The Pennsylvania State University

The Graduate School

Plant Biology Graduate Program

A STRATEGY TO FACILITATE THE CONVERSION OF LIGNOCELLULOSE TO BIOETHANOL VIA LIGNIN MODIFICATION

A Dissertation in

Plant Biology

by

Fang Cong

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

December 2011

The dissertation of Fang Cong was reviewed and approved* by the following:

Ming Tien
Professor of Biochemistry and Molecular Biology
Dissertation Advisor
Chair of Committee

John E. Carlson
Professor of Molecular Genetics

Nicole Brown
Associate Professor of Wood Chemistry

Jeffrey Catchmark
Associate Professor of Agricultural and Biological Engineering

Tom L. Richard
Professor of Agricultural and Biological Engineering

Teh-hui Kao
Professor of Biochemistry & Molecular Biology
Chair of Intercollege Graduate Degree Program in Plant Biology

*Signatures are on file in the Graduate School

ABSTRACT

This thesis focuses on developing a novel strategy to increase the digestibility of lignocellulosic biomass in bioethanol production by introducing peptides to the plant cell wall. First, the present study tested an experimental approach to verify the subcellular localization of apoplastic signal peptides identified by blasting and proteomic study in poplar. Second, the present study characterized the chemistry of cross linking between amino acids and proteins with lignin. Third, poplar was transformed to express and secrete proteins into the lignifying cell wall. The digestibility of most transgenic lines was increased compared to wildtype.

In total, eight signal peptides were tested for protein subcellular localization using enhanced green fluorescence protein (EGFP) as a reporter. One of the signal peptides is from a putative *Populus* coniferin-β-glucosidase (CβG), which was first identified by blasting a cell wall-localized protein, lodgepole pine (*Pinus contorta*) CβG, against the *Populus trichocarpa* genome. The other seven signal peptides were selected from a poplar xylem sap proteome study. Eight signal peptides were inserted into EGFP-containing vector for plant transformation. Transformants of onion epidermal cells from the eight signal peptides showed extracellular localization of EGFP, however at different ratios comparing to total EGFP fluorescence intensity. *Populus* CβG was first identified in this study and fasciclin-like arabinogalactan protein 10 precursor (AGP) exhibited the highest extracellular to total EGFP intensity ratio in onion. Accordingly these two were also transiently transformed to native plant, hybrid poplar (*Populus deltoides*×*nigra*). In the transformants, EGFP was detected in the endoplasmic reticulum (ER) secretion pathway and extracellular space. Our experimental results are highly consistent with

proteomics database and this study provides a fast and quantitative way for detecting extracellular localization of apoplastic proteins.

Our results show that three amino acids (Cys, Tyr, and Thr) are able to react with intermediates of *in vitro* lignin synthesis: incubations of coniferyl alcohol, peroxidase and H_2O_2. LC/MS results show that the mechanism of cross linking of Cys, Tyr and Thr with lignin, is through the quinone methide intermediate of synthesis. In addition to free amino acids, our results also show the ability of whole proteins to cross link with lignin *in vitro* through Cys or through Tyr residues. Our findings provide a mechanism by which proteins and lignin can cross link in the plant cell wall and reveal the chemical nature of the interactions between protein and lignin.

Poplar was transformed to express and secrete proteins into the lignifying cell wall. One construct (NTRG2) contains a gene encoding a peptide with high Tyr content (17%) with the poplar phenylalanine ammonia lyase (PAL2) promoter (Gray-Mitsumune et al. 1999) and a signal peptide from *Pinus* coniferin β glucosidase (CβG) (Samuels et al. 2002). The other vector (MGRP) has a gene encoding a peptide rich in Cys (11% Cys), which is able to participate in cross linking reactions with lignin *in vitro*. This vector is driven by the double 35S CaMV promoter and poplar CβG signal peptide. Transgenic poplars of NTRG2 showed normal plant growth, morphology, stem structure, and unaltered cell wall composition. Transgenic poplars of MGRP showed slow plant growth, narrow leaf shape, and small stem radius. The digestibility of most transgenic lines of NTRG2 and MGRP surveyed was increased compared to wildtype.

TABLE OF CONTENTS

List of figures……ix

List of tables……..xi

Acknowledgements……xii

Chapter 1 Literature review……..……..……...…1

Overview of bioethanol production…….……..1

The plant cell wall……...…2

Lignocellulose……...…..4

Cellulose…….....4

Hemicellulose…….…..…....5

Lignin…….....…....5

Biosynthesis of monolignols…….……..…8

Transport of monolignols……...10

Dehydrogenation and polymerization……...……10

Proteins in the plant cell wall…….13

Hydroxyproline-rich glycoproteins……...…13

Arabinogalactan proteins……...13

Glycine-rich proteins…….14

Proline-rich proteins……..14

Extracellular transportation of proteins…….14

Pectin…….…….17

Poplar as a biofuels crop……17

Bioethanol production using lignocellulose……...18

Pretreatment……...18

Hydrolysis……..21

Acid hydrolysis……..21

Enzymatic hydrolysis…….21

Fermentation……..22

Improvement of bioethanol production through lignin modification……22

Summary……25

Reference……..…..27

Chapter 2 Screening potential signal peptides for apoplastic secretion of heterologously-expressed proteins in poplar……...……..32

Introduction…….…….…..32

Materials and methods…….…..35

Indentification and synthesis of the signal-peptide-encoding sequence…35

Plasmid construction……..35

Transient expression in onion epidermal cells……...36

Poplar infiltration……...36

EGFP visualization by confocal microscopy…….36

Histogram quantification with ImageJ……...37

Results……38

Identification of signal peptides…….38

Analysis of signal peptides with onion epidermal cells…….42

Analysis of signal peptides in *Populus*……..47

Discussion……..50

Reference…….....53

Chapter 3 Characterization of cross linking between amino acids/proteins with lignin..55

Introduction……55

Materials and methods……...58

Chemicals……...58

Non-radioactive cross linking reaction mixtures……....58

HPLC and LC-MS…….58

Transient state kinetics……...…...61

Radioactive cross linking reaction mixtures……..61
Expression and purification of a Cys-rich peptide…….62
Protein cross-linking reactions…….……...64
Blocking of Cys and Tyr residue with iodoacetamide…….……..65
Autoradiogram of protein cross-linking reactions…….65
Results……66
Amino acid/coniferyl alcohol cross linking……...66
Lack of free radical coupling between Tyr and coniferyl alcohol…….73
Crosslinking of Cys and Tyr into the DHP polymer…….…77
Evidence for cross linking of lignin with proteins…….81
Autoradiogram of SDS-PAGE of cross-linking products of protein and coniferyl alcohol……..……84
Discussion……..86
Reference……...90
Chapter 4 Improved digestibility of poplar by expression of a Tyr/Cys rich peptide in the cell wall……..…….....93
Introduction……...……93
Materials and methods……...96
Binary vector construction…….96
Poplar transformation……...….96
Transcription of the transgene……...97
UV detection and histochemical staining of lignin……97
Compositional analysis of the cell wall…….98
Digestibility assays of steam-exploded pretreated samples……...…...99
Digestibility assays of protease and hot water pretreated samples…...…99
Results……..……101
Expression of transgene……...……101

Phenotype of transgenic poplar……105

Cell wall composition of transgenic poplar……..109

Digestibility and pretreatment of transgenic poplar……...….113

Discussion……117

References……...…….120

Chapter 5 Future research……...…….123

LIST OF FIGURES

Figure 1-1 Schematic representation of the lignified secondary wall……..…..3

Figure 1-2 Structure of three monolignols……..7

Figure 1-3 The biosynthesis pathway of monolignols in flowering plants……...9

Figure 1-4 Radical delocalization following monolignol dehydrogenation, demonstrated for the coniferyl alcohol radicals ….....11

Figure 1-5 Radical coupling of lignin precursors and formation of the β-O-4 dimer…...12

Figure 1-6 Signal peptides and signal-recognition particle (SRP) direct Ribosomes to the ER membrane……...……16

Figure 1-7 Diagram of pretreatment on lignocellulosic biomass……..19

Figure 2-1 BLAST and alignment analysis with *Pinus* CβG sequence (Pc_AAC69619) in *Populus trichocarpa* genome……39

Figure 2-2 EGFP localization in plasmolyzed onion epidermal cells transformed with signal peptide:EGFP observed by confocal microscope……...43

Figure 2-3 The ratio of green fluorescence intensity in apoplast to that in the whole cell, quantified by ImageJ……..46

Figure 2-4 EGFP localization in cells of hybrid poplar leaves transformed with CβG:EGFP and AGP:EGFP observed by confocal microscope……...48

Figure 3-1 HPLC chromatogram of the reaction products……69

Figure 3-2 Mass spectrum of products isolated from amino acid/coniferyl alcohol cross linking reaction…….……71

Figure 3-3 Mechanism of oxidation of horseradish peroxidase……...74

Figure 3-4 Radioactivity in eluent from Sephacryl S-100 column of cross

linking reaction mixtures ……..……..…..79

Figure 3-5 Molarity of coniferyl alcohol and amino acids in eluent from Sephacryl S-100 column of cross linking reaction mixtures……80

Figure 3-6 Cross linking of coniferyl alcohol with either GRP or MGRP……83

Figure 3-7 Cross linking of GRP with coniferyl alcohol as determined by SDS-PAGE and autoradiography…….…….85

Figure 4-1 PCR of genomic DNA from wildtype and NTRG2 and MGRP……102

Figure 4-2 RT-PCR detection of transgene transcription……104

Figure 4-3 Phenotype of transgenic and wildtype plants grown under greenhouse condition……..…106

Figure 4-4 Visualization of lignin in transverse stem sections of wildtype, NTRG2, and MGRP……..…….108

LIST OF TABLES

Table 2-1 List of signal peptides tested in this study ……..……..41

Table 3-1 Gradient program for the reverse-phase HPLC analysis……...60

Table 3-2 DNA and protein sequences of GRP and MGRP……..63

Table 3-3 Rate constants of horseradish peroxidase compound I to compound II and compound II to the resting enzyme……76

Table 4-1 Acid-soluble lignin, acid-insoluble lignin, glucan, and xylan content of wildtype, NTRG2-1, NTRG2-7, and NTRG2-11……110

Table 4-2 Lignin content of wildtype, MGRP-1, MGRP-2, and MGRP-8…….…....112

Table 4-3 Percent residue remaining after digestion of stem tissues of wildtype, NTRG2 and MGRP..……...114

Table 4-4 Percent glucose released from wildtype and NTRG2 after steam-explosion and enzymatic hydrolysis..……....116

ACKNOWLEDGEMENTS

First, I would like to thank all those that have helped and supported me over the past five years with my doctoral study. It has been a long and challenging journey for me and without your encouragement, I could not have come to this far and be where I am.

I own my deepest gratitude to my doctoral advisor Ming Tien for his invaluable support and guidance. He has been a great mentor to take me into the field of biochemistry and inspire my scientific thought and interest of knowledge. Furthermore, I have learned much from him about taking responsibility for both my research and my life. Not only have I obtained my degree, but also become a more mature person during the past five years with him.

Besides my advisor, I would like to thank my thesis committee members, Dr. John Carlson, Dr. Tom Richard, Dr. Nicole Brown, and Dr. Jeff Catchmark for their helpful comments, suggestions, and continual encouragement. This work would not have been possible without them. Also, I would like to gratefully acknowledge my lab members: Yufeng Qian, Prashanti Iyer, Camille Stephen, Joseph Hill, Ashley Hill, John McManus and Qin Chen (Emma) for their precious suggestions and generous help. I truly cherish the time working with them. I feel thankful to Dr. Haiying Liang at Clemson University for her great help with my doctoral research through the past five years.

I also would like to take this opportunity to thank my friends for their constant support. I would like to thank Yili Sun, Yufeng Qian, Junlei Sun, Ye Tian, Meng Zhang, Li Wang, Zi Shi, Yuannian Jiao, Jing Guo, Jia Wen, Shu Li for being there in my times of

need. At Penn State, you guys are my family. To Yufeng, you have helped me in every way possible. I cannot imagine where I would be without your presence in my life. When I was down, you never turned your back on me. Your comfort always dragged me out of the dark time and made me believe that my hard work will pay off.

Finally, and most importantly, I wish to thank my father, my step mother and my grandmother for your boundless love and endless care throughout my life. To my mother and grandfather, you gave me the strongest determination to finish my doctoral degree. To them, I dedicate this thesis.

CHAPTER 1

LITERATURE REVIEW

Overview of bioethanol production

The rapid and extensive industrial development experienced in the mid to late 20th Century was driven, in part, by the harnessing of energy from fossil fuels (Mcconnell et al., 2007). That era is predictably coming to an end within the next 50 to 100 years (Topal and Shafiee, 2009). This is based on predictions from decreasing new discoveries of oil sources on Earth (Topal and Shafiee, 2009). The problems associated with dependence on a declining energy source are wide spread and will have impact in national security, the economy, and of the environment (Demirbas, 2008). Since the energy crisis in the 1970s, the interest in finding sustainable energy source to substitute fossil fuel has emerged (Worldwatch Institute, 2006).

There are several energy alternatives to fossil fuels such as wind power, hydropower, solar energy, and nuclear energy. Outside of fossil fuels, modern biomass, is the largest contributor (1.9%) of the world total energy supply (Goldemberg, 2007). Modern biomass refers to biomass that is produced in a sustainable way (Goldemberg, 2007). Modern biomass can be transformed into a variety of energy products, including heat, methanol, biodiesel, and bioethanol (Brown, 2003). Among them, bioethanol has been recognized as the most favorable renewable fuel for transportation (Zaldivar et al., 2001).

Bioethanol can be produced from a wide range of feedstock. Feedstock for bioethanol can be grouped into three major classes based on their chemical properties (Gray et al., 2006): sucrose-based feedstock (sugarcane and sugar beet), starch-based feedstock (corn grain and wheat), and lignocellulose-based feedstock (corn stover and woody plant). The use of sucrose and starch as substrates for bioethanol production is more widespread than that using lignocellulose (Zaldivar et al., 2001). Over 90% of the world's bioethanol is derived from those two types of feedstock (Zaldivar et al., 2001). As a feedstock, lignocellulosic material is cheaper and more abundant than sucrose and starch, however the processing of lignocellulosic material is most costly (Wyman, 1994).

The plant cell wall

The plant cell wall is composed of cellulose, hemicellulose, lignin, pectin, and proteins, forming a rigid structure. The composition of each component differs depending on cell types. The plant cell wall plays essential roles in determining the morphology of single cells and even the whole plant (Zhong and Ye, 2007). There are two types of cell wall, differing in function and composition: primary cell walls and secondary cell walls (Mcneil et al., 1984). Primary cell walls form first, to provide mechanical support and also have the ability to expand, allowing for cell growth and division (Mcneil et al., 1984). Secondary cell walls are deposited between the plasma membrane and the primary cell wall and in doing so, is part of the process that occurs when the cells stop growing (Cosgrove, 2005). These walls account for most of the carbohydrates found in biomass (Himmel et al., 2007). Each component of plant cell wall is described as follows. Figure 1-1 shows the schematic structure of lignified plant secondary cell wall. Cellulose forms the skeleton and is associated with lignin and hemicelluloses, forming a rigid cell wall structure.

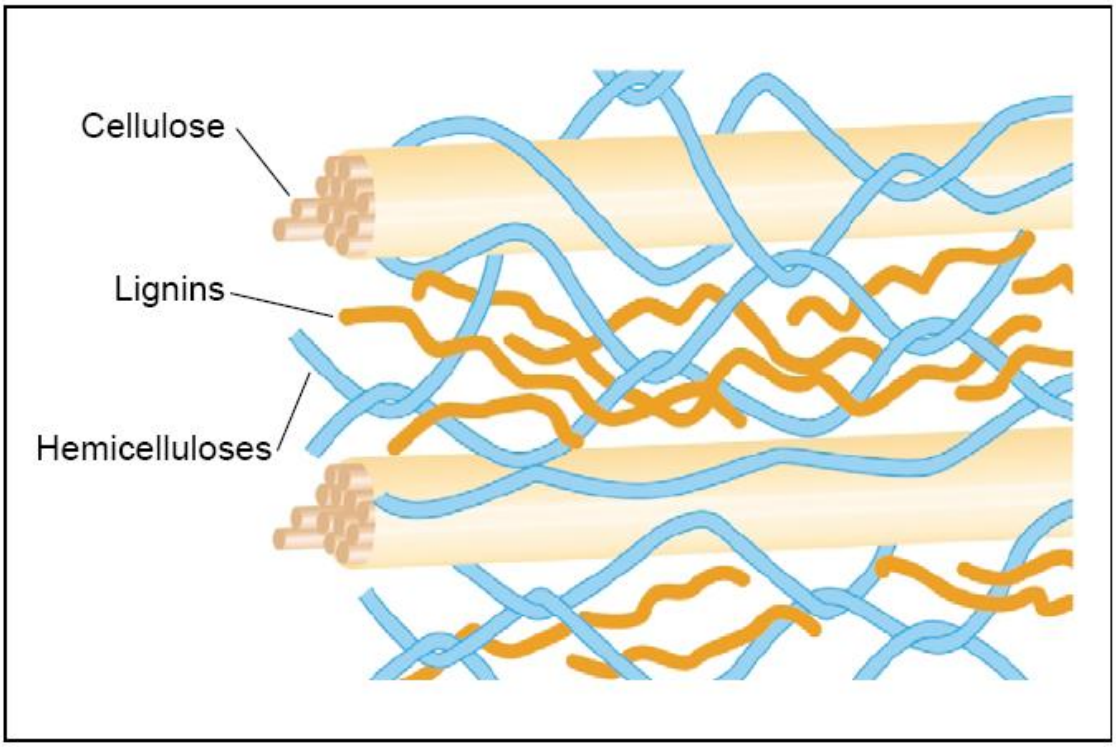


Figure 1-1. Schematic representation of the lignified secondary wall (Boudet 2003)

Lignocellulose

Cellulose, hemicellulose, and lignin together constitute lignocellulose, the most abundant biomass in nature (Wyman, 1994). As a feedstock for biofuels, a variety of sources can provide the lignocellulosic material. These include agricultural and forestry residues, the paper and much of the remaining organic fraction of municipal solid waste (MSW), industrial processing residues, and herbaceous and woody plants grown as feedstocks (Wyman, 1994). Lignocellulose is a biomass feedstock in a wide range of industrial applications, such as pulp, livestock forage, and bioethanol production.

