* digested unique peptides are detectable by LC-MS/MS, a shotgun proteomic assay by LC-MS/MS[15] was applied on the native SDX extraction and pure recombinant target proteins. This step allows a validated list of unique and detectable peptides to be created. Some more criteria were applied on these unique MS detectable peptides to ensure their storage stability and ability to easily be synthesized. Also this peptide list was used to avoid sites for known post-translational modifications (e.g., Met, Asn, Gln, Cys)[16, 17] Peptides with higher ion abundance were also preferred. A final list of surrogate peptides is shown below (**Table 3.1**):

Table 3.1 Surrogate Peptide List

Peptide	Sequence
4CL3.262-273	FDIGTLLGLIEK
4CL5.262-273	FEIGSLLGLIEK
C3H3.125-134	VCTLELFSPK
C4H1.255-261	DYFVDER
C4H2.255-261	DYFVEER
CAD1.184-198	GGILGLGGVGHMGVK
CAD2.177-185	YFGLDEPGK
CAld5H1.426-435	FLEPGVPDFK
CAld5H2.L.427-436	FLKPGVPDFK
CAld5H2.M.427-436	FMKPGVPDFK
CCoAOMT1.182-206	VGGLIGYDNTLWNGSVVAPPDAPMR
CCoAOMT2.182-206	VGGLIGYDNTLWNGSVVAPADAPMR
CCoAOMT1|2.115-126	ENYELGLPVIQK
CCoAOMT3.59-65	FLSMLLK
CCR2.299-308	DLGFEFTPVK
COMT2.51-69	AGPGAFLSTSEIASHLPTK
HCT1.338-354	SALDFLELQPDLSALVR
HCT6.338-354	SALDYLELQPDLSALVR
PAL1.664-675	EELGTGLLTGEK
PAL2.661-672	EELGTILLTGEK
PAL3.665-676	EELGTVLLTGEK
PAL4/5.614-622	IGSFEEELK
PO1.136-149	DGIVSLGGPHIPLK
PO2.213-230	IYPTVDPTMDPDYAEYLK
PO3.300-310	MSSITGGQEVR
PO8.113-121	AFEIIEDLR

For enzyme CAld5H2, CAld5H2 427-436 was the only surrogate peptide to meet all of the criteria for SIL synthesis and LC-MS/MS detection. However, this peptide contained an amino acid substitution (L428M), hence both forms of CAld5H2 427-436 was quantified and the real quantity of CAld5H2 was the sum of these two forms9. For CCoAOMT1 and CCoAOMT2, the only surrogate candidate peptide that could be synthesized was

CCoAOMT2.153-166. However, this peptide has an isobaric co-eluting contaminant which would not allow accurate quantification for CCoAOMT2. So instead, we chose one of the unique peptides for both CCoAOMT1 and CCoAOMT2 as surrogate peptide, which could designate us the ratio of these two proteins. In the meanwhile we added a shared surrogate peptide CCoAOMT1|2.115-126 to measure the sum amount of these two proteins[9]. According the ratio and the sum quantity, we can get the specific quantity of both proteins. For PAL4 and PAL5, neither of the two surrogate peptide candidate could be synthesized, so we can only quantify the sum of both protein using a shared peptide: PAL4|5.614-622[9].

SRM transition development, transition validation using relative abundance, and FASP digestion condition optimization were all worked out and described by Shuford *et. al*[9]. Optimally, the 6 most abundant transitions were picked out for each peptide. After transition validation using relative abundance comparison strategy, non-specific transition or transition with poor signal to noise ratio would be excluded. 6 digestion parameters: digestion time, presence of methanol, urea and calcium, enzyme to substrate ratio, and the concentration of trypsin was chosen for FASP digestion condition optimization using fractional factorial design (FracFD) of experiments. In summary, 10 mM Ca^{2+}, 2 M Urea presence, absence of methanol, 8 hour digestion, 400 μg/ml trypsin and 1:5 enzyme: substrate ratio were concluded as the optimized condition for complete proteolysis and accurate quantification.

3.3.2 Transgenic Gene Constructs

Four batches of transgenic tree samples have been processed. Each batch contains 3 or 4 native wild type biological replicates as control. All gene knock-down was driven by a P. trichocarpa xylem-specific promoter (Ptr4CLXP) of Ptr4CL3 gene. For each gene construct, transgenics with 3 distinct knock-down levels of the target gene were selected. Many of these genes are not only from the same gene family but also phylogenetically paired. The

gene constructs that were chosen for knock-down include 3 types: individual gene level, phylogenetic level and gene family level. The target gene knock-down constructs that were included in these 4 Batches were: PAL1, PAL2, PAL3, PAL4, PAL5, PAL1|3, PAL2|4|5, PAL1|3|4|2|5, CAD1, CAD2, CAD1|2, CAD2 overexpress, C3H, C4H1, C4H2, C3H|C4H, CCR2 MYB 156, MYB156|221. RNAi and amiRNA were used as method for gene perturbation. amiRNA is used for individual gene members that share very high sequence similarity such as individual PAL and C4H genes. RNAi was used for phylogenetic pair gene constructs and gene family, or individual gene that has no other gene family member (C3H3) or distantly related gene family member (COMT2 and COMT 24). Each knock-down level of each gene construct has 3 biological replicates except PAL 1, PAL 2, PAL 3, PAL 4, PAL 5.

3.3.3 Absolute Quantification on the Target Protein of the Transgenic Tree

The gene transcription level of each gene knock-down construct was quantified by qRT-PCR[6] . PCIDMS was used as the strategy to implement absolute quantification on every target protein in each biological replicate of each gene knock-down level. We compared the transcript level of gene constructs, and the gene corresponding to protein abundance of each level. Wild type was always used as control for every comparison. The protein abundance difference significance between native wild type and transgenic was tested using student t test, which was aimed to have a general view of the feedback and feed-forward effect of the knock-down gene to the other pathway involved protein.

CAD catalyzes the reduction of hydroxycinnamyl aldehydes to the corresponding[18] monolignols[18]. 16 CAD gene models were found in *P. trichocarpa*, and among the 16 gene models, PtrCAD1 was chosen to perturb for the reason of: PtrCAD1 had a very high transcript level specifically for differentiating xylem (its absolute transcript level in SDX is >30% of the sum of transcripts in: stem differentiating xylem (X), stem differentiating phloem (P), shoot tip (S), and fully expanded leaf (L)[6]). Together with PtrCAD1, PtrCAD2,

even for its very low transcript level, was also added in the list of gene perturbation for its importance for the formation of sinapyl alcohol[19].

The PtrCAD1 gene transcript level and CAD1 protein abundance correlation was shown below (**Figure 3.1**):