Cellulose

Cellulose is a linear homopolymer of anhydroglucose linked by β-1,4- glycosidic bonds. The disaccharide cellobiose is the basic repeating unit (Taylor, 2008). Single glucan chains of cellulose form microfibrils through inter- and intra-molecular hydrogen bonding and hydrophobic interactions (Burton et al., 2010). The structure of cellulose microfibril is partially crystalline and partially amorphous (Zaldivar et al., 2001). The amorphous regions appear in regularity within the liner crystalline polymer (Taylor, 2008).

Cellulose is synthesized by membrane-associated enzymes (Persson and Endler, 2011). In higher plants, the cellulose synthase is part of a larger complex, located on plasma membrane. Visualization of the complex by electron microscopy reveals a hexameric rosette, with a diameter of 25-30 nm (Doblin et al., 2002). It is thought that each hexameric rosette is composed with six subunits and each subunit contains six cellulose synthase catalytic subunits (CESA) (Lerouxel et al., 2006). Little is known about

the biochemistry of cellulose synthesis in regards to the subunit composition or the identity of other protein partners.

Hemicellulose

Hemicellulose has more complex structure than cellulose. There are four types of hemicelluloses: xyloglucans (XyG), (gluco)mannans, glucuronoarabinoxylans (GAX), and mixed linkage glucans (MLG) (Lerouxel et al., 2006). Unlike cellulose, hemicellulose is a highly branched polymer, amorphous in nature and more heterogeneous than cellulose (Zaldivar et al., 2001). Hemicellulose is synthesized by a family of Golgi-localized glycan synthases encoded by *cellulose synthase like* (CSL) genes that synthesizes the backbone of the β-linked hemicellulose. After the backbone is biosynthesized, side chains are added by glycosyltransferases (Lerouxel et al., 2006). In the plant cell wall, hemicelluloses are associated with cellulose through non-covalent bonds, which is significant to the formation of cell wall structure. According to the model of McCann and Roberts (Mccann Mc, 1991), in primary cell walls, XyG can bind to cellulose by being embedded in the cellulose microfibril during the crystallization, and/or through hydrogen bonds. Collapsed xylem vessels and impaired plant growth are observed in xylan-deficient mutants, so it is believed that hemicelluloses also play an essential role in maintaining secondary cell walls (Turner et al., 2009).

Lignin

Lignin is the most abundant aromatic polymer in nature, derived mainly from three hydroxycinnamyl alcohol monomers differing in their degree of methoxylation. The three monolignol precursors of lignin are *p*-coumaryl alcohol (H), coniferyl alcohol (G), and sinapyl alcohol (S), whose structure are shown in Figure 1-2 (Boerjan et al., 2003). Lignin

plays important roles in mechanical support, water transport, and pathogen defense in vascular plant (Campbell and Sederoff, 1996).

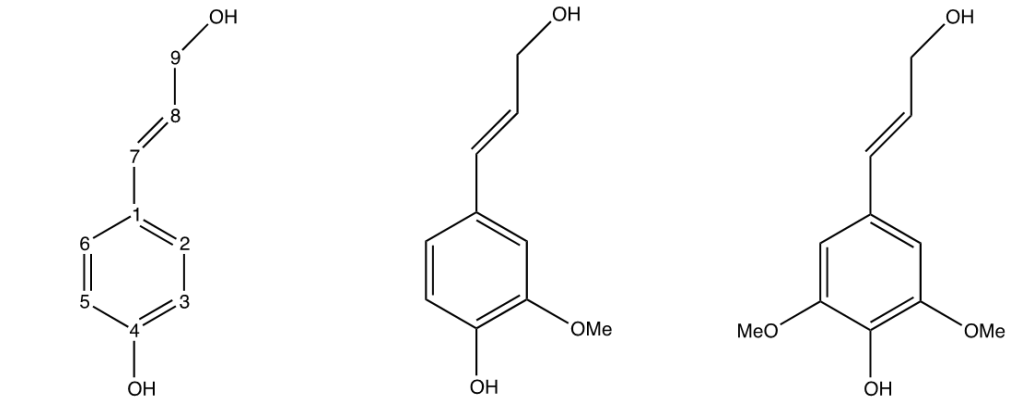


Figure 1-2. Structure of three monolignols *Number indicates carbon position

Biosynthesis of monolignols

Monolignols biosynthesis has been intensively studied over the past two decades. Monolignols are synthesized through the phenolpropanoid pathway which is also responsible for the production of flavonoids, a diverse collection of compounds involved in pigmentation, plant defense, and plant-microbe interactions (Li et al. 2010). The biosynthesis of the monolignols starts with the deamination of phenylalanine, followed by hydroxylation of the aromatic ring, *O*-methylations and modification of side chain (Figure 1-3). To date, it is commonly accepted that 10 enzymes are required for monolignol biosynthesis: phenylalanine ammonia lyase (PAL), cinnamic acid 4-hydroxylase (C4H), 4-(hydroxy)cinnamoyl CoA ligase (4CL), hydroxycinnamoyl CoA:shikimate hydroxycinnamoyl transferase (HCT), *p*-coumaroylshikimate 3'-hydroxylase (C3'H), caffeoyl CoA *O*-methyltransferase (CCoAOMT), (hydroxy)cinnamoyl CoA reductase (CCR), ferulic acid 5-hydroxylase (F5H), caffeic acid/5-hydroxyferulic acid *O*-methyltransferase (COMT) and (hydroxy)cinnamyl alcohol dehydrogenase (CAD) (Boerjan et al. 2003; Li et al. 2008).

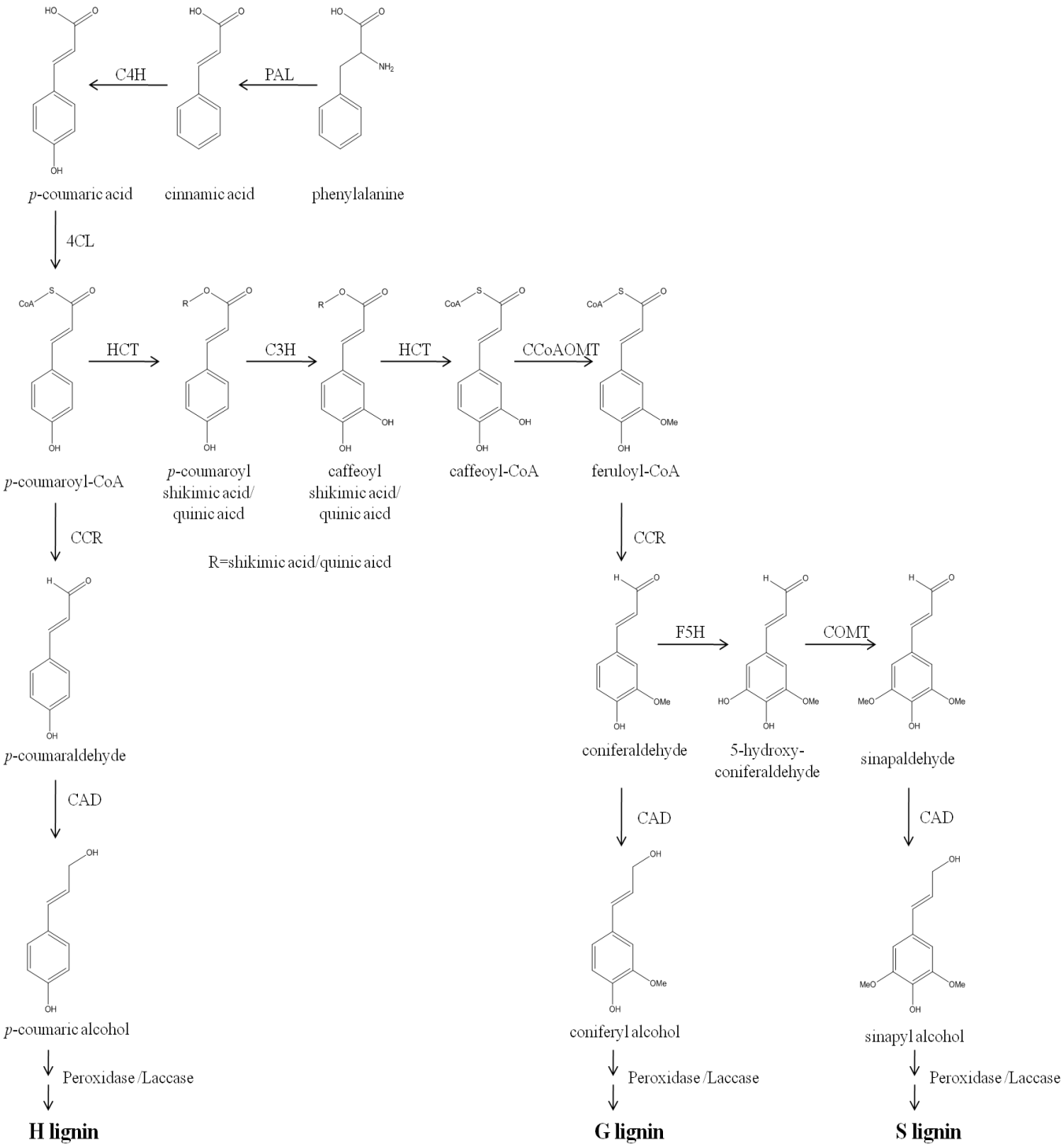


Figure 1-3. The biosynthesis pathway of monolignols in flowering plants (adapted from Li et al. 2008)

Transport of monolignols: possible role of the β-glucoside.

After their synthesis in cytoplasm, monolignols are transported to the cell wall where they are polymerized. In gymnosperms and some angiosperms, monolignol 4-O-β glucosides have been observed to accumulate to high levels in the cambial sap (Steeves et al. 2001). The addition of glucose to form the glucoside makes the monolignol more hydrophilic. As such, the monolignol glucosides have been proposed to be the storage or transport forms of the lignin precursors (Dharmawardhana et al. 1995). Before being oxidized and polymerized into lignin structure, monolignols must to be hydrolyzed to free the phenolic hydroxyl. In the recent studies, β-glucosidases capable of hydrolyzing monolignol glucosides have been detected in suspension culture systems (Hosel and Todenhagen 1980), seedlings (Marcinowski and Grisebach 1978), and differentiating xylem of mature trees (Dharmawardhana et al. 1995).

Dehydrogenation and polymerization

Lignin formation is through dehydrogenative polymerization of the monolignols, which has been proposed to be catalyzed by a number of different enzymes, such as peroxidases, laccases, polyphenol oxidases, and coniferyl alcohol oxidase (Boerjan et al. 2003). These enzymes can oxidize monolignols and generate free radicals with different cosubstrates. For example, peroxidases use hydrogen peroxide and laccases use oxygen. The resulting monolignol radical which is relatively stable due to electron delocalization has single electron density at its 1-, 3-, O-4-, 5-, and β-positions, as shown in Figure 1-4 (Morreel et al. 2010). Following oxidation, two free radicals can be coupled together to form a dimer (Figure 1-5). Radical coupling at the propyl side chain β-position is favored

and in native lignin polymers, around 50% of the linkages are of β-*O*-4 alkyl aryl ether type, which is the most easily cleaved biochemically (Adler 1977).

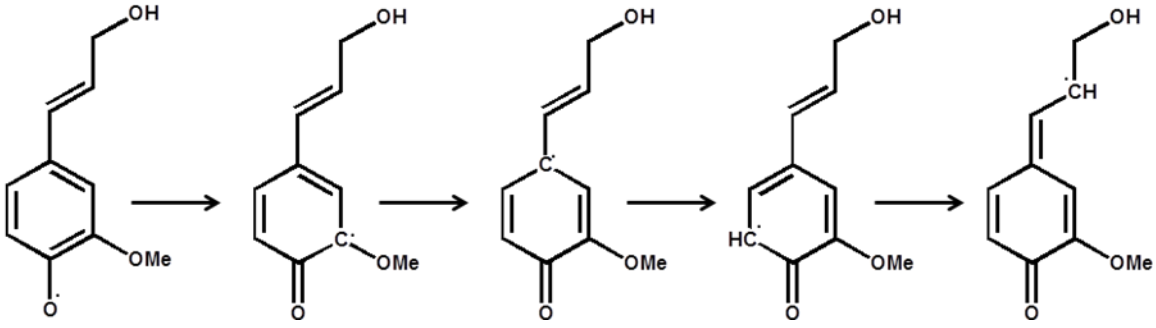


Figure 1-4. Radical delocalization following monolignol dehydrogenation, demonstrated for the coniferyl alcohol radicals. Single electron density is localized at O-4-, 3-, 1-, 5- and β- position.

The most widely accepted model for lignin polymerization is that radicals couple under chemical control (rather than biochemical control) between phenolic radicals in an essentially combinatorial fashion. Recently, a new model has been proposed by Lewis' group (Davin and Lewis 2000). They discovered a class of dirigent proteins involved in biosynthesis of lignans, which are dehydrodimers of monolignols and are typically optically active. Thus they believe polymerization of lignin is under rigidly biological control rather than "random coupling" of monolignols (Davin and Lewis 2000).

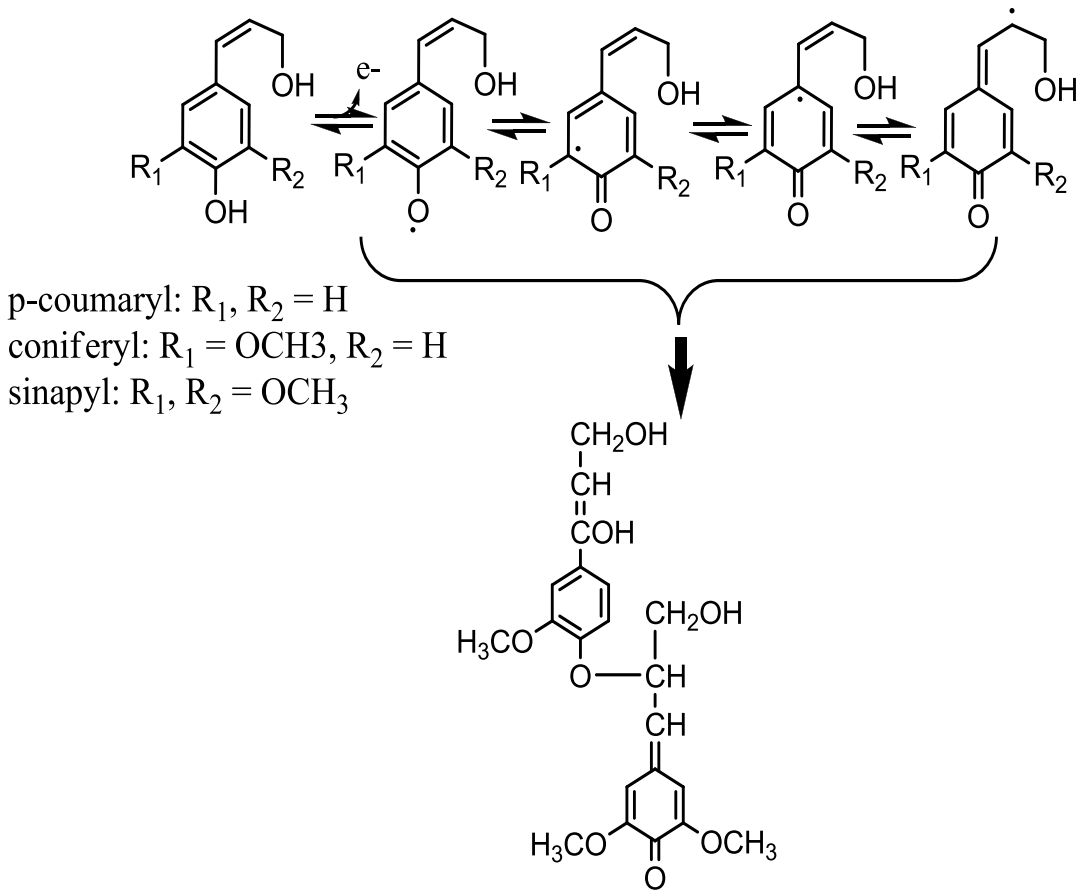


Figure 1-5. Radical coupling of lignin precursors and formation of the β-O-4 dimer.

Proteins in the plant cell wall

In the plant cell wall, there exists a variety of proteins cross linked with other cell wall components (Showalter 1993). Proteins located in the cell wall are grouped into five classes, including hydroxyproline-rich glycoproteins (HRGPs), the arabinogalactan proteins (AGPs), the glycine-rich proteins (GRPs), the proline-rich proteins (PRPs), and chimeric proteins that contain extensins like domains (Cassab 1998).

Hydroxyproline-rich glycoproteins (HRGPs)

A family of HRGPs found in the cell walls of higher plants, is the most well-studied plant cell wall proteins (Showalter 2001). HRGPs are rich in hydroxyproline mostly and some other amino acids, such as serine, valine, tyrosine, lysine, and histidine. HRGPs usually contain the repeating peptide motif of Ser-$(Hyp)_4$, which is significant to define the structure (Showalter 2001). Most of the hydroxyproline residues are glycosylated with one to four arabinosyl residues, and some of the serine residues are glycosylated with a single galactose unit (Showalter 1993).

HRGPs are rapidly insolublized once they are secreted to the cell wall due to the formation of isodityrosine (Fry 1986). HRGPs seem to be covalently cross linked with other cell wall components. Besides the structural significance, HRGPs also function in plant development, wound healing, and plant defense (Showalter 1993).

Arabinogalactan proteins (AGPs)

AGPs, found in higher plant and liverwort, are very soluble and highly glycosylated. They generally contain less than 10% (wt/wt) protein and more than 90% (wt/wt)

carbohydrate. The protein moiety of AGPs is typically rich in hydroxyproline, serine, alanine, threonine, and glycine (Cassab 1998). Due to the stickiness and high sugar content, AGPs play an significant role in pollination, serving as adhesive and nutrient (Lord and Sanders 1992).

Glycine-rich proteins (GRPs)

GRPs, a class of plant cell wall structural proteins, are characterized by high content of glycine (Ringli et al. 2001). Most GRPs contain the signal peptide at N-terminus which results in the secretion of proteins, and extracellular localization was verified with immune-labeled microscope (Showalter 1993). Some studies suggest that GRPs participate in the development of vascular tissues, nodules, and flowers and in the defense of wound and freezing (Cassab 1998).

Proline-rich proteins (PRPs)

PRPs characteristically contain a repetitive motif: $(Pro\text{-}Hyp\text{-}Val\text{-}Tyr\text{-}Lys)_n$ (Chen and Varner 1985). It is noteworthy that PRPs lack serine residue and are lightly glycosylate (Marcus et al. 1991). The PRPs may be implicated in plant normal development and in nodule formation (Showalter 1993).

Extracellular transportation of proteins

All the plant cell wall proteins are synthesized inside the cell and transported out of the cell through the secretory pathway (Walter and Johnson 1994). The process of transport is complex, involving the ribosomes, the mRNA, endoplasmic reticulum (ER) and Golgi complex (Walter and Johnson 1994). All secretory proteins contain a hydrophobic

sequence at N-terminus, called signal peptide (N. Raikhel 2000). Once the signal peptide is translated by ribosome, it binds to a signal-recognition particle (SRP), which exists between ER membrane and cytosol (see Figure 1-6). The SRP-ribosome complex is anchored to SRP receptor, an ER membrane associated protein. Secretory protein enters ER lumen cotranslationally through an aqueous pore in protein translocator on ER (Bruce Alberts 1994). Once the signal peptide is crossed the ER membrane, it is cleaved by signal peptidase, an enzyme located on the lumenal surface of the ER membrane (Vitale and Denecke 1999).

After the secretory protein is synthesized and modified in ER, it enters Golgi complex at the *cis*-face and subsequently travels to *trans*-face to exit to plasma membrane which is default destination, or vacuolar system led by sorting signal (Neumann et al. 2003).

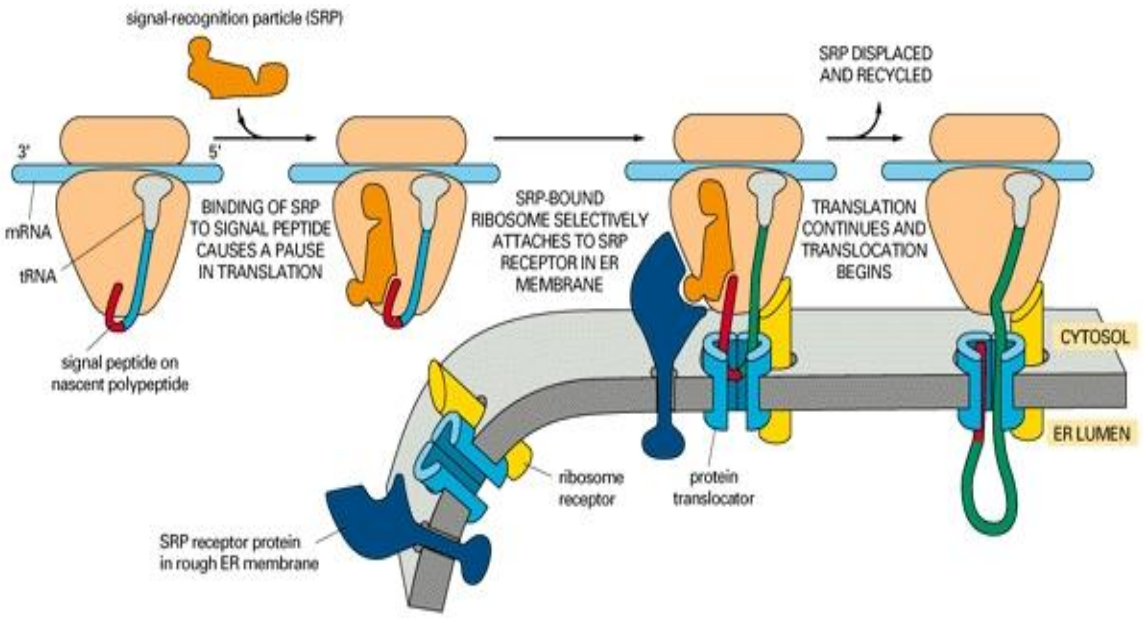


Figure 1-6. Signal peptides and signal-recognition particle (SRP) direct Ribosomes to the ER membrane (Bruce Alberts 1994)

Pectin

Pectin consists a family of complex colloidal polysaccharides that all contain 1, 4-linked α-D-galacturonic acid. Pectins can be extracted from the cell wall with water, or with solutions of chelating agents (Mohnen 2008). Various pectic polysaccharides can be detected in the cell wall, including homogalacturonan (HG), xylogalacturonan (XGA), apiogalacturonan, rhamnogalacturonan I (RGI), and rhamnogalacturonan II (RGII) (Harholt et al. 2010).

Poplar as a biofuels crop

Poplar, together with cottonwood and aspen belongs to the genus *populous*, which contains approximately 30 species, found in the Northern hemisphere. Poplar grows fast with 3-5 feet elongation every year and yield of biomass is 10 dry ton/acre/year (Thompson et al., 2011). Recently, intensive investigation of poplar is motivated by the great economic potential of woody material. Because poplar has relatively small genome size (450-550 Mbp) and is genetically transformable, it was selected as a model to study woody plant and first sequenced (Taylor, 2002).

Poplar provides many good opportunities to investigate questions which are not easy to be addressed in Arabidopsis and rice, the two well-studied model plants. These questions include wood formation and secondary growth, seasonality, and interactions with other organisms (Jansson and Douglas, 2007). All wood-based industries, such as wood processing industries, pulping and paper industries, and biofuels industries will benefit from the better understanding of these questions as researched through poplar.

Bioethanol production using lignocellulose

Because of the low cost and high abundance, lignocellulose has been recognized as a potential feedstock for bioethanol production. Generally, there are three steps in the process of bioethanol production using lignocelluloses: pretreatment, hydrolysis, and fermentation (Wyman 1994). A brief summary for each step is given as follows.

Pretreatment

The carbohydrate components in lignocellulosic biomass exist in the form of cellulose and hemicellulose which are tightly bound together and cross-linked with lignin (Wyman 1994). The association of cellulose and hemicellulose with lignin limits the accessibility of hydrolytic enzymes to polysaccharides (Zaldivar et al. 2001). Thus, lignin is recognized as the major recalcitrance for production of lignocellulosic bioethanol, and there is a great interest in being able to circumvent the lignin recalcitrance.

Pretreatment is a process that breaks the lignin seal and disrupts the crystalline structure of cellulose (Mosier 2005). The purpose of pretreatment is to alter the structure of lignocellulosic biomass to make polysaccharides more accessible during hydrolysis and furthermore increase the efficiency of enzymatic hydrolysis and the yield of fermentable sugars, as shown in Figure 1-7.

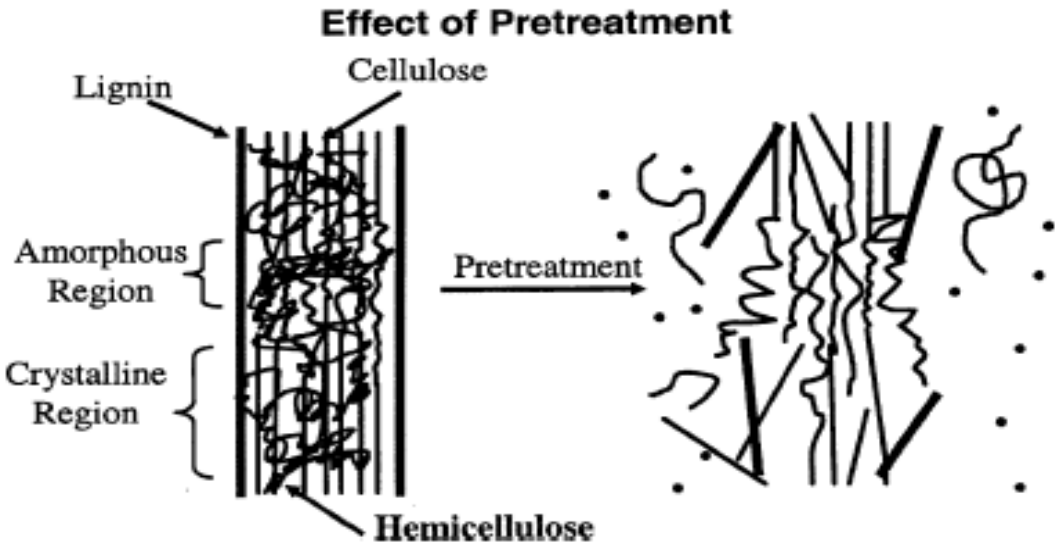


Figure 1-7. Diagram of pretreatment on lignocellulosic biomass (Mosier et al. 2005)

Pretreatment process is generally divided into four major categories: physical process (e.g. milling, grinding and irradiation), chemical process (e.g. alkali, dilute acid, oxidizing agents and organic solvents), physical-chemical process (e.g. steam pretreatment/autohydrolysis, hydrothermolysis and wet oxidation) and biological process (e.g. fungi, enzymes) (Galbe and Zacchi 2007).

To evaluate the efficiency and economy of a pretreatment process, a number of criteria have been developed. Some criteria are summarized from Mosier's paper in 2005: (a) avoiding the need for reducing the size of biomass particles, (b) preserving the pentose (hemicellulose) fractions, (c) limiting formation of degradation products that inhibit growth of fermentative microorganism, (d) minimizing energy demands, (e) lowering pretreatment cost and (f) generating higher-value lignin co-product.

Although current pretreatment methods are effective in increasing the yield of lignocellulosic bioethanol production, there are still some drawbacks. First, harsh chemical reagents and the large volume of water consumed makes pretreatment processes environmentally unfriendly. Second, pretreatment is one of the most expensive steps in the conversion from lignocellulosic biomass to ethanol with costs as high as 30 cent/gallon ethanol (Mosier 2005). Third, high temperature and high pressure required in some processes make it highly demanding in regard to energy input, which causes net energy yield to be lowered. Fourth, inhibitors produced in the process of pretreatment can affect the fermentative microorganism growth and function (Galbe & Zacchi, 2007). Therefore, it has a great potential to improve efficiency and lower cost for pretreatment.

Hydrolysis

The aim of hydrolysis process is to convert the cellulose and hemicellulose into simple sugars which can be fermented to ethanol, and there are two main methods to conduct hydrolysis process, which are acid hydrolysis and enzymatic hydrolysis (Cheng and Sun 2002).

Acid hydrolysis

The acid breaks down the cellulose fraction to glucose and the hemicellulose to xylose and other sugars. However, if hydrolysis conditions are very severe, a large amount of sugars will be converted into other chemicals which not only decrease the sugar yield but also inhibit downstream fermentation. Under these conditions, glucose could be degraded into hydroxymethyl furfural (HMF), and xylose could be degraded to form furfural, tars and other undesirable compounds (Wyman 1994).

Enzymatic hydrolysis

Enzymatic hydrolysis breaks down polysaccharides by enzymes, such as cellulase and hemicellulase. In contrast to acid hydrolysis, enzyme-based method shows many advantages: (a) avoiding the generation of undesired by-products due to the catalytic specificity of enzymes, (b) reducing of the cost of waste treatment, (c) conducting at mild conditions, (d) high yield and efficiency, and (e) economical for large-scale application (Wyman 1994).

Fermentation

The objective of fermentation is to convert sugars which are from hydrolysis process into ethanol by microorganisms. Both five and six carbon sugars are formed from hydrolysis of lignocellulose. The formulas for the conversion of the six and five carbon sugars to ethanol by microorganisms are shown below:

$$C_6H_{12}O_6 \rightarrow 2\ C_2H_5OH + 2\ CO_2$$

$$3C_5H_{10}O_5 \rightarrow 5\ C_2H_5OH + 5\ CO_2$$

Genetic engineering of microorganisms for improvements (rate, yield, conditions, etc) is an active area of research. Researchers are developing the engineering bacteria to increase the fermentation efficiency of both six carbon sugars and five carbon sugars (Brown 2003).

Improvement of bioethanol production through lignin modification

Lignin has been recognized for its negative impact on livestock forage quality, the pulp and paper industry, and, more recently, the conversion of lignocellulosic material to bioethanol. Motivated by the central role that lignin plays in the economics of these industries, much research in recent years have focused on the biosynthesis and conversion of lignin (Li et al. 2008). Most if not all of the steps involved in the biosynthesis of monolignols have been identified (Boerjan et al. 2003). Thus there are many potential genetic targets to alter the content and composition of lignin. In the last decades, by taking advantage of plant transformation technologies, researchers have successfully changed

lignin content and composition in a variety of plant species using genetic tools (Li et al. 2008). Details about their studies will be discussed as follows.

Generally, down regulation of genes involved in the early steps of monolignol biosynthetic pathway, such as PAL, C4H, HCT and C3'H, shows a large reduction of lignin content (Boerjan et al. 2003). Down regulation of F5H, CCoAOMT or COMT, which participate in converting G to S, greatly reduces the lignin S/G ratio but has little effect on lignin content (Li et al. 2008). However, altering lignin content or composition has its potential negative impacts. Many studies have demonstrated the negative impact of lowering lignin content yielding biomass with increased digestibility (Jones et al. 2001; Preston et al. 2004; Pedersen et al. 2005; Reddy et al. 2005). However, in respect to altering lignin composition, it is still debated on the relative benefits of plant digestibility in plants with altered lignin S/G ratio (Li et al. 2008). Chemically, G lignin can be more highly cross linked than S lignin. Therefore, the lignin network with lower S/G ratio could be more rigid and protect polysaccharides from degradation more efficiently. However there are several findings showing altered S/G radio does not affect cell wall digestibility. Arabidopsis *fah1* mutant, which lacks S lignin but has unaltered lignin content, does not show altered cell-wall digestibility (Jung et al., 1999). The study of Dixon's group even showed that lower S/G ratio resulted in higher digestibility. A 27% reduction in S/G ratio in COMT-suppressed tobacco caused a 6% increase in digestibility (Vailhe et al., 1996). Thus only examples which increase plant digestibility by decreasing lignin content are discussed here.

PAL and C4H are two enzymes, functioning at the early stage of the monolignol biosynthesis pathway. PAL uses phenylalanine as substrate to produce cinnamic acid

(Henderson and Friend 1979), which is converted to *p*-coumaric acid by C4H (Sewalt et al. 1997). Down regulation of PAL and C4H leads to large decrease of lignin content in transgenic tobacco (Sewalt et al. 1997).

HCT catalyzes the conversion of 4-coumaroyl CoA to 4-coumaroyl shikimate (Legrand et al. 2007). Down regulation of HCT causes decreased lignin content and high forage quality in transgenic alfalfa (Shadle et al. 2007).

The enzyme 4CL catalyzes the conversion from *p*-coumarate to *p*-coumaroyl-CoA (Bowles et al. 2001). Downregulation of 4CL causes lignin content reduction in tobacco (Kajita et al. 1997), Arabidopsis (Lee et al. 1997), and aspen (Hu et al. 1999).

C3'H is a cytochrome P450-dependent monooxygenase that catalyzes the 3'-hydroxylation of *p*-coumaroyl shikimate and p-coumaroyl quinate (Chapple et al. 2002). The lignin content of *Ref8*, a C3'H mutant of Arabidopsis, was reduced to 20-40% of wild-type levels, and *ref8* cell wall residue remaining after digestion of polysaccharide hydrolases was only one fifth of that found with wild-type cell walls (Franke et al. 2002). Mansfield's group also obtained similar result with their work with hybrid poplar (Coleman et al. 2008).

Even though the transgenic plants with low lignin content show improved digestibility, there is still concern about the fitness of the transgenic plants. Lignin is critically important to both plant structure and function. First, lignin provides the mechanical support for terrestrial plants (Campbell and Sederoff 1996). Second, lignin is involved in response to mechanical damage or pathogen infection in different spices (Bhuiyan et al. 2009). Pleiotropic effects on plants caused by lowered lignin content have

been observed (Pedersen et al. 2005). The effects include changes in stunted growth (Reddy et al. 2005), reduced pollen viability (Preston et al. 2004), and collapsed vessels (Jones et al. 2001). A recent review by Pedersen et al. summarized the impact of reduced lignin on plant fitness from eight different species, including annual crops and perennial forages (Pedersen et al, 2005). While most of the lignin-altered plants show negative impact, whereas some examples represent neutral or even positive impact. A higher carbohydrate content was observed in lignin-reduction lines, which is explained to be due to compensation mechanism on a mass balance basis (Chen and Dixon 2007). As the demand for bioethanol feedstock grows, there will be greater need to determine the impact of altered lignin in response to a variety of factors relevant to the environment in which plant are grown and genetic background in which they are placed (Pedersen et al, 2005).

Summary

The goal of this research focuses on developing a novel method which can facilitate circumventing the lignin barrier without compromising plant fitness. The strategy is by introducing proteins, which could be cross-linked into lignin network during lignification process, to the plant cell wall by genetic transformation. Protease treatment would then be applied to these transgenic poplars and the proteins associated with the lignin would then be hydrolyzed, resulting in a loosened lignin structure which facilitates subsequent steps in bioethanol production. Our previous work with this strategy successfully increased the poplar digestibility by introducing Tyr rich peptide (12% Tyr) to the plant cell wall (Liang et al., 2008). In the present study, I aim to further test this strategy and to seek transgenic plants with even more enhanced digestibility. In order to accomplish this goal, this research addresses the following topics:

1. Screening potential signal peptides for apoplastic secretion of heterologously-expressed proteins in poplar

2. Characterization of cross linkinig between amino acids/proteins with lignin

3. Characterization of digestibility of poplar by expression of a Tyr/Cys rich peptide in the cell wall

REFERENCES

Adler, E. (1977). Lignin Chemistry - Past, Present and Future. *Wood Sci Technol* **11**(3): 169-218.

Bhuiyan, N. H., G. Selvaraj, Y. Wei and J. King (2009). Role of lignification in plant defense. *Plant Signal Behav* **4**(2): 158-9.

Boerjan, W., J. Ralph and M. Baucher (2003). Lignin Biosynthesis. *Annual Review of Plant Biology* **54**(1): 519-546.

Boudet, A. (2003). Lignins and lignocellulosics: a better control of synthesis for new and improved uses. *Trends in Plant Science* **8**(12): 576-581.