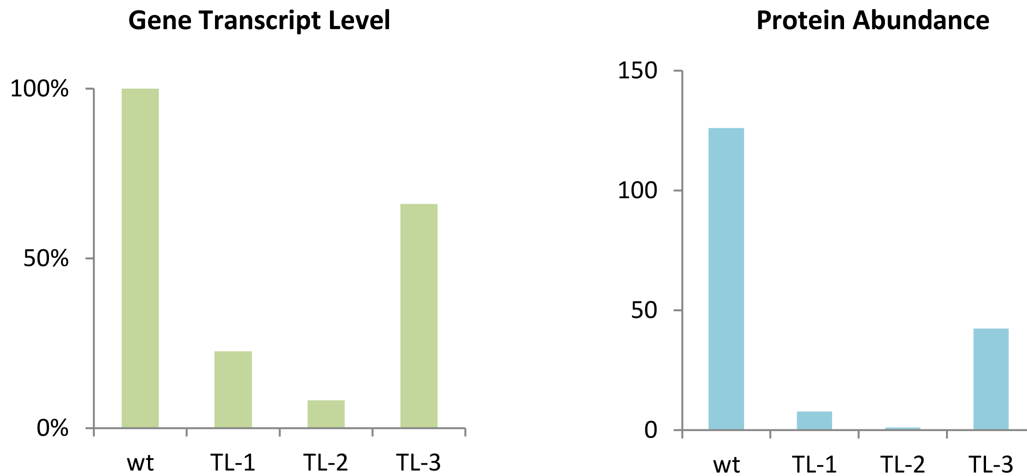


Figure 3.1 - CAD1 gene knock down transcript level and CAD1 protein abundance to corresponding level. CAD1 gene was knocked down in three different levels. Comparing to wild type, level-1 has transcript level of 22.67%, level-2 had transcript level of 8.14% and level-3 had transcript level of 66.06%. Level-2 is the level got gene knocked down the most significantly. The CAD 1 protein abundance in each level was: 7.70, 1.04 and 42.3 pmol/µg of total protein. CAD 1 in wild type is 126.03 pmol/ µg of total protein.

The correlation of CAD1 gene transcript level and protein abundance in 3 level transgenic knock down sample is respectively correlated (**Figure 3.1**). Level-2 shows the gene knocked down the most and correspondingly, the CAD1 protein in level-2 sample is the lowest abundant. Level-3 was knock down the least and correspondingly, contains highest abundant of CAD1 protein other than wild typ.

To have a more comprehensive understanding of the effect of gene perturbation to the other pathway related proteins, we also compared the mean of every target protein in the wild type and the mean of every target protein in the transgenic under each transcript level. We used the student t-test to examine whether the difference between two means is significant. And also calculated the percentage difference of each protein between wild type and transgenic under the lowest transcript level. The student t-test result and percentage difference is showing in **Table 3.2**:

Table 3.2 Student t-test Result

P-Value	TL-1	TL-2	TL-3
PAL1			
PAL2			
PAL3			
PAL4|5			
C4H1			
C4H2			
4CL3			
4CL5			
HCT1			
HCT6			
C3H3			
CCoAOMT1			
CCoAOMT2			
CCoAOMT3			
CCR2			
CAD1			
CAld5H1			
CAld5H2			
COMT2			
PO1			
PO8			

The cutoff of significance test is P=0.05. P-value lower than 0.05 means the protein abundance mean of wild type and the mean of transgenic is different significantly. From the significance test showing below, we can see that CAD1 is the one that differed most significantly showing the lowest P-value, since CAD1 is the one we artificially down regulated. Proteins were knocked down most significantly in Level-2 where only 8.14% transcript level compared to the native wild type. The table showed that in level-2, except the protein CAD1, the abundance of most of target proteins is significantly different with the wild type target protein (P-Value <0.05). In level-3, the level got knock down much less than level-2, the abundance of most of the target proteins didn't show significant differences with the wild type. All of the target proteins are involved in the monolignol synthesis. When we tried to knock down one of the target genes to a certain level, from the protein abundance significance test we can see that the transgenic target protein is not the only protein affected by the specific gene perturbation. Other proteins within the pathway were showing significant difference with the wild type even though their encoded gene was not down regulated. To have a general idea how much the protein abundances were affected, we calculated the percentage difference of each target enzyme family in the level-2.