Bowles, D. J., E. K. Lim, Y. Li, A. Parr, R. Jackson and D. A. Ashford (2001). Identification of glucosyltransferase genes involved in sinapate metabolism and lignin synthesis in Arabidopsis. *J Biol Chem* **276**(6): 4344-4349.

Brown, R. C. (2003). Biorenewable resources.

Bruce Alberts, D. B., Julian Lewis, Martin Raff, Keith Roberts, and James D Watson (1994). *Molecular Biology of the Cell*. New York.

Burton, R. A., M. J. Gidley and G. B. Fincher (2010). Heterogeneity in the chemistry, structure and function of plant cell walls. *Nat Chem Biol* **6**(10): 724-732.

Campbell, M. M. and R. R. Sederoff (1996). Variation in lignin content and composition - Mechanism of control and implications for the genetic improvement of plants. *Plant Physiol* **110**(1): 3-13.

Cassab, G. I. (1998). Plant cell wall proteins. *Annu Rev Plant Phys* **49**: 281-309.

Chapple, C., R. Franke, J. M. Humphreys, M. R. Hemm, J. W. Denault, M. O. Ruegger and J. C. Cusumano (2002). The Arabidopsis REF8 gene encodes the 3-hydroxylase of phenylpropanoid metabolism. *Plant J* **30**(1): 33-45.

Chen, F. and R. A. Dixon (2007). Lignin modification improves fermentable sugar yields for biofuel production. *Nat Biotechnol* **25**(7): 759-761.

Chen, J. and J. E. Varner (1985). Isolation and Characterization of Cdna Clones for Carrot Extensin and a Proline-Rich 33-Kda Protein. *P Natl Acad Sci USA* **82**(13): 4399-4403.

Cheng, J. Y. and Y. Sun (2002). Hydrolysis of lignocellulosic materials for ethanol production: a review. *Bioresource Technology* **83**(1): 1-11.

Coleman, H. D., J. Y. Park, R. Nair, C. Chapple and S. D. Mansfield (2008). RNAi-mediated suppression of p-coumaroyl-CoA 3'-hydroxylase in hybrid poplar impacts lignin deposition and soluble secondary metabolism. *Proceedings of the National Academy of Sciences* **105**(11): 4501-4506.

Cosgrove, D. J. (2005). Growth of the plant cell wall. *Nat Rev Mol Cell Bio* **6**(11): 850-861.

Davin, L. B. and N. G. Lewis (2000). Dirigent proteins and dirigent sites explain the mystery of specificity of radical precursor coupling in lignan and lignin biosynthesis. *Plant Physiol* **123**(2): 453-461.

Demirbas, A. (2008). Biofuels sources, biofuel policy, biofuel economy and global biofuel projections. *Energy Conversion and Management* **49**(8): 2106-2116.

Dharmawardhana, D. P., B. E. Ellis and J. E. Carlson (1995). A Beta-Glucosidase from Lodgepole Pine Xylem Specific for the Lignin Precursor Coniferin. *Plant Physiol* **107**(2): 331-339.

Doblin, M. S., K. I, D. Jacob-Wilk and D. P. Delmer (2002). Cellulose biosynthesis in plants: from genes to rosettes. *Plant Cell Physiol* **43**(12): 1407-1420.

Franke, R., M. R. Hemm, J. W. Denault, M. O. Ruegger, J. M. Humphreys and C. Chapple (2002). Changes in secondary metabolism and deposition of an unusual lignin in the ref8 mutant of Arabidopsis. *Plant J* **30**(1): 47-59.

Fry, S. C. (1986). Cross-Linking of Matrix Polymers in the Growing Cell-Walls of Angiosperms. *Annu Rev Plant Phys* **37**: 165-186.

Galbe, M. and G. Zacchi (2007). Pretreatment of lignocellulosic materials for efficient bioethanol production. *Adv Biochem Eng Biot* **108**: 41-65.

Goldemberg, J. (2007). Ethanol for a Sustainable Energy Future. *Science* **315**(5813): 808-810.

Gray, K. A., L. S. Zhao and M. Emptage (2006). Bioethanol. *Curr Opin Chem Biol* **10**(2): 141-146.

Harholt, J., A. Suttangkakul and H. Vibe Scheller (2010). Biosynthesis of Pectin. *Plant Physiol* **153**(2): 384-395.

Henderson, S. J. and J. Friend (1979). Increase in Pal and Lignin-Like Compounds as Race-Specific Resistance Responses of Potato-Tubers to Phytophthora-Infestans. *Phytopathol Z* **94**(4): 323-334.

Himmel, M. E., S. Y. Ding, D. K. Johnson, W. S. Adney, M. R. Nimlos, J. W. Brady and T. D. Foust (2007). Biomass recalcitrance: Engineering plants and enzymes for biofuels production. *Science* **315**(5813): 804-807.

Hosel, W. and R. Todenhagen (1980). Characterization of a Beta-Glucosidase from Glycine-Max Which Hydrolyses Coniferin and Syringin. *Phytochemistry* **19**(7): 1349-1353.

Hu, W. J., S. A. Harding, J. Lung, J. L. Popko, J. Ralph, D. D. Stokke, C. J. Tsai and V. L. Chiang (1999). Repression of lignin biosynthesis promotes cellulose accumulation and growth in transgenic trees. *Nat Biotechnol* **17**(8): 808-812.

Jansson, S. and C. J. Douglas (2007). Populus: A Model System for Plant Biology. *Annual Review of Plant Biology* **58**(1): 435-458.

Jones, L., A. R. Ennos and S. R. Turner (2001). Cloning and characterization of irregular xylem4 (irx4): a severely lignin-deficient mutant of Arabidopsis. *Plant J* **26**(2): 205-216.

Jung, H. J. G., W. T. Ni, C. C. S. Chapple and K. Meyer (1999). Impact of lignin composition on cell-wall degradability in an Arabidopsis mutant. *J Sci Food Agr* **79**(6): 922-928.

Kajita, S., S. Hishiyama, Y. Tomimura, Y. Katayama and S. Omori (1997). Structural characterization of modified lignin in transgenic tobacco plants in which the activity of 4-coumarate:coenzyme A ligase is depressed. *Plant Physiol* **114**(3): 871-879.

Lee, D., K. Meyer, C. Chapple and C. J. Douglas (1997). Antisense suppression of 4-coumarate:coenzyme A ligase activity in Arabidopsis leads to altered lignin subunit composition. *Plant Cell* **9**(11): 1985-1998.

Legrand, M., S. Besseau, L. Hoffmann, P. Geoffroy, C. Lapierre and B. Pollet (2007). Flavonoid accumulation in Arabidopsis repressed in lignin synthesis affects auxin transport and plant growth. *Plant Cell* **19**(1): 148-162.

Lerouxel, O., D. M. Cavalier, A. H. Liepman and K. Keegstra (2006). Biosynthesis of plant cell wall polysaccharides - a complex process. *Curr Opin Plant Biol* **9**(6): 621-630.

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

Li, X., N. D. Bonawitz, J. K. Weng and C. Chapple (2010). The Growth Reduction Associated with Repressed Lignin Biosynthesis in Arabidopsis thaliana Is Independent of Flavonoids. *The Plant Cell Online* **22**(5): 1620-1632.

Liang, H. Y., C. J. Frost, X. P. Wei, N. R. Brown, J. E. Carlson and M. Tien (2008). Improved sugar release from lignocellulosic material by introducing a tyrosine-rich cell wall peptide gene in poplar. *Clean-Soil Air Water* **36**(8): 662-668.

Lord, E. M. and L. C. Sanders (1992). Roles for the Extracellular-Matrix in Plant Development and Pollination - a Special Case of Cell-Movement in Plants. *Dev Biol* **153**(1): 16-28.

Marcinowski, S. and H. Grisebach (1978). Enzymology of Lignification - Cell-Wall-Bound Beta-Glucosidase for Coniferin from Spruce (Picea-Abies) Seedlings. *Eur J Biochem* **87**(1): 37-44.

Marcus, A., J. Greenberg and V. Averyhartfullard (1991). Repetitive Proline-Rich Proteins in the Extracellular-Matrix of the Plant-Cell. *Physiol Plantarum* **81**(2): 273-279.

Mccann Mc, R. K. (1991). Architecture of the primary cell wall. *In The Cytoskeletal Basis of Plant Growth and Form, ed. CW Lloyd*: 21.

Mcconnell, J. R., R. Edwards, G. L. Kok, M. G. Flanner, C. S. Zender, E. S. Saltzman, J. R. Banta, D. R. Pasteris, M. M. Carter and J. D. W. Kahl (2007). 20th-century industrial black carbon emissions altered arctic climate forcing. *Science* **317**(5843): 1381-1384.

Mcneil, M., A. G. Darvill, S. C. Fry and P. Albersheim (1984). Structure and Function of the Primary-Cell Walls of Plants. *Annu Rev Biochem* **53**: 625-663.

Mohnen, D. (2008). Pectin structure and biosynthesis. *Curr Opin Plant Biol* **11**(3): 266-277.

Morreel, K., H. Kim, F. C. Lu, O. Dima, T. Akiyama, R. Vanholme, C. Niculaes, G. Goeminne, D. Inze, E. Messens, J. Ralph and W. Boerjan (2010). Mass Spectrometry-Based Fragmentation as an Identification Tool in Lignomics. *Anal Chem* **82**(19): 8095-8105.

Mosier, N. (2005). Features of promising technologies for pretreatment of lignocellulosic biomass. *Bioresource Technology* **96**(6): 673-686.

Mosier, N., C. Wyman, B. Dale, R. Elander, Y. Y. Lee, M. Holtzapple and M. Ladisch (2005). Features of promising technologies for pretreatment of lignocellulosic biomass. *Bioresource Technology* **96**(6): 673-686.

N. Raikhel, M. J. C. (2000). Protein sorting and vesicle traffic. *Biochemistry & Molecular Biology of Plants*.

Neumann, U., F. Brandizzi and C. Hawes (2003). Protein transport in plant cells: In and out of the Golgi (vol 92, pg 167, 2003). *Ann Bot-London* **92**(3): 475-475.

Pedersen, J. F., K. P. Vogel and D. L. Funnell (2005). Impact of reduced lignin on plant fitness. *Crop Sci* **45**(3): 812-819.

Persson, S. and A. Endler (2011). Cellulose Synthases and Synthesis in Arabidopsis. *Mol Plant* **4**(2): 199-211.

Preston, J., J. Wheeler, J. Heazlewood, S. F. Li and R. W. Parish (2004). AtMYB32 is required for normal pollen development in Arabidopsis thaliana. *Plant J* **40**(6): 979-995.

Reddy, M. S. S., F. Chen, G. Shadle, L. Jackson, H. Aljoe and R. A. Dixon (2005). Targeted down-regulation of cytochrome P450 enzymes for forage quality improvement in alfalfa (Medicago sativa L.). *P Natl Acad Sci USA* **102**(46): 16573-16578.

Ringli, C., B. Keller and U. Ryser (2001). Glycine-rich proteins as structural components of plant cell walls. *Cell Mol Life Sci* **58**(10): 1430-1441.

Sewalt, V. J. H., W. T. Ni, J. W. Blount, H. G. Jung, S. A. Masoud, P. A. Howles, C. Lamb and R. A. Dixon (1997). Reduced lignin content and altered lignin composition in transgenic tobacco down-regulated in expression of L-phenylalanine ammonia-lyase or cinnamate 4-hydroxylase. *Plant Physiol* **115**(1): 41-50.

Shadle, G., F. Chen, M. S. Srinivasa Reddy, L. Jackson, J. Nakashima and R. A. Dixon (2007). Down-regulation of hydroxycinnamoyl CoA: Shikimate hydroxycinnamoyl transferase in transgenic alfalfa affects lignification, development and forage quality. *Phytochemistry* **68**(11): 1521-1529.

Showalter, A. M. (1993). Structure and Function of Plant-Cell Wall Proteins. *Plant Cell* **5**(1): 9-23.

Showalter, A. M. (2001). Arabinogalactan-proteins: structure, expression and function. *Cell Mol Life Sci* **58**(10): 1399-1417.

Steeves, V., H. Forster, U. Pommer and R. Savidge (2001). Coniferyl alcohol metabolism in conifers - I. Glucosidic turnover of cinnamyl aldehydes by UDPG: coniferyl alcohol glucosyltransferase from pine cambium. *Phytochemistry* **57**(7): 1085-1093.

Taylor, G. (2002). Populus: Arabidopsis for forestry. Do we need a model tree? *Ann Bot-London* **90**(6): 681-689.

Taylor, N. G. (2008). Cellulose biosynthesis and deposition in higher plants. *New Phytol* **178**(2): 239-252.

Thompson, R., J. Swana, Y. Yang and M. Behnam (2011). An analysis of net energy production and feedstock availability for biobutanol and bioethanol. *Bioresource Technology* **102**(2): 2112-2117.

Topal, E. and S. Shafiee (2009). When will fossil fuel reserves be diminished? *Energ Policy* **37**(1): 181-189.

Turner, S. R., D. M. Brown, Z. N. Zhang, E. Stephens and P. Dupree (2009). Characterization of IRX10 and IRX10-like reveals an essential role in glucuronoxylan biosynthesis in Arabidopsis. *Plant J* **57**(4): 732-746.

Vailhe, M. A. B., C. Migne, A. Cornu, M. P. Maillot, E. Grenet, J. M. Besle, R. Atanassova, F. Martz and M. Legrand (1996). Effect of modification of the O-methyltransferase activity on cell wall composition, ultrastructure and degradability of transgenic tobacco. *J Sci Food Agr* **72**(3): 385-391.

Vitale, A. and J. Denecke (1999). The endoplasmic reticulum - Gateway of the secretory pathway. *Plant Cell* **11**(4): 615-628.

Walter, P. and A. E. Johnson (1994). Signal Sequence Recognition and Protein Targeting to the Endoplasmic-Reticulum Membrane. *Annu Rev Cell Biol* **10**: 87-119.

Worldwatch Institute, C. F. A. P. (2006). The Renewable Path to Energy Security.

Wyman, C. E. (1994). Ethanol from Lignocellulosic Biomass - Technology, Economics, and Opportunities. *Bioresource Technology* **50**(1): 3-16.

Zaldivar, J., J. Nielsen and L. Olsson (2001). Fuel ethanol production from lignocellulose: a challenge for metabolic engineering and process integration. *Appl Microbiol Biot* **56**(1-2): 17-34.

Zhong, R. Q. and Z. H. Ye (2007). Regulation of cell wall biosynthesis. *Curr Opin Plant Biol* **10**(6): 564-572.

CHAPTER 2

Screening Potential Signal Peptides for Apoplastic Secretion of Heterologously-Expressed Proteins in Poplar

INTRODUCTION

The apoplast is the region of plant cell outside the cell membrane, including cell walls as well as the extracellular spaces. Apoplastic proteins are transported out of the plant cells via secretory pathway, after being synthesized inside the cell (Dafoe et al. 2009). As part of this process, all secretory proteins contain a hydrophobic signal peptide at the N-terminus directing entry of the protein into the endoplasmic reticulum (ER) (Walter et al. 1994). Preceded by a signal peptide, secretory proteins enter ER lumen co-translationally through an aqueous pore in the protein translocator on ER. Once the signal peptide has crossed the ER membrane, it is cleaved by a signal peptidase, an enzyme located on the lumenal surface of the ER membrane (Vitale et al. 1999). Apoplastic proteins are involved in a number of essential cellular functions: cell extension and growth (Humphrey et al. 2007), cell-cell recognition (Rubinstein et al. 1995), and plant response to biotic and abiotic stresses (Zhang et al. 2000).

Given the significance of apoplast-associated processes, efforts have been made to identify apoplast-localized proteins and to determine their role in the plant cell. Attempts have been made to identify apoplastic proteins through bioinformatic analysis. As a means of identifying new cell wall proteins, Davis et al. (Davis et al. 2009) mined the Arabidopsis

genome for proteins with unknown functions with N-terminal signal peptide. The top ten genes passing the bioinformatic analysis were then tested for subcellular localization using green fluorescent protein (GFP). While six of the genes resulted in expression of GFP, none were above to express the GFP into the cell wall. The study by Lease et al. has also used bioinformatic tools to identify apoplastic proteins in Arabidopsis (Lease et al. 2006). However, this study did not obtain experimental verification of subcellular localization.

An alternative method for identifying cell-wall-associated proteins is through proteomics. Recently, proteomic method was used to identify novel apoplastic proteins in Arabidopsis, rice and poplar (Watson et al. 2004; Boudart et al. 2005; Dafoe and Constabel 2009). While proteomics can be argued to provide the final proof of subcellular localization (superior to predictive bioinformatic methods), proteomic methods are only as good as the method of isolation. This method can be problematic due to the contamination from cytosolic proteins during the sample preparation. Indeed, the criteria for identifying subcellular localization using bioinformatics typically involve stringent controls (e.g. marker enzyme assays) (Guo et al. 2009).

In the present study, we characterize eight putative signal peptides for their ability to direct the protein's apoplast localization. One of the eight signal peptides is from a putative *Populus* coniferin β-glucosidase (CβG) homolog, which was screened from the poplar genome with BLAST and alignment analysis with *Pinus* CβG sequence. The other seven proteins were chosen by mining the poplar xylem sap proteomic data (Dafoe and Constabel 2009). We fused signal peptides with enhanced green fluorescent protein (EGFP) to visualize subcellular localization. Subcellular localization of these signal peptides was investigated with transient expression experiments by bombardment of onion epidermal

cell and vacuum infiltration of poplar leaves. Our results showed that all eight signal peptides are able to direct the EGFP into the apoplastic space and that proteomic data is much more reliable as an indicator of apoplastic localization.

MATERIALS AND METHODS

Identification and synthesis of the signal-peptide-encoding sequences. A TBLASTN search of the masked genome of *Populus trichocarpa* in Phytozome (v6.0, http://www.phytozome.net/) with the *Pinus* CβG (Dharmawardhana et al. 1995) resulted in 15 scaffolds (e-value cut off = 4.2e-7). The top three sequences with most similarity were chosen for an alignment analysis with *Pinus* CβG. DNA fragment encoding the signal peptides chosen for this study plus one more extra amino acid following predicted cleavage site was identified with SignalP 3.0 server (http://www.cbs.dtu.dk/services/SignalP/) and then sent for synthesis by Integrated DNA Technologies (Integrated DNA Technologies, IA, USA). Signal peptides in this study were also screened by Scan-Prosite (http://www.expasy.ch/tools/scanprosite/) to eliminate the possibility of ER retention. DNA fragment were received in the pSMART plasmid, flanked by KpnI at 3' and XmaI and NcoI at 5'.

Plasmid construction. The EGFP-containing plasmid used in this study is P2113 (Guiltinan et al. 2010) (generously provided by Mark Guiltinan, Penn State University), which contains a CaMV35S promoter, the coding sequence of EGFP, and a 35S terminator in the PUC19 backbone. Each DNA sequence encoding signal peptide was cut from pSMART (Integrated DNA Technologies, Inc, USA), using KpnI and NcoI and cloned between the CaMV35S promoter and EGFP coding sequence. The constructs were verified by sequencing. P2113 without insertion of signal peptide was used as control. The gene cassette of CaMV35S promoter, the coding sequence of signal peptide, the coding sequence of EGFP, and a 35S terminator was then cloned into binary vector pCAMBIA 1300 for poplar transformation. pCAMBIA 1300 cloned with the gene cassette of

CaMV35S promoter, the coding sequence of EGFP, and a 35S terminator was used as control.

Transient expression in onion epidermal cells. The P2113 vectors containing the signal peptide-EGFP fusion and only containing EGFP were introduced into onion epidermal cells with a Model PDS-1000/He Biolistic particle delivery system (Biorad, Hercules, CA, USA). A tungsten M-10 (0.7 micron) (Biorad, Hercules, CA, USA) and an 1100-psi rapture disk (Biorad, Hercules, CA, USA) was used for bombardment. A total of 1.8 mg of tungsten particles coated with 1.5 μg of DNA was prepared per shot. After bombardment, onion peels were incubated for 18 h at 22 ℃ in the dark. Plasmolysis of onion cells was induced by soaking onion peels in 0.8 M sucrose for 5 min and then 0.5 M sucrose for 10 min.

Poplar infiltration. This procedure was performed according to Meximova et al. with modification (Guiltinan et al. 2003). *Agrobacterium tumefaciens* strain AGL1 was grown at 25 ℃ in 20 ml of Luria-Bertani (LB) media supplemented with 50 μg/ml kanamycin. The cells were harvested ($A_{420\ nm}$=1) by centrifugation at 5,000×g for 5 min at room temperature and resuspended in 40 ml of induction media, containing 3% sucrose, 0.1% glucose, 0.0001 M acetosyringone, 0.0013 M proline, and DKW basal medium (PhytoTechnology Laboratories, KS, USA). After incubation for 5 h at 25 ℃, poplar leaves (20 mm×40 mm) were added and a vacuum was applied for three minutes. The infiltrated poplar leaves were placed on three layers of pre-wet filter papers and incubated in a sealed Petri dish for three days under light at 25 ℃.

EGFP visualization by confocal microscopy. The subcellular localization of signal peptide-EGFP fusions was observed with an Olympus FV1000 Laser Scanning Confocal Microscope (Olympus America Inc., Melville, NY, USA). EGFP-488nm was applied as excitation filter and 500-600 nm wavelength of emission light were detected. Images were taken and exported with the OLYMPUS FLUOVIEW software.

Histogram quantification with ImageJ. EGFP-associated green fluorescence in apoplastic region and in the whole cell was quantified by ImageJ.

RESULTS

Identification of signal peptides. Our previous work utilized a CβG signal peptide from *Pinus* for extracellular transportation (Liang et al. 2008). *Pinus* CβG (Dharmawardhana et al. 1995; Dharmawardhana et al. 1999) is a cell-wall-localized protein, involved in releasing coniferyl alcohol from a glycosylated form to a monolignol form (Dharmawardhana et al. 1999). In the present study, we aimed to identify and utilize a CβG signal peptide from *Populus*. To identify the CβG family in *Populus*, we conducted BLAST analysis in genome of *Populus trichocarpa* with the *Pinus* CβG sequence. A total of 15 scaffolds were obtained from the *Populus* genome. The top three sequences with the most similarity to the *Pinus* CβG were subjected to alignment analysis. The BLAST and alignment results are shown in Fig. 2-1. The Pt555358 shows the highest similarity with the *Pinus* CβG. Additionally, among the top three hits to the *Pinus* CβG, the same homolog, Pt555358, is the only sequence encoding a protein containing a putative signal peptide predicted with SignalP (Fig. 2-1B). Therefore, Pt555358 is most likely the CβG in *Populus* and predicted to have a similar function and localization pattern with *Pinus* CβG. The signal peptide portion of Pt555358 was chosen for further study.

A:

Homologs in *Populus trichocarpa*	*E*-value to *Pinus* CβG (Pc_AAC69619)
Pt555358	5e-149
Pt196698	8e-141
Pt284289	3e-136

B:

```
Pt284289        ----------------------------------SRNSFPDDFVFGTSSSAYQYEGETNKH
Pc_AAC69619     MEVSVLMWVLLFYSLLGFQVTTA--------RLDRNNFPSDFMFGTASSAYQYEGAVRED
Pt555358        MGISSLCKALILLELILLPLFASSDTKTLHESSDSSSFPANFLFGTASSSYQFEGAYLSD
Pt196698        ----------------------------------DNSAFPSNFLFGTASSSYQFEGAYLSH
                                                  . . ** :*:***:**:**:**   .

Pt284289        GRGPAIWDTFTVEHTERINDHSNGNVAVDFYHRYKEDVQRMKEMGMDAFRFSISWSRVLP
Pc_AAC69619     GKGPSTWDALT-HMPGRIKDSSNGDVAVDQYHRYMEDIELMASLGLDAYRFSISWSRILP
Pt555358        GKGLSNWDVHT-HKPGNIIDGSNGDIAVDQYHRYLEDIELMASLGVNSYRFSMSWARILP
Pt196698        GKGLSNWDVFT-HKPGTIMDGTNGDIAVDHYHRYPEDLDLMEYIGVNSYRVSISWARILP
                *.* : **. *   .  * * :**::*** **** **:: *  :*::::*.*:**:*:**

Pt284289        HGRLSAGVNEEGIKFYNDLIDDLLKNGLQPYVTLFHWDTPQALEDKYGGFLSPNIVNDFR
Pc_AAC69619     EGR--GEINMAGIEYYNNLIDALLQNGIQPFVTLFHFDLPKALEDSYGGWLSPQIINDFE
Pt555358        KGRF-GGVNMAGISYYNKLINALLLKGIQPFVSLTHFDVPQELEDRYGGFLSPKSQEDFG
Pt196698        KGRF-GTANRAGINHYNKFINELLRRGIQPFVTLTHYDIPQELEDRYGAWLSPEIQQDFK
                 **  .  *  **..**.:*: ** .*:**:*:* *:* *: *** **.:***:  :**

Pt284289        DFVDLCFQNFGDRVKKWITLNEPWMFSVQGYDMGTMAPGRISVVVNDPHRSLNT-GATEV
Pc_AAC69619     AYAEICFRAFGDRVKYWATVNEPNLFVPLGYTVGIFPPTRCAAPHANPLCMTGNCSSAEP
Pt555358        YYVDICFKYFGDRVKYWATFNEPNFQAIYGYRVGECPPKRCSKPFGN--CSHGD-SEAEP
Pt196698        YYADICFKSFGDRVKYWTTFNEPNVAAIRGYRSGIFPPSRCSGTFGY--CSSGD-SEREP
                 :.::**. ****** * *.***..    **  *  .* * :          .  .  *

Pt284289        YTVSHHLLLAHAAAVKLYKEKYQSCQGGQIGITLVSHWFEPYSNSEADQNATKRSLDFML
Pc_AAC69619     YLAAHHVLLAHASAVEKYREKYQKIQGGSIGLVISAPWYEPLENSPEERSAVDRILSFNL
Pt555358        FIAAHNIILAHATAVDIYRTKYQREQRGSIGIVMNCMWYEPISNSTANKLAVERALAFFL
Pt196698        FIAAHNMILSHAAAVNVYRTKYQKKQGGSIGIVMNAIWHEPISDSLEDKLAVERANAFYM
                : .:*::::*:**:**. *. ***  * *.**:.: . *.** .:*  :. *..*   * :

Pt284289        GWFMDPLTNGDYPRNMHDFVGGRLPEFTAEESKMLKGSYDFIGINYYTTYYAQNI--DAN
Pc_AAC69619     RWFLDPIVFGDYPQEMRERLGSRLPSISSELSAKLRGSFDYMGINHYTTLYATSTPPLSP
Pt555358        RWFLDPIIFGRYPEEMKKVLGSTLPEFSRNDMNKLRKGLDFIGMNHYTSYYVQDCILSVC
Pt196698        NWFLDPIILGKYPTEMRETLGSDLPVFSKYELEKLKSGVDFIGINQYTSFYVKDCMFSTC
                 **:**:  * ** :*.. :*. ** ::      *. . *::*:* **: *. .

Pt284289        YQSVGFMSDARANW-TGERNGIPIGPQAGVKWLYIYPEGISRLLNYTKDLYGSPTIYITE
Pc_AAC69619     DHTQYLYPDSRVYL-TGERHGVSIGERTGMDGLFVVPHGIQKIVEYVKEFYDNPTIIIAE
Pt555358        EPGKGSTRTEGSSLLTQEKDGVPIGKPSEVDWLHVYPQGMEKMVTYVKERYNNTPMIITE
Pt196698        EQGPGVSKTEGLYLRTAQKDGFFIGQPTALDWLHVYPQGMEKLVTYFKDRYNNIPMYITE
                              * :. *. **  : :. *.: * *:..:: * *: *.. .: *:*

Pt284289        NGVDDVNNNASSLKEALNDPIREKSYKDHLKNVLRSINEHGVDVKGFFAWSLMDNFEWGS
Pc_AAC69619     NGYPESEESSSTLQENLNDVRRIRFHGDCLSYLSAAIKN-GSDVRGYFVWSLLDNFEWAF
Pt555358        NGYAQVSNSNGNIEEFLHDTGRVEYMSGYLDALLTAMKK-GADVRGYFAWSFLDNFEWTF
Pt196698        NGYCDEENVNVTTKAVLKDVQRVEYMSSYLDALETAVRK-GADVRGYFAWSLLDNFEWTS
                **  : .:   . :  *:*  *     . *. :  ::.: * **.*:*.**::*****

Pt284289        GYAVRFGLYYVDYKNDLKRYPKQSVKWFKQFL--------------
Pc_AAC69619     GYTIRFGLYHVDFISDQKRYPKLSAQWFRQFLQHDDQGSIRSSSSI
Pt555358        GYTRRFGLYHVDY-TTMKRTPRLSATWYKEFIARYKVDKSQM----
Pt196698        GYTIRFGLYHVDF-STLKRTRKLSATWYKDYI--------------
                **: *****:**: .  **  . *. *:.:::
```

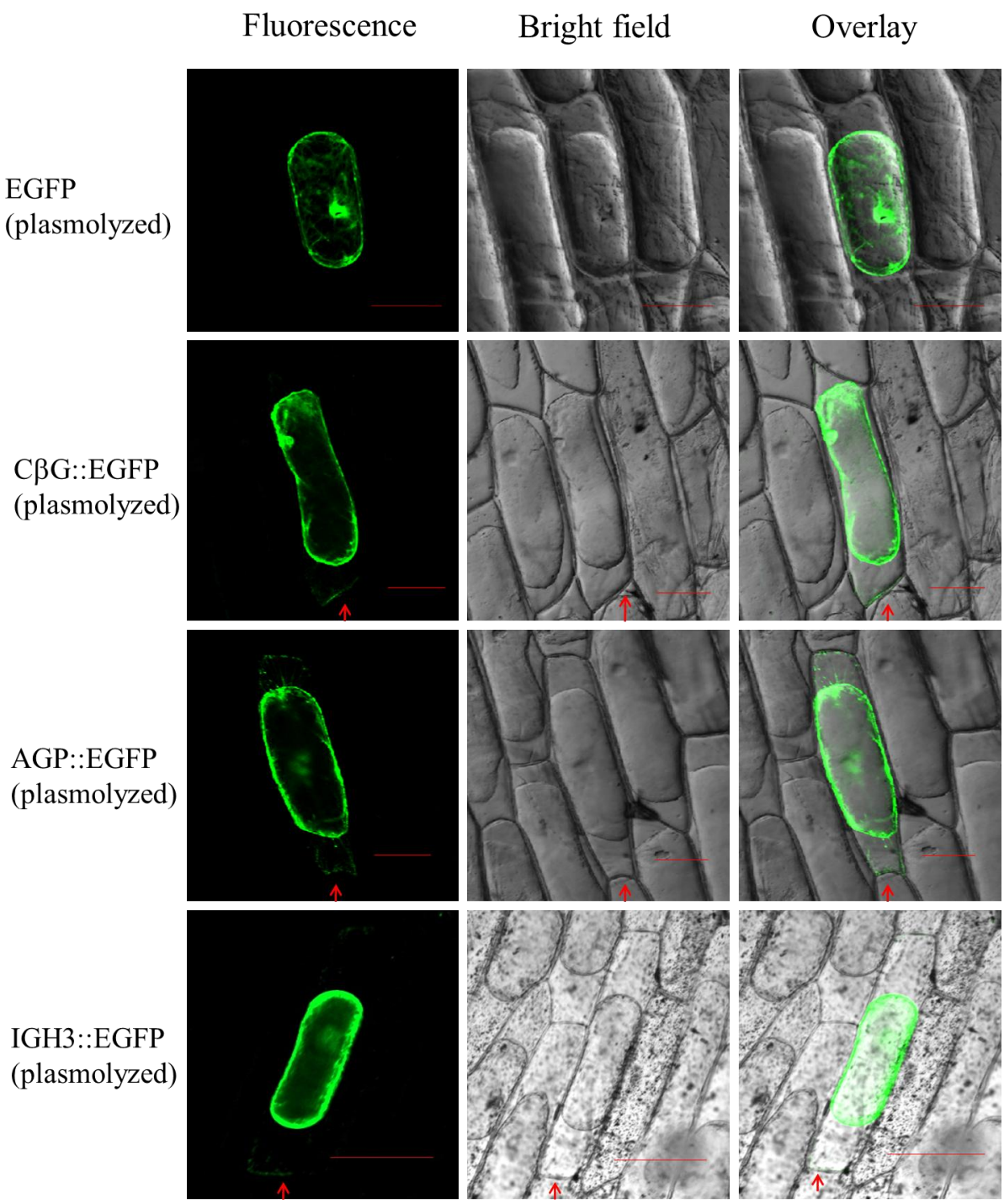
Figure 2-1. BlAST and alignment analysis with *Pinus* CβG sequence (Pc_AAC69619) in *Populus trichocarpa* genome. (A) E values of the top three poplar CβG homologs (Pt555358, Pt196698, and Pt284289) with highest similarity to *Pinus* CβG sequence (query sequence) in the BLAST result. (B) Alignment analysis of Pc_AAC69619, Pt555358, Pt196698, and Pt284289 with MUSCLE (Edgar 2004).

In addition to the *Populus* CβG, we also selected seven other *Populus* signal peptides from proteins found to be localized in the xylem sap (Dafoe and Constabel 2009). The portion of the signal peptide was predicted with SignalP and also screened by Scan-Prosite to eliminate the possibility of ER retention. The signal peptide plus one additional amino acid after the predicted cleavage site was fused with EGFP for the purpose of testing its ability to direct the protein's apoplast secretion. The eight signal peptides tested in this study are listed in Table 2-1.

Origin of signal peptide	JGI protein ID	Secretion prediction by signalP	Signal peptide portion
Putative Coniferin β-glucosidase (CβG)	Not available	0.998	MGISSLCKALILLELILLPLFASSDT/K
Fasciclin-like arabinogalactan protein 10 precursor (AGP)	730906	0.999	MKHHFSVFLFPAILLLLHCTQTLS/Q
Beta-Ig-H3/fasciclin (IGH3)	728480	1.000	MKQQISLFSFSLVLLFLHCTQTLS/Q
Polygalacturonase (glycoside hydrolase, family 28) (POLY)	644497	1.000	MSLQVVLALLVVAFISFRGAES/R
Plant basic secretory protein (BASIC)	549955	1.000	MFHHVYVLSLLVLLAINTVSA/V
Serine/threonine protein kinase (SERINE)	662643	1.000	MMFLVVFLLTLLSHVPALDA/C
Cupin (CUPIN)	591472	1.000	MRSVHFLLAFVLLTLASSIASA/Y
Kunitz trypsin inhibitor (KUNITZ)	746481	1.000	MNYPMLLLCLLLLAFACTKQSIAAA/E

Table 2-1. The list of signal peptides tested in this study. Predicted cleavage site were indicated by slash.

Analysis of signal peptides with onion epidermal cells. These eight signal peptides were cloned into P2113 vector. The P2113 vector contains the strong 35S promoter (Guiltinan et al. 2010), which is then followed by the coding sequence of the eight candidate signal peptides and then the EGFP coding sequence. These were then transformed into onion epidermal cells. The empty P2113 vector was used as a negative control. The ability of the signal peptides to direct secretion of EGFP was examined by confocal microscope. The transformed onion epidermal cells showed strong green fluorescence, as shown in Fig. 2-2. To confirm that the putative signal peptides indeed directed apoplastic secretion of EGFP, we induced plasmolysis, to separate the cell membrane from the cell wall. This allowed for better resolution between a protein that is secreted to apoplast and that associated with the cellular membrane. After plasmolysis with 0.5 M and 0.8 M sucrose, the green fluorescence was clearly visible, extracellularly in the apoplastic region, as shown in Fig. 2-2. For the empty vector control, EGFP was not present in the extracellular region (Fig. 2-2). Our result indicates that the eight signal peptides are able to transport heterologous protein to apoplastic region.