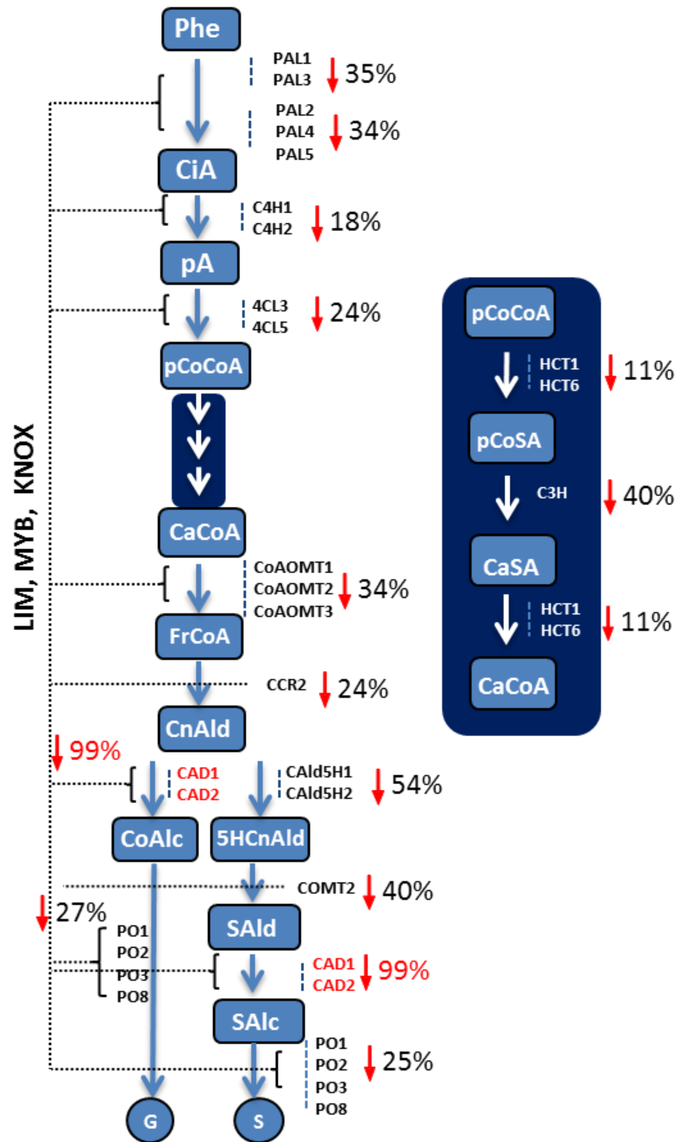


Figure 3.2 - the percentage difference of each enzyme family between transcript level-2 and native wild typ.

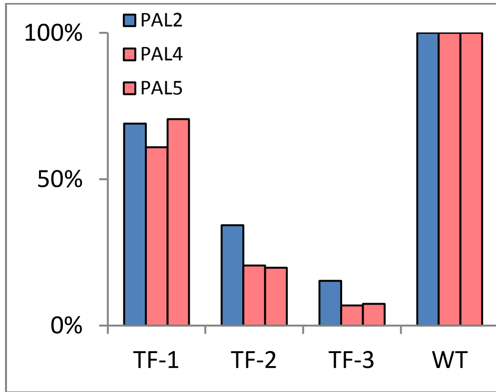
The percentage difference is calculated in the formula:

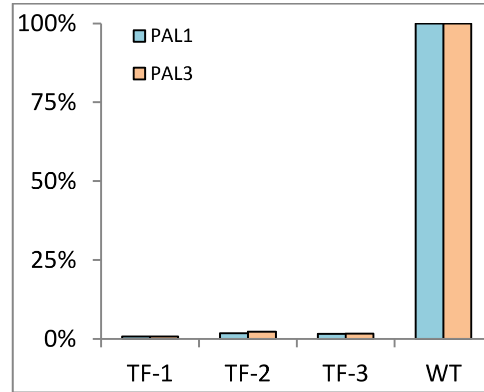
$$d_r = \frac{|\overline{trasngenic} - \overline{wild\ type}|}{\overline{wild\ type}}$$