Fluorescence
Bright field
Overlay
EGFP (plasmolyzed)
CβG::EGFP (plasmolyzed)
AGP::EGFP (plasmolyzed)
IGH3::EGFP (plasmolyzed)

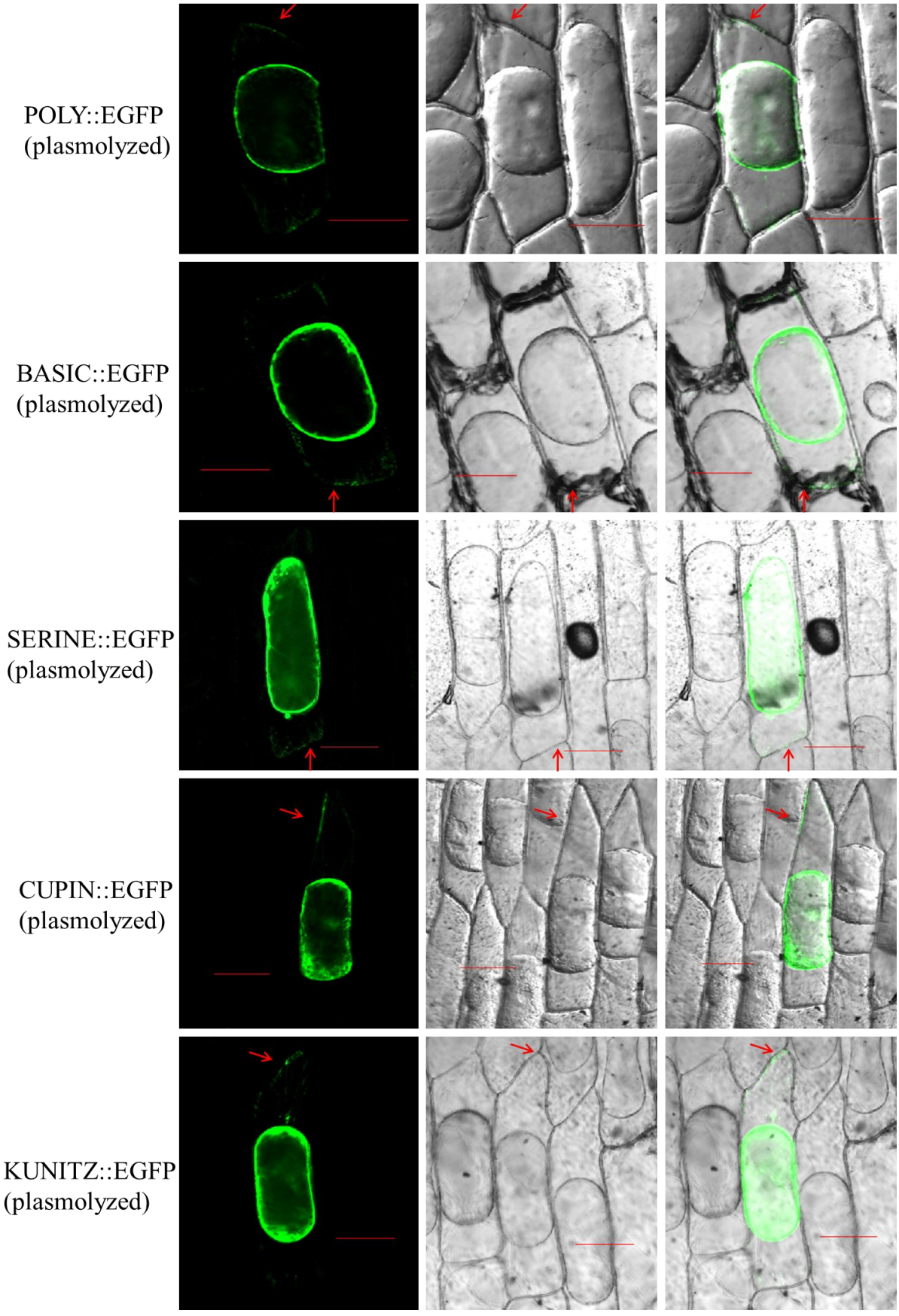

POLY::EGFP
(plasmolyzed)
BASIC::EGFP
(plasmolyzed)
SERINE::EGFP
(plasmolyzed)
CUPIN::EGFP
(plasmolyzed)
KUNITZ::EGFP
(plasmolyzed)

Fig. 2-2. EGFP localization in plasmolyzed onion epidermal cells transformed with signal peptide:EGFP observed by confocal microscope. Plasmolysis of onion epidermal cells was induced by soaking onion peels in 0.8M sucrose for 5 min and 0.5M sucrose for 10 min. EGFP vector served as the negative control. Scale bar represents 100 μm

To quantify the amount of fusion protein transported out of the cell, we used ImageJ to count the pixels in the color range of green in the whole cell and extracellular space. The ratio of the extracellular EGFP intensity to that in the whole cell was then determined. As shown in Fig. 2-3, significant differences in ratios of the extracellular EGFP to that in the whole cell was observed between expression of EGFP with and without signal peptides (based on ANOVA one factor analysis with $p<0.05$).

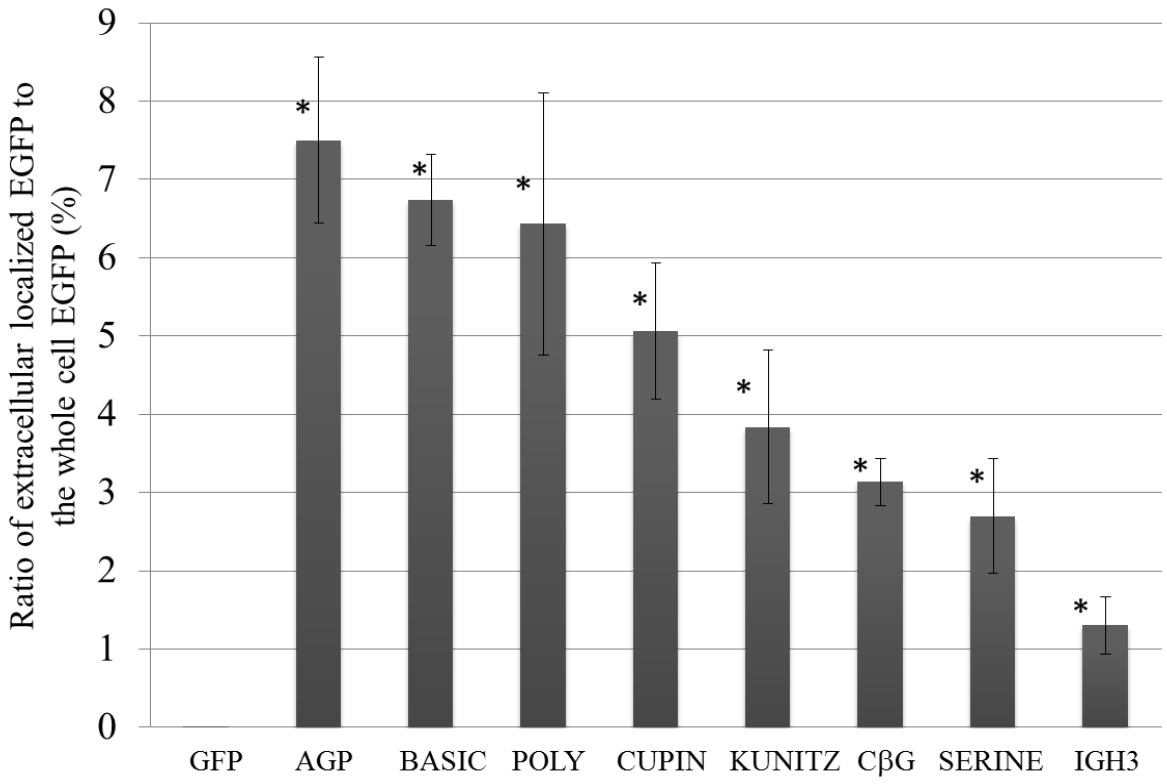
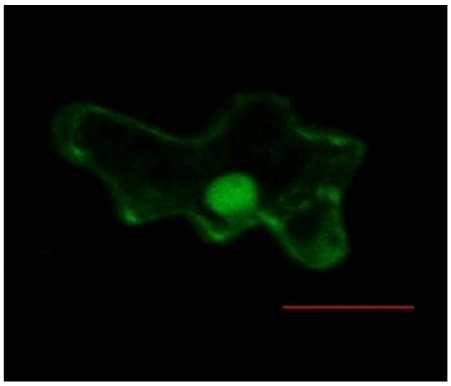
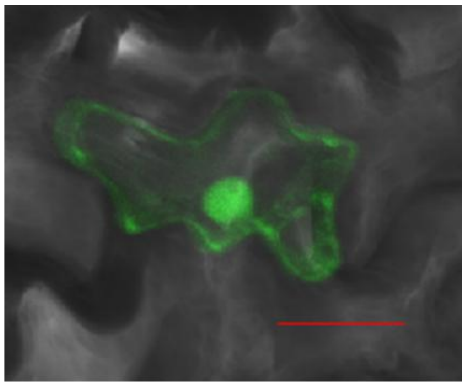


Figure 2-3. The ratio of green fluorescence intensity in apoplast to that in the whole cell, quantified by ImageJ. Data represent means ± SE of three replicates. Significant difference between EGFP and each signal peptide fusion protein was determined by ANOVA one factor test ($p<0.05$).

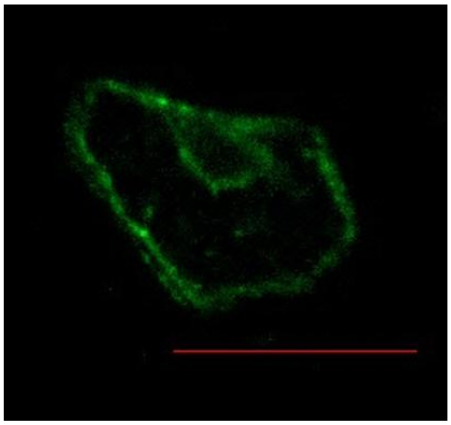
Analysis of signal peptides in *Populus*. The results from the onion epidermal cells indicated that all eight signal peptides were able to direct the extracellular secretion of EGFP. The highest ratio of secreted *versus* the whole cell EGFP was obtained with the fasciclin-like arabinogalactan protein 10 precursor (AGP) signal peptide. Therefore, we decided to test this signal peptide in *Populus* along with the poplar CβG signal peptide. Gene cassettes of the poplar CβG-EGFP and the AGP-EGFP were cloned into the binary vector pCAMBIA1300 and then transformed into *A. tumefaciens* strain AGL1. CβG-EGFP and AGP-EGFP were transformed with poplar leaves by vacuum-infiltration. As shown in Fig. 2-4, in transformed epidermal cells, signal peptide-driven EGFP were localized to the cell edge. Accumulation of green fluorescence in the ER network was also observed around the nucleus rather than inside the nucleus. This is, however, a characteristic for proteins designated for secretion to the apoplast via the ER-mediated secretory pathway (Brandizzi et al. 2003). In contrast, EGFP expression without a signal peptide showed strong florescence only in the nucleus and the cytoplasm. In vascular cells, signal peptide-driven EGFP was detected in the cell wall of cells neighboring the transformed cell. This is consistent with EGFP being transported out of the cell and moving to the adjacent cells (see Fig. 2-4).

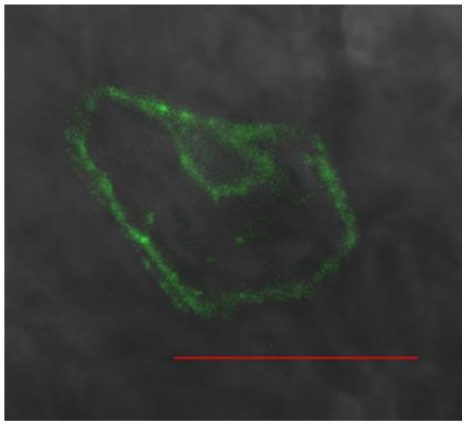
EGFP
(epidermal cell)



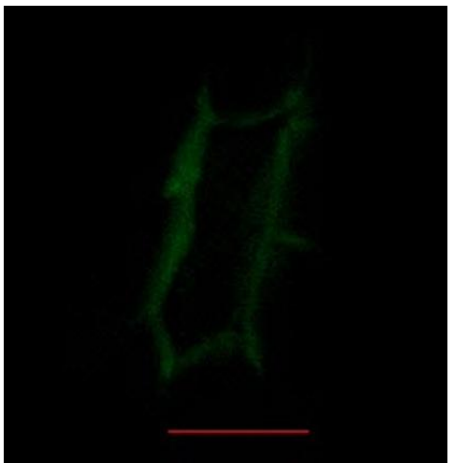


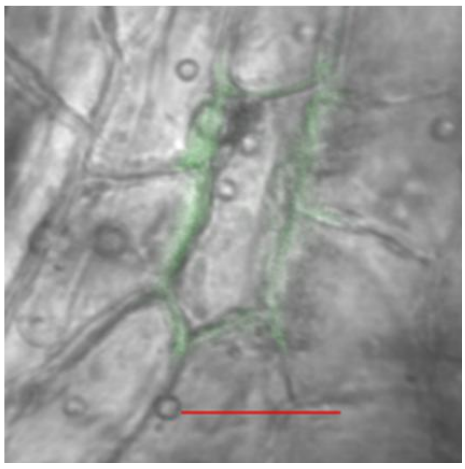
CβG::EGFP
(epidermal cell)





CβG::EGFP
(vascular cell)





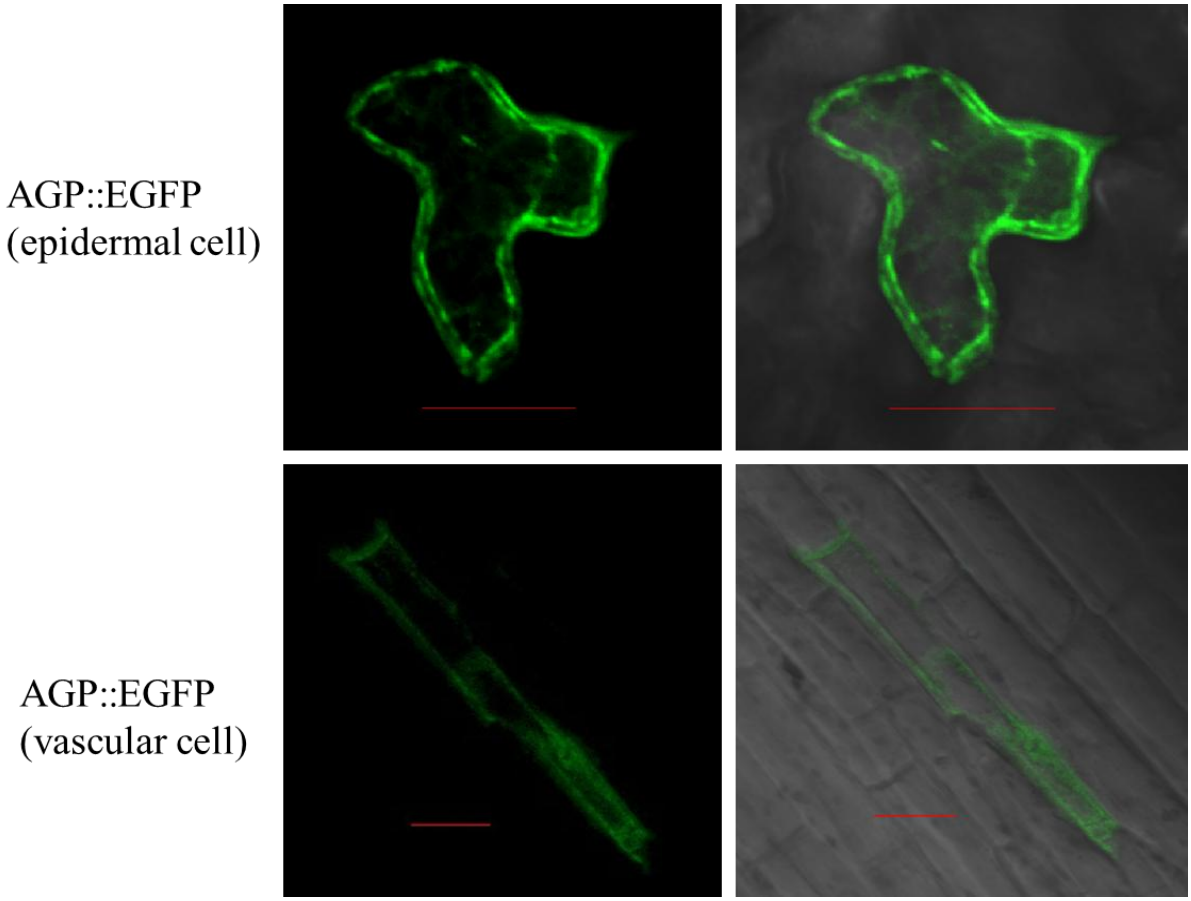


Fig. 2-4. EGFP localization in cells of hybrid poplar (*Populus deltoides*×*nigra*) leaves transformed with CβG:EGFP and AGP:EGFP observed by confocal microscope. Scale bar represents 20 μm.

DISCUSSION

The apoplast is comprised of the cell wall and extracellular space where many important physiological processes and structural assembling occur (Pechanova et al. 2010). Of importance in recent years due to interest in the use of cellulose as a biofuel, is the process of cell wall synthesis. Our interests (Liang et al. 2008), and those of others (Brunecky et al. 2011), in modifying cell wall synthesis, has led us to pursue the characterization of signal peptides that direct the secretion of proteins into the apoplast. Our interest has been on secreting proteins into the lignifying cell wall in order to cross link proteins with the growing lignin polymer (Liang et al. 2008).

Previous studies have shown that bioinformatics is not a good (perhaps even poor) predictor of cell-wall-associated proteins based on the presence of a signal peptide (Davis et al. 2009). For this reason, the present study utilizes a proteomic database (Dafoe and Constabel 2009) to select the candidate signal peptide to test. For verification, we performed subcellular localization studies with EGFP as the heterologously-expressed protein. Of the eight signal sequences we characterized, one is from the *Populus* CβG, first indentified in this study by blasting *Pinus* CβG against poplar genome. Our previous work with poplar involved secretion of a tyrosine-rich protein into the growing, lignifying cell wall (Liang et al. 2008). For this, we utilized the signal peptide of *Pinus* CβG. In *Pinus*, CβG was localized on the plant cell wall and found to catalyze the hydrolysis of coniferin to glucose and coniferyl alcohol (Dharmawardhana et al. 1995). However, the presence or localization of CβG in poplar is still unknown. In the present study, we identified a putative *Populus* CβG to use as one of the signal peptides. The other seven signal peptides

we obtained from the proteomic database of cell-wall-associated protein (Dafoe and Constabel 2009).

From the set of eight signal peptides, we then selected two, the CβG and the AGP to assess their ability to direct secretion into the apoplast in native plant, hybrid poplar. Our result shows the apoplastic localization of the *Populus* CβG signal peptide is consistent with this enzyme being involved in cell wall formation. The eight out of eight success rate is much higher than the success rate provided by bioinformatic prediction used by Davis et al. (Davis et al. 2009). We have no apparent reason for this, however, after the proteins are transported to extracellular space, their localization is also determined by their interactions with other cell wall components (Jamet et al. 2006). Within the extracellular apoplast region, proteins can be localized into three compartments (Jamet et al. 2006). When secretory proteins have little interaction with the cell wall, they are able to flow in the extracellular space freely (Jamet et al. 2006). A second group of proteins are weakly associated with cell wall components. This association can occur by Van der Waals interactions, hydrogen bonds, and hydrophobic or ionic interactions (Jamet et al. 2006). Most enzymes involving in cell wall formation and degradation, responding to biotic and abiotic stress, and plant signaling, exist in the apoplast in those two forms (Cassab 1998; Pennell 1998; Brownlee 2002). Finally, there is a group of apoplastic proteins that are covalently bonded to cell wall components (Brady et al. 1996). Such proteins are suggested to be cell wall structural proteins, integrated into cell wall structure during the process of cell wall formation (Jamet et al. 2006). High-resolution microscope and co-localization with other cell wall components are required to determine if the protein is cell wall structural protein (Keller et al. 1989).

In summary, our study has utilized a proteomic database to identify signal peptides that direct secretion into the apoplast. The success rate obtained here is much higher than that obtained from a previous bioinformatic study (Davis et al. 2009). The success of this study could allow large-scale and fast searching and evaluation of other putative apoplast-localized proteins.

REFERENCES

Boudart, G., E. Jamet, M. Rossignol, C. Lafitte, G. Borderies, A. Jauneau, M. T. Esquerre-Tugaye and R. Pont-Lezica (2005). Cell wall proteins in apoplastic fluids of Arabidopsis thaliana rosettes: Identification by mass spectrometry and bioinformatics. *Proteomics* **5**(1): 212-221.

Brady, J. D., I. H. Sadler and S. C. Fry (1996). Di-isodityrosine, a novel tetrameric derivative of tyrosine in plant cell wall proteins: A new potential cross-link. *Biochemical Journal* **315**: 323-327.

Brandizzi, F., S. Hanton, L. L. P. Dasilva, P. Boevink, D. Evans, K. Oparka, J. Denecke and C. Hawes (2003). ER quality control can lead to retrograde transport from the ER lumen to the cytosol and the nucleoplasm in plants. *Plant Journal* **34**(3): 269-281.

Brownlee, C. (2002). Role of the extracellular matrix in cell-cell signalling: paracrine paradigms. *Current Opinion in Plant Biology* **5**(5): 396-401.

Brunecky, R., M. J. Selig, T. B. Vinzant, M. E. Himmel, D. Lee, M. J. Blaylock and S. R. Decker (2011). In planta expression of A. cellulolyticus Cel5A endocellulase reduces cell wall recalcitrance in tobacco and maize. *Biotechnology for Biofuels* **4**: -.

Cassab, G. I. (1998). Plant cell wall proteins. *Annual Review of Plant Physiology and Plant Molecular Biology* **49**: 281-309.

Dafoe, N. J. and C. P. Constabel (2009). Proteomic analysis of hybrid poplar xylem sap. *Phytochemistry* **70**(7): 856-863.

Davis, E. J., D. Chen, D. B. Dewald, J. Shope, J. Macadam and Y. J. Wu (2009). Searching for new cell wall protein genes in Arabidopsis. *Plant Physiology and Biochemistry* **47**(2): 81-85.

Dharmawardhana, D. P., B. E. Ellis and J. E. Carlson (1995). A Beta-Glucosidase from Lodgepole Pine Xylem Specific for the Lignin Precursor Coniferin. *Plant Physiol* **107**(2): 331-339.

Dharmawardhana, D. P., B. E. Ellis and J. E. Carlson (1999). cDNA cloning and heterologous expression of coniferin beta-glucosidase. *Plant Molecular Biology* **40**(2): 365-372.

Edgar, R. C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *Bmc Bioinformatics* **5**: 1-19.

Guiltinan, M. J., S. Maximova, C. Miller, G. A. De Mayolo, S. Pishak and A. Young (2003). Stable transformation of Theobroma cacao L. and influence of matrix attachment regions on GFP expression. *Plant Cell Rep* **21**(9): 872-883.

Guiltinan, M. J., Z. Shi, S. N. Maximova, Y. Liu and J. Verica (2010). Functional analysis of the theobroma cacao NPR1 gene in arabidopsis. *Bmc Plant Biol* **10**.

Guo, Y., L. Zhang, L. H. Tian, J. F. Zhao, Y. Song and C. J. Zhang (2009). Identification of an Apoplastic Protein Involved in the Initial Phase of Salt Stress Response in Rice Root by Two-Dimensional Electrophoresis. *Plant Physiol* **149**(2): 916-928.

Humphrey, T. V., D. T. Bonetta and D. R. Goring (2007). Sentinels at the wall: cell wall receptors and sensors. *New Phytologist* **176**(1): 7-21.

Jamet, E., H. Canut, G. Boudart and R. F. Pont-Lezica (2006). Cell wall proteins: a new insight through proteomics. *Trends in Plant Science* **11**(1): 33-39.

Keller, B., M. D. Templeton and C. J. Lamb (1989). Specific Localization of a Plant-Cell Wall Glycine-Rich Protein in Protoxylem Cells of the Vascular System. *Proceedings of the National Academy of Sciences of the United States of America* **86**(5): 1529-1533.

Lease, K. A. and J. C. Walker (2006). The Arabidopsis unannotated secreted peptide database, a resource for plant peptidomics. *Plant Physiol* **142**(3): 831-838.

Liang, H. Y., C. J. Frost, X. P. Wei, N. R. Brown, J. E. Carlson and M. Tien (2008). Improved sugar release from lignocellulosic material by introducing a tyrosine-rich cell wall peptide gene in poplar. *Clean-Soil Air Water* **36**(8): 662-668.

Pechanova, O., C.-Y. Hsu, J. P. Adams, T. Pechan, L. Vandervelde, J. Drnevich, S. Jawdy, A. Adeli, J. C. Suttle, A. M. Lawrence, T. J. Tschaplinski, A. Séguin and C. Yuceer (2010). Apoplast proteome reveals that extracellular matrix contributes to multistress response in poplar. *BMC Genomics* **11**(1): 674.

Pennell, R. (1998). Cell walls: structures and signals. *Current Opinion in Plant Biology* **1**(6): 504-510.

Rubinstein, A. L., J. Marquez, M. Suarezcervera and P. A. Bedinger (1995). Extensin-like glycoproteins in the maize pollen tube wall. *Plant Cell* **7**(12): 2211-2225.

Vitale, A. and J. Denecke (1999). The endoplasmic reticulum - Gateway of the secretory pathway. *Plant Cell* **11**(4): 615-628.

Walter, P. and A. E. Johnson (1994). Signal Sequence Recognition and Protein Targeting to the Endoplasmic-Reticulum Membrane. *Annual Review of Cell Biology* **10**: 87-119.

Watson, B. S., Z. T. Lei, R. A. Dixon and L. W. Sumner (2004). Proteomics of Medicago sativa cell walls. *Phytochemistry* **65**(12): 1709-1720.

Zhang, Y., R. R. Sederoff and I. Allona (2000). Differential expression of genes encoding cell wall proteins in vascular tissues from vertical and bent loblolly pine trees. *Tree Physiology* **20**(7): 457-466.

CHAPTER 3

Characterization of cross linking between amino acids/proteins with lignin

INTRODUCTION

Lignocellulosic material is composed of three major constituents, cellulose, hemicellulose and lignin (Wyman 1994). Cellulose is the most abundant of the components, approximately accounting for 50 % of woody biomass(Bull 1991). The nature of β-1,4 linkages causes cellulose to be highly crystalline (Taylor 2008). Hemicellulose is more heterogeneous in structure than cellulose and is composed of a polymer of xylans (Scheller et al. 2010). Structurally, lignin is the most heterogeneous component of lignocellulose. It is the most abundant aromatic polymer in nature, accounting for approximately 30% of the organic carbon on earth (Boerjan et al. 2003).

In the plant cell wall, the non-covalent interactions are central toward cellulose crystallinity (Chundawat et al. 2011) and the interaction between cellulose microfibrils and hemicellulose (Scheller and Ulvskov 2010). To date, no evidence exists for covalent interactions between cellulose and hemicellulose. Perhaps the extensive intra-molecular interactions found in cellulose render the extensively hydrogen-bonded hydroxyl groups less reactive. The interactions between lignin and polysaccharides have been studied with "lignin-carbohydrate complex" (LCC) isolated from the plant cell walls (Merewether 1957; Morrison 1974; Morrison 1974; Neilson et al. 1982; Iiyama et al. 2006). The presence of covalent bonds between lignin and cellulose has been detected by NMR analysis of

carboxymethylated LCC (Iiyama et al. 2006). In addition, studies in *Lolium perenne* (Morrison 1974; Morrison 1974) and bovine rumen (Neilson and Richards 1982) have provided evidences of covalent linkages between hemicellulose and lignin. It has been suggested that the free hydroxyl groups of hemicellulose act as nucleophiles, attacking the quinone methide intermediate of lignin synthesis (Freudenberg 1959; Leary 1980; Helm 1998). The quinone methide is an intermediate formed from every coupling that involves reaction through a β-position (Freudenberg 1959). The electro-positive carbon of the quinone methide is susceptible to nucelophilic attack. When the nucleophile is water, the α-hydroxyl group is the resultant product.

Another essential component of the plant cell wall is protein. Proteins account for approximately 0.5% of the dry weight of the plant cell wall (Laidlaw et al. 1965; Martius 1992). Research has shown that a large number of different proteins are cell wall-associated (Cassab 1998). Some plant cell wall proteins may arise simply from the necessity of having proteins participate in plant cell wall synthesis. Other proteins may arise from the role of proteins in cell-cell interaction (Cheung et al. 1995; Rubinstein et al. 1995), morphogenesis (Bonilla et al. 1997), and plant differentiation (Knox et al. 1991; Kreuger et al. 1993). Once the cell wall is lignified, the ultimate of these proteins is not well known. However, studies have also shown that most of the cell wall-localized proteins are eventually cross-linked with the wall as structural components (Cassab 1998). That component may be lignin, researchers have shown that protein contaminations are always found along with lignin isolations from plant tissue (Jamet et al. 2006). Another study by Keller et al. showed that the antibody of glycine rich protein displayed a similar

staining pattern with lignified secondary cell wall, suggesting that glycine rich protein may cross link with lignin during cell wall formation (Keller et al. 1989).

The mechanism of this cross linking may be similar to that between hemicellulose and lignin and involves the quinone methide. In an *in vitro* study of interaction of peptide and synthetic lignin, researchers demonstrated that polylysine becomes embedded into the growing lignin polymer (McDougall et al. 1996). These researchers also showed Tyr augmentation of the polylysine enhances the aggregation of peptides into lignin (McDougall et al. 1996). However, the chemical nature of the interactions between protein and other cell wall components are still undefined.

In this study, we determined the reactivity of amino acids and lignin and showed the ability of three amino acids (Cys, Tyr, and Thr) to be covalently cross linked with lignin in coniferyl alcohol/peroxidase reaction mixtures. In addition, we demonstrated the capacity of proteins to cross link with lignin *in vitro* through Cys or through Tyr residues. Our findings provide a mechanism by which proteins and lignin can interact in the plant cell wall. Our results also suggest that this reaction can be exploited to modify cell wall structure for easier utilization of lignocellulosic biomass in industries.

MATERIALS AND METHODS

Chemicals. ^{14}C-cystine (>250 mCi/mmol), ^{3}H-tyrosine (40–60 Ci/mmol), ^{14}C-glycine (>100 mCi/mmol), and ^{3}H -sodium borohydride (5–15 Ci/mmol) were purchased from PerkinElmer (MA, USA). Horseradish peoxidase (HRP) Type II and tris(2-carboxyethyl) phosphine (TCEP) were obtained from Sigma-Aldrich. All other chemicals were purchased from Sigma-Aldrich and were reagent grade (USA).

Non-radioactive cross linking reaction mixtures. Coniferyl alcohol (CA), Tyr, Cys, Ser, and Thr were dissolved in 0.01 M sodium phosphate, pH 7.5. Hydrogen peroxide (H_2O_2) stock was prepared in 0.01 M sodium phosphate, pH 7.5, to the final concentration of 55 mM. Cross linking reactions were catalyzed by HRP. For reactions containing only CA, 10 ml reaction mixtures contained 2.75 mM CA and 1 mg HRP in 0.01 M sodium phosphate, pH 7.5. Then a total of ten aliquots of 500 μl of 55 mM H_2O_2 were added to this reaction mixture in intervals of 5 min. Reaction mixtures were incubated for an additional 4 h at room temperature with shaking. Reaction mixtures containing amino acids were identical except that they also contained 2.75 mM of the amino acid in the reaction mixture.

HPLC and LC-MS. Cross linking reaction mixtures described above were centrifuged at 13,000 ×g for 5 min. The supernatant was filtered by Amicon Ultracel (Millipore, Carrigtwohill, Co. Cork, Ireland) with a cut-off of 3,000 molecular weight (M.W.). The flow-through was collected for HPLC analysis (Shimadzu HPLC with UV-Vis detection). The reverse-phase column used in this study is XTerra RP18 with 5μm particle size, 125 Å pore size, and overall dimensions of 4.6 ×150 mm (Waters, USA). The

aqueous solvent, organic solvent, flow rate, and HPLC gradient are listed in Table 3-1. For LC/MS, we used a Shimadzu pump system, coupled with 3200 Q TRAP from AB Sciex Instruments. Samples were dissolved in HPLC solvent A, and subjected to the following gradient: 0 to 3 min, 2% solvent B; 3 to 5 min, 2% to 22%; 5 to 25 min, 22-75%; 25 to 27 min, 75-100%; 27 to 30 min, 100%. The flow rate was 0.9 ml/min. For MS analysis, the flow rate to the 3200 Q TRAP was 0.5 ml/min. Positive ionization was attained with electron spray ionization (ESI). Mass spectra were screened from 100 to 1000 *m/z* in the positive mode.

Solvent A: 100% water and 0.2% formic acid
Solvent B: 100% methanol and 0.2% formic acid
Flow rate: 1ml/min

Time (min)	A %	B%
0	100	0
10	90	10
30	10	90
40	0	100
45	0	100
47	100	0
50	100	0

Table 3-1. Gradient program for the reverse-phase HPLC analysis of reaction mixtures of reaction products of coniferyl alcohol, coniferyl alcohol/Tyr, coniferyl alcohol/Cys, and coniferyl alcohol/Thr

Transient state kinetics. Rate constants between peroxidase intermediates compound I and compound II and phenolic substrates were obtained as previously described according to Kuan et al. (Kuan et al. 1993). The reaction of compound I and compound II with the substrate was studied under pseudo-first order conditions and was monitored at 426 nm. The trace was fitted using the single-exponential decay model provided by KinTek to extract the overall observed electron transfer rate constant (k_{obs}). Rate constants were then obtained from plots of observed rates *versus* substrate concentrations. The slope of linear fit lines equals to the forward rate constant (k_1).

Radioactive cross linking reaction mixtures. ^{14}C-Cys was obtained by reduction of ^{14}C-cystine by TCEP at room temperature. Reaction solution containing the equal molar ratio of ^{14}C-cystine and TCEP was prepared by mixing 100 μl of ^{14}C-cystine (11 nmol, $3.8 \cdot 10^6$ cpm) in 2% ethanol and 100 μl of TCEP (11 nmol in 0.01 M sodium phosphate, pH 7.5).

^{3}H-CA was synthesized by reducing coniferyl aldehyde with ^{3}H-sodium borohydride, following the procedure (Ludley and Ralph 1996). Sodium borohydride (0.56mmol, $3.75 \cdot 10^8$ cpm) was added to 100 ml of coniferyl aldehyde (0.28 mmol) in 10 ml ethyl acetate. The solution was stirred for 1 h at room temperature and then poured into 50 ml water. The organic layer was removed and the aqueous layer was extracted three times with an equal volume of ethyl acetate. The organic layers were combined and dried over anhydrous $MgSO_4$. Removal of ethyl acetate by rotovap yielded a pale yellow solid. The final CA had a specific activity of $1.1 \cdot 10^6$ cpm/μmol.

Cross linking reaction mixtures (0.5 ml) contained 5.5 mM ^{3}H-CA and 55 nM of either ^{14}C-Cys or ^{14}C-Gly with 1 mg/ml HRP in 0.01 M sodium phosphate, pH 7.5. For reaction mixtures containing Tyr, cold CA was used with ^{3}H-Tyr under identical conditions. Reactions were initiated by adding ten 5 μl aliquots of H_2O_2 to give a final concentration of 5.5 mM.

Reaction mixture was centrifuged at 13,000 ×g for 10 min. After three times washing by water, the pellet was dissolved in dimethylformamide (DMF) and loaded to a column (1 ×40 cm) packed with Sephacryl S-100 (GEHealth, USA). The column was eluted with DMF containing 0.1 M lithium chloride. Fractions (1 ml) were collected for measuring radioactivity by liquid scintillation counter (LSC).

Expression and purification of a Cys-rich peptide. A glycine-rich cell wall structural protein (GRP) from maize (GenBank ACG24587) was selected for heterologous expression. The protein sequence except the portion of signal peptide predicted by SignalP (http://www.cbs.dtu.dk/services/SignalP/) was reversely translated to DNA sequence based on maximizing codon usage of *E. coli*. Protein sequence and DNA sequence are listed in Table 3-2. The DNA was synthesized by Integrated DNA Technologies, Inc. (IA, USA). The modified glycine-rich protein (MGRP) with increased Cys and Thr residues content was made with the following mutations in the GRP: E8C, D16T, D57T, E67C and E70C. The DNA sequences were received from IDT in carrying vectors. The desired gene was cloned into pET21a vector and transformed into *E. coli* BL21 (DE3) for expression. Both genes contained a his-tag at the C-terminus.