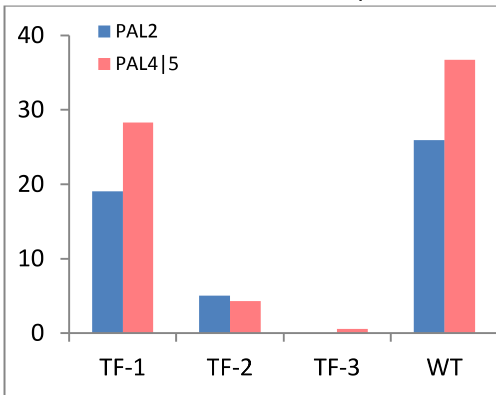
We calculated the percentage difference for all of the target proteins within all the enzyme families. We can see that CAD1, as its target gene was knocked down in level 2, decreased about 99% compared to the wild type. The absolute quantification for CAD1 in wild type is about 126.03 pmol/μg of total protein, but only 1.043 pmol/μg in the level-2 sample. Except CAD1, only C3H, CAld5H and COMT2 were the enzymes that decreased over 40% between level-2 and wild type. C3H decreased from 46.00 pmol/μg to 27.14 pmol/μg, CAld5H decreased from 42.88 pmol/μg to 19.67 pmol/μg and COMT decreased from 3450.33 pmol/μg to 2065.33 pmol/μg. We can see that except C3H, the enzymes that decreased more than 40% when CAD1 was knocked down were the proteins close to where CAD1 is located in the pathway: CAld5H and COMT. CAld5H catalyzes the hydroxylation primarily of coniferaldehyde, leading to the biosynthesis of sinapyl alcohol, the syringyl monolignol. COMT was thought to catalyze the methylation of caffeate and 5-hydroxyferulate for the biosynthesis of monolignols. One possible hypothesis is that when CAD1 is down regulated, the enzyme activity decreased and its substrate SAld did not transfer to SAlc sufficiently. This may inhibit the gene transcript or expression response of the upstream enzymes.

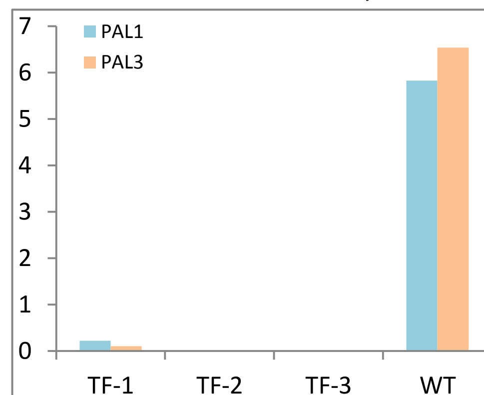
We also checked the loop effect of two phylogenetic gene-pairs in the PAL family at protein abundance level. The PAL gene family can be classified into two phylogenetic groups: PAL2|4|5 as group A and PAL1|3 as group B. Each group was down regulated using RNAi[10] approach. When we knock down group A, group B was not targeted and vice versa. Our collaborator discovered a loop effect when knocking down the gene in group A to the group B gene transcript abundance[7]. Transcripts of group B were up-regulated in transgenic trees of group A as target. But the transcripts of group A were not affected in transgenic trees of group B as target[7]. Rohde also described a null mutation in either one or two PAL genes resulted in an increase in the transcript abundance of the non-mutated PAL genes[20]. Based

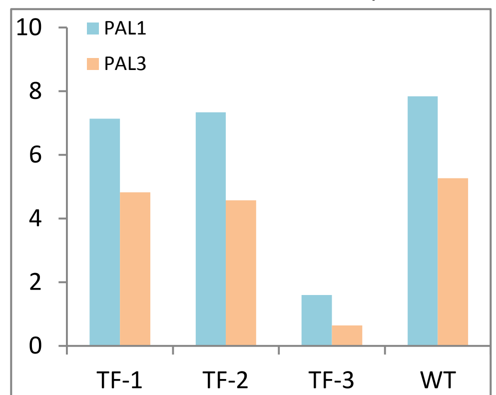
on the result from our collaborator at the gene transcript level, we compared the protein abundance of these two PAL groups. The comparison result is shown below:











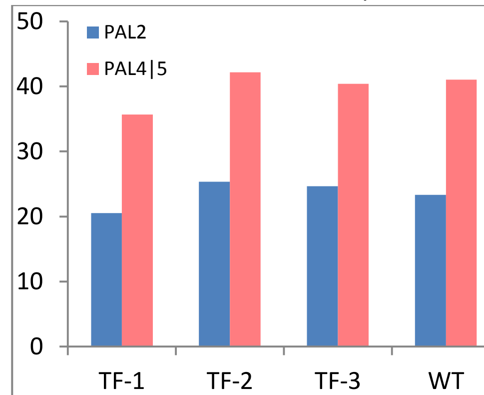


Figure 3.3 - the gene transcript abundance and protein abundance of PAL gene family. A, B, and C are PAL2|4|5 targeted knock down. D, E, and F are PAL1|3 targeted knock down.

We knocked down PAL2|4|5 at different transcript levels and did absolute quantification for all of the PAL proteins. We can see that the protein abundance of group A is corresponding to the gene transcript level of the PAL2|4|5 targeted sample (Fig A, B). In the meanwhile, the protein abundance of PAL1|3 is also affected when PAL2|4|5 were targeted. PAL1 is slightly increased in level 2 compared to level1. In level 3, when PAL2|4|5 were knocked down the most significant, PAL1 and PAL3 protein abundance were decreased significantly. When we knock down PAL1|3 separately in a different set of transgenic trees, we saw that the protein abundance of group B is corresponding to the gene transcript level (fig D, E). But the protein abundance of the PAL2|4|5, the non-targeted group was almost not affected. Not only at the genome level, but at the proteome level, we demonstrated the nonreciprocal loop effect of PAL2|4|5 to PAL1|3.

3.4 Conclusion

A comprehensive understanding of the gene monolignol biosynthetic pathway requires not only fully identifying pathway related molecules such as the enzyme, metabolite, gene, transcript factor component, but also requires accurate and precise quantification information of all of the components while each of the gene was artificially perturbed. Mass spectrometry based absolute quantification applied on the pathway related enzyme component has been demonstrated as a viable approach to accurately and precisely quantity differences in the pathway system under study. We already achieved protein quantity data on more than half of the designed gene constructs used specifically for perturbation. The correlation of protein abundance and gene transcript level and the effect of targeted perturbed gene constructs to other pathway related genes and corresponding proteins has been studied. All of these data helps the construction of the integrated statistical-model to have a systematic understanding lignin synthesis.

3.5 References

1. Boerjan, W.; Ralph, J.; Baucher, M., Lignin biosynthesis. *Annu Rev Plant Biol* **2003,** *54*, 519-46.

2. Sarkanen, K., Precursors and their polymerization. *Lignins, occurrence, formation, structure and reactions* **1971**.

3. Ralph, J.; Brunow, G.; Harris, P. J.; Dixon, R. A.; Schatz, P. F.; Boerjan, W., Lignification: are Lignins Biosynthesized via simple Combinatorial Chemistry or via Proteinaceous Control and Template Replication? *Rec Adv Polyphen Res* **2008,** *1*, 36-66.

4. Higuchi, T., Pathways for monolignol biosynthesis via metabolic grids: coniferyl aldehyde 5-hydroxylase, a possible key enzyme in angiosperm syringyl lignin biosynthesis. *P Jpn Acad B-Phys* **2003,** *79* (8), 227-236.

5. Dixon, R. A.; Chen, F.; Guo, D.; Parvathi, K., The biosynthesis of monolignols: a "metabolic grid", or independent pathways to guaiacyl and syringyl units? *Phytochemistry* **2001,** *57* (7), 1069-84.

6. Shi, R.; Sun, Y. H.; Li, Q.; Heber, S.; Sederoff, R.; Chiang, V. L., Towards a systems approach for lignin biosynthesis in Populus trichocarpa: transcript abundance and specificity of the monolignol biosynthetic genes. *Plant Cell Physiol* **2010,** *51* (1), 144-63.