DNA sequence of GRP	GAG CAG GCT AAG GCT GCT AAG GAG GAG AAG AAG GTG GAG GTG CAG GAC TGG CAT GGT GGA GGA GGT TAC CCT GGT CAT GGT GGT GGT GGT GGA GGT GGT TAC CCT CAT TGC AGA TGG GGTTGC TGC AAC AGA GGTTAC CAT GGT GGT TGC AGA TGC TGC TCT CAT CCT GAC CAG ATC CCT GAG CCT ATG TAC AGA CCT GAG TTG GTG GAG GTG CAT AAC
Protein sequence of GRP	EQAKAAKEEKKVEVQDWHGGGGYPGHGGGGGG GYPHCRWGCCNRGYHGGCRCCSHPDQIPEPMYRP ELVEVHN
DNA sequence of MGRP	GAG CAG GCT AAG GCT GCT AAG TGC GAG AAG AAG GTG GAG GTG CAG ACC TGG CAT GGT GGA GGA GGT TAC CCT GGT CAT GGT GGT GGT GGT GGA GGT GGT TAC CCT CAT TGC AGA TGG GGT TGC TGC AAC AGA GGT TAC CAT GGT GGT TGC AGA TGC TGC TCT CAT CCT ACC CAG ATC CCT GAG CCT ATG TAC AGA CCT TGC TTG GTG TGC GTG CAT AAC
Protein sequence of MGRP	EQAKAAKCEKKVEVQTWHGGGGYPGHGGGGGG GYPHCRWGCCNRGYHGGCRCCSHPTQIPEPMYRP CLVCVHN

Table 3-2. DNA and protein sequences of GRP and MGRP. Modified nucleotides and amino acids of MGRP are shown in red.

Cells containing the pET21a plasmid were grown in Luria Broth with 50 μg/ml ampicillin. When the absorbance at 600 nm reached 0.6, isopropyl-β-D-thiogalactopyranoside (IPTG) was added to the final concentration of 1 mM and incubated for 16 h at 15 °C. Cells were collected by centrifugation at 6,000 ×g for 15 min. The pellet was suspended in bind buffer (20 mM Tris-Cl, pH 8.0, 100mM NaCl, and 10 mM imidazole), plus 20 mM β-mechaptoethanol (BME) and sonicated 5 times for 30 s each time. The disrupted cells were centrifuged at 13,000 ×g for 30 min at 4 °C. The supernatant was saved for purification, which followed the handbook of the QIAexpressionist (Qiagen, USA) with modifications. BME (20 mM) was added to all buffers used in the procedure. The desalt column was packed with Sephadex G-25 (Sigma, USA), and equilibrated with degassed 0.01 M sodium phosphate, pH 7.5 in an anaerobic chamber. Protein in elute buffer was loaded to the column for buffer exchange with 0.01M sodium phosphate, pH 7.5.

Protein cross-linking reactions. Reaction mixtures of 0.2 ml containing ^{3}H-coniferyl alcohol (1.8 mM) and either GRP (10 μM) or MGRP (10 μM) were cross linked as above with adding ten aliquots (2 μl) of H_2O_2 (18 mM). Reaction mixture was centrifuged at 13,000 ×g for 10 min. The supernatant was saved and incubated with Ni-NTA resin of 20 μl for 20 min. The mixture was centrifuged at 1,000 ×g for 20 s and the supernatant was discarded and then 100 μl of wash buffer (20 mM Tris-Cl, pH 8.0, 100mM NaCl, and 20 mM imidazole) was added and mixed well. The sample was centrifuged at 1,000 ×g for 20 s. The supernatant was saved for measuring radioactivity. The washing steps were repeated four times more, with 100 μl for each time. After washing, 100 μl of

elution buffer (20 mM Tris-Cl, pH 8.0, 100 mM NaCl, and 250 mM imidazole) was added and mixed well. The sample was centrifuged at 1,000 ×g for 20 s. The elution steps were repeated four times more, 100 μl for each time. The supernatant was used for subsequent characterization.

Blocking of Cys and Tyr residue with iodoacetamide. To samples of 50 μl of either GRP (50 μM) or MGRP (50 μM) in 0.01 M sodium phosphate, pH 7.5 was added 5 μl of 2% SDS and 45 μl of 200 mM ammonium bicarbonate, pH 8.0. To reduce the disulfide bonds, 5 μl of 200 mM TCEP was added and incubated at 55 ℃ for 1 h. Then 5 μl of freshly prepared 375 mM iodoacetamide in 200 mM ammonium bicarbonate, pH 8.0 was added and incubated for 30 min. The sample was protected from light during reaction. After reaction, treated protein was desalted on a Sephadex G-25 column in 10 mM sodium phosphate, pH 7.5.

Autoradiogram of protein cross-linking reactions. ^{3}H-CA/GRP cross linked protein described above was concentrated to 20 μl in a vacuum Speedvac. The sample was then subjected to SDS-PAGE (17% acrylamide) (Laemmli 1970). GRP (0.5 mg) without the incubation with ^{3}H-CA was loaded to SDS-PAGE (17% acrylamide) as a control. After running, the gel was fixed with 5% glacial acetic acid, 5% isopropanol, and 90% water for 20 min. After the gel was rinsed in a continuous flow of tap water for 15 min, it was soaked in Autofluor (National Diagnostics, USA) for 30 min with shaking. The gel was directly placed on three layers of filter paper and dried on gel dryer, under 80 °C and vacuum. After drying, the gel with filter paper was placed on film and expose at -80 °C for 11 days before development.

RESULTS

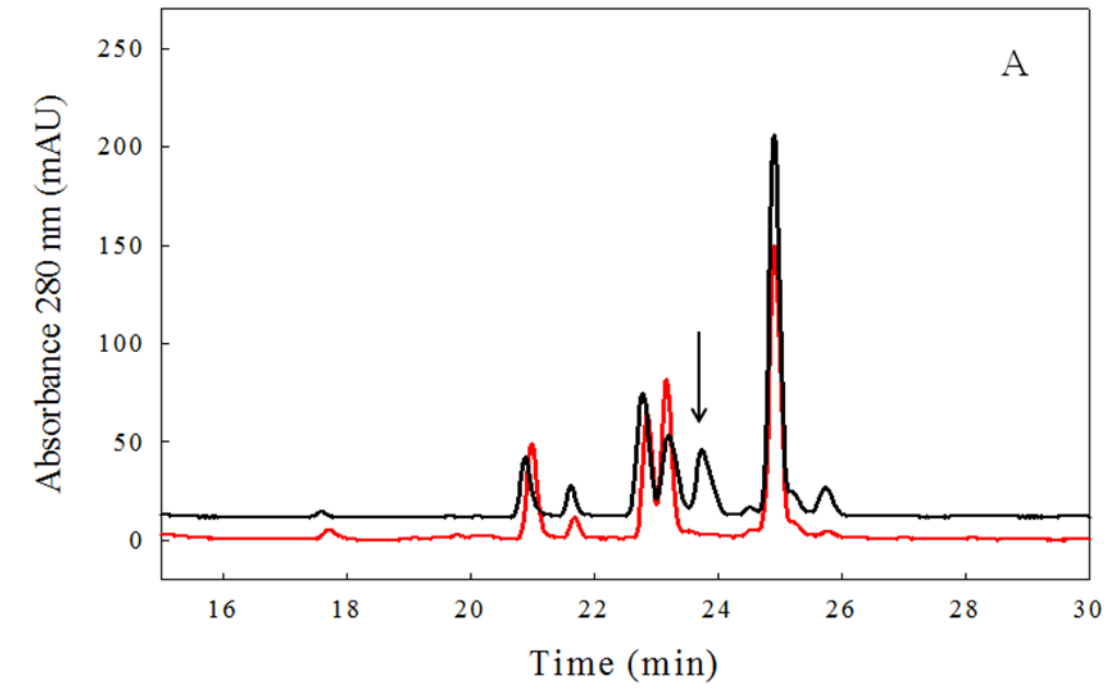
Amino acid/coniferyl alcohol cross linking. We previously generated transgenic plants where proteins containing a high Tyr content were secreted into the lignifying plant cell wall (Liang et al. 2008). A protein with a high Tyr content was chosen because we reasoned that since Tyr contains a phenolic group like that of the lignin precursors, it could potentially be linked into lignin by a free radical coupling process. In the present study, we aim at determining the mechanism of Tyr cross linking with lignin. As such, Tyr was added to reaction mixtures previously developed for synthesis of DHP lignin (Cathala et al. 1998). These incubations contained coniferyl alcohol and horseradish peroxidase. Our initial experiments sought to identify the low molecular reaction products because we reasoned their structure would be easier to determine by mass spectroscopy (MS) or by nuclear magnetic resonance (NMR). The reaction mixtures were analyzed by reverse phase HPLC. HPLC chromatograms of coniferyl alcohol-containing reaction mixtures with and without added Tyr are shown in Fig. 3-1A. Products unique to the incubation containing Tyr (arrow in Fig. 3-1A) was purified and then analyzed by LC/MS. From the Tyr-containing reaction mixture, the unique peak had a mass to charge ratio (*m/z*) of the major ion as 540.3. This corresponds to the mass of a product containing two coniferyl alcohols and one Tyr (Fig. 3-2A). The MS spectra of the new peak from CA/Tyr reaction showed mass losses of 181 Da and 211 Da, representing the $[\underline{M+H^+} - C_9H_{11}NO_3(Tyr)]^+$ and $[\underline{M+H^+} - C_9H_{11}NO_3(Tyr) - CH_2O]^+$. According to the work by Morreel et al. (Boerjan et al. 2010; Morreel et al. 2010) about dissolving the structure of lignin oligomers by mass spectrometry-based fragmentation, we proposed the new product's structure as shown in Fig. 3-2A. In Morreel's work, the MS fragmentation of coniferyl alcohol dimer linked with

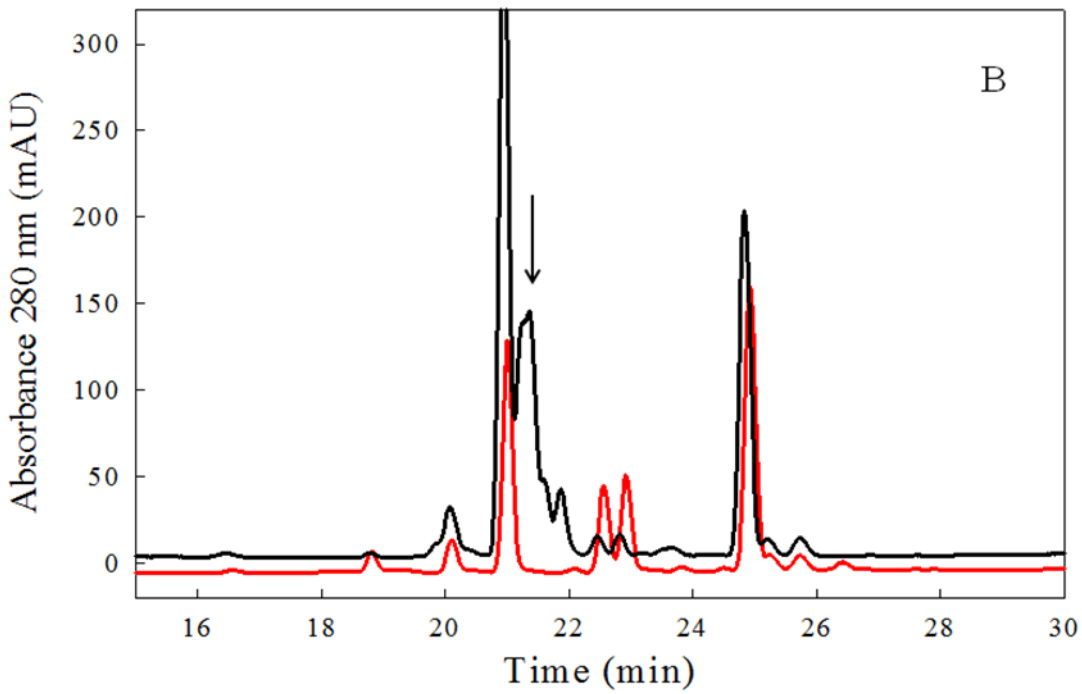
β-O-4 linkages showed the loss of water and the combined loss of water and formaldehyde. When two monolignol radicals couple to form a lignin dimer, the coupling intermediate will go through the rearomatization to form a stable dimer. β-O-4 linked dimer involves an addition of external water at α-position for rearomatization. They believed that during the MS process, the hydroxyl group at α-position is expelled when quinone methide structure is formed again, resulting in the loss of water. After the elimination of water, 8-9 bond is continuously fragmented subjected to 1, 2-elimination mechanism, leading the release of formaldehyde. Similarly, in our case, the amino acid added at α-position is expelled when quinone methide structure is formed again, yielding the *m/z* peak 359.2. After the elimination of amino acid, 8-9 bond is continuously fragmented subjected to 1, 2-elimination mechanism, leading the release of CH_2O. The structure of the adduct is also shown in Fig. 3-2A. MS data of new product indicates that amino acid as an external nucleophile attacks the α position of quinone methide of β-O-4 linked coniferyl alcohol dimer to form the amino acid-lignin crosslinking product.

Quinone methides are unstable intermediates generated during lignin dehydrogenative polymerization (Leary 1980). With the realization that nucleophilic attack of the quinone methide dominated the reaction mechanism of Tyr with the lignin intermediates, we then examined other amino acids with hydroxyl or thiol group. As such, Ser, Thr, and Cys were examined for their cross linking ability with lignin. Reaction mixtures containing the above amino acids were added to coniferyl alcohol/ horseradish peroxidase reaction mixtures. All reactions were initiated by addition of H_2O_2. In comparison to coniferyl alcohol only reaction mixtures, the HPLC chromatograms showed new peak(s) in reaction mixtures containing Cys and Thr (Fig. 3-1B and Fig. 3-1C).

However, the chromatogram of the reaction mixture containing Ser did not show any new peaks.

As for the Tyr-containing adducts, reaction products unique to the incubation containing the amino acid, compared to the reaction mixture containing only coniferyl alcohol, were isolated and analyzed by LC/MS. For the new peaks from the Cys-containing reaction mixture, an m/z of 480 was obtained which corresponds to the trimer where the Cys reacts with the quinone methide. The m/z peak of 350.2 corresponds to loss of Cys (Fig. 3-2B). Fig. 3-2C shows similar results obtained with Thr, again showing the Thr adduct with the quinone methide and then loss of Thr.


A
Absorbance 280 nm (mAU)
250
200
150
100
50
0
16
18
20
22
24
26
28
30
Time (min)


B
Absorbance 280 nm (mAU)
300
250
200
150
100
50
0
16
18
20
22
24
26
28
30
Time (min)

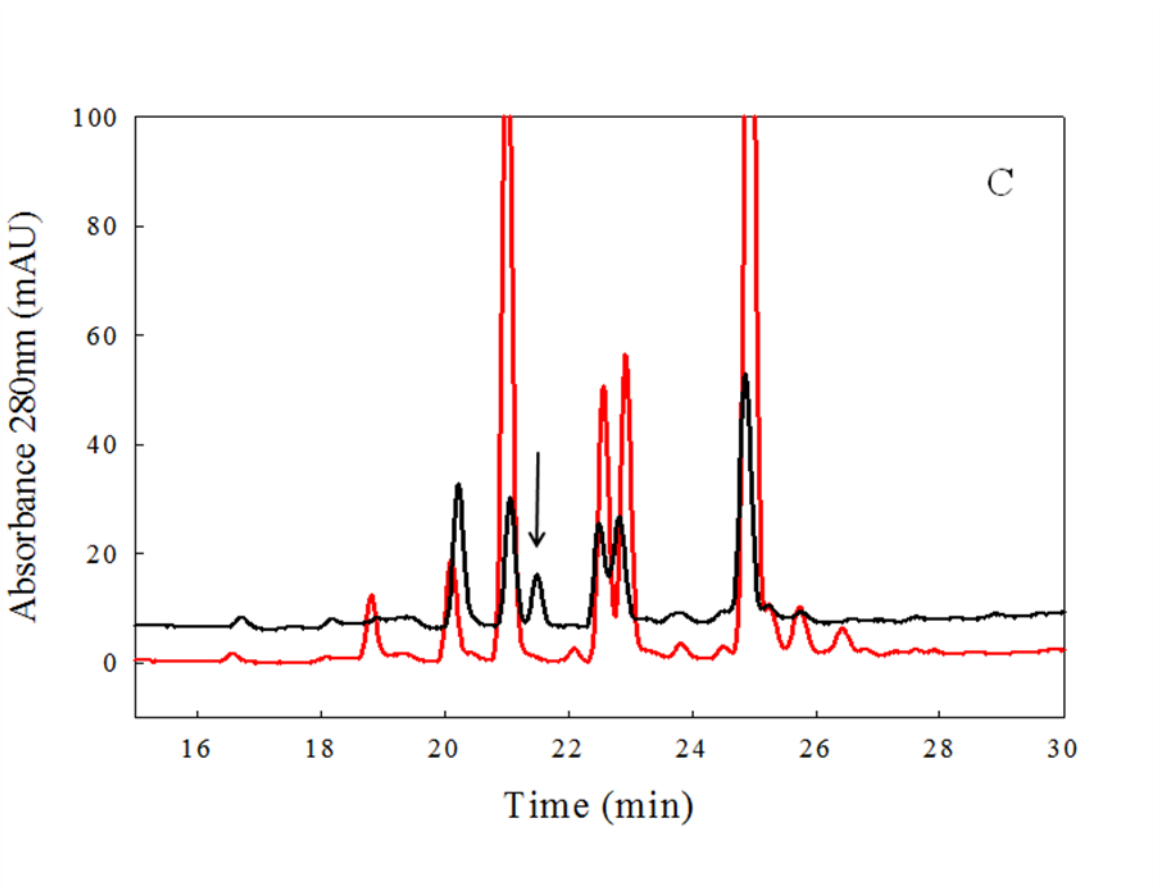