7. Shi, R.; Yang, C. M.; Lu, S. F.; Sederoff, R.; Chiang, V. L., Specific down-regulation of PAL genes by artificial microRNAs in Populus trichocarpa. *Planta* **2010,** *232* (6), 1281-1288.

8. Barr, J. R.; Maggio, V. L.; Patterson, D. G., Jr.; Cooper, G. R.; Henderson, L. O.; Turner, W. E.; Smith, S. J.; Hannon, W. H.; Needham, L. L.; Sampson, E. J., Isotope dilution--mass spectrometric quantification of specific proteins: model application with apolipoprotein A-I. *Clin Chem* **1996,** *42* (10), 1676-82.

9. Shuford, C. M.; Li, Q.; Sun, Y. H.; Chen, H. C.; Wang, J.; Shi, R.; Sederoff, R. R.; Chiang, V. L.; Muddiman, D. C., Comprehensive Quantification of Monolignol-Pathway Enzymes in Populus trichocarpa by Protein Cleavage Isotope Dilution Mass Spectrometry. *J Proteome Res* **2012**.

10. Miki, D.; Itoh, R.; Shimamoto, K., RNA silencing of single and multiple members in a gene family of rice. *Plant Physiol* **2005,** *138* (4), 1903-13.

11. Li, Q.; Min, D.; Wang, J. P.; Peszlen, I.; Horvath, L.; Horvath, B.; Nishimura, Y.; Jameel, H.; Chang, H. M.; Chiang, V. L., Down-regulation of glycosyltransferase 8D genes in Populus trichocarpa caused reduced mechanical strength and xylan content in wood. *Tree Physiol* **2011,** *31* (2), 226-36.

12. Gietz, R. D.; Schiestl, R. H., High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat Protoc* **2007,** *2* (1), 31-4.

13. Wisniewski, J. R.; Zougman, A.; Nagaraj, N.; Mann, M., Universal sample preparation method for proteome analysis. *Nat Methods* **2009,** *6* (5), 359-62.

14. MacLean, B.; Tomazela, D. M.; Shulman, N.; Chambers, M.; Finney, G. L.; Frewen, B.; Kern, R.; Tabb, D. L.; Liebler, D. C.; MacCoss, M. J., Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **2010,** *26* (7), 966-8.

15. Shevchenko, A.; Tomas, H.; Havlis, J.; Olsen, J. V.; Mann, M., In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat Protoc* **2006,** *1* (6), 2856-60.

16. Blom, N.; Gammeltoft, S.; Brunak, S., Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J Mol Biol* **1999,** *294* (5), 1351-62.

17. Blom, N.; Sicheritz-Ponten, T.; Gupta, R.; Gammeltoft, S.; Brunak, S., Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics* **2004,** *4* (6), 1633-49.

18. Higuchi, T., *Biochemistry and molecular biology of wood*. Springer series in wood science. Berlin, etc.:]. Springer-Verlag: 1997.

19. Li, L.; Cheng, X. F.; Leshkevich, J.; Umezawa, T.; Harding, S. A.; Chiang, V. L., The last step of syringyl monolignol biosynthesis in angiosperms is regulated by a novel gene encoding sinapyl alcohol dehydrogenase. *Plant Cell* **2001,** *13* (7), 1567-86.

20. Rohde, A.; Morreel, K.; Ralph, J.; Goeminne, G.; Hostyn, V.; De Rycke, R.; Kushnir, S.; Van Doorsselaere, J.; Joseleau, J. P.; Vuylsteke, M.; Van Driessche, G.; Van Beeumen, J.; Messens, E.; Boerjan, W., Molecular phenotyping of the pal1 and pal2 mutants of Arabidopsis thaliana reveals far-reaching consequences on phenylpropanoid, amino acid, and carbohydrate metabolism. *Plant Cell* **2004,** *16* (10), 2749-71.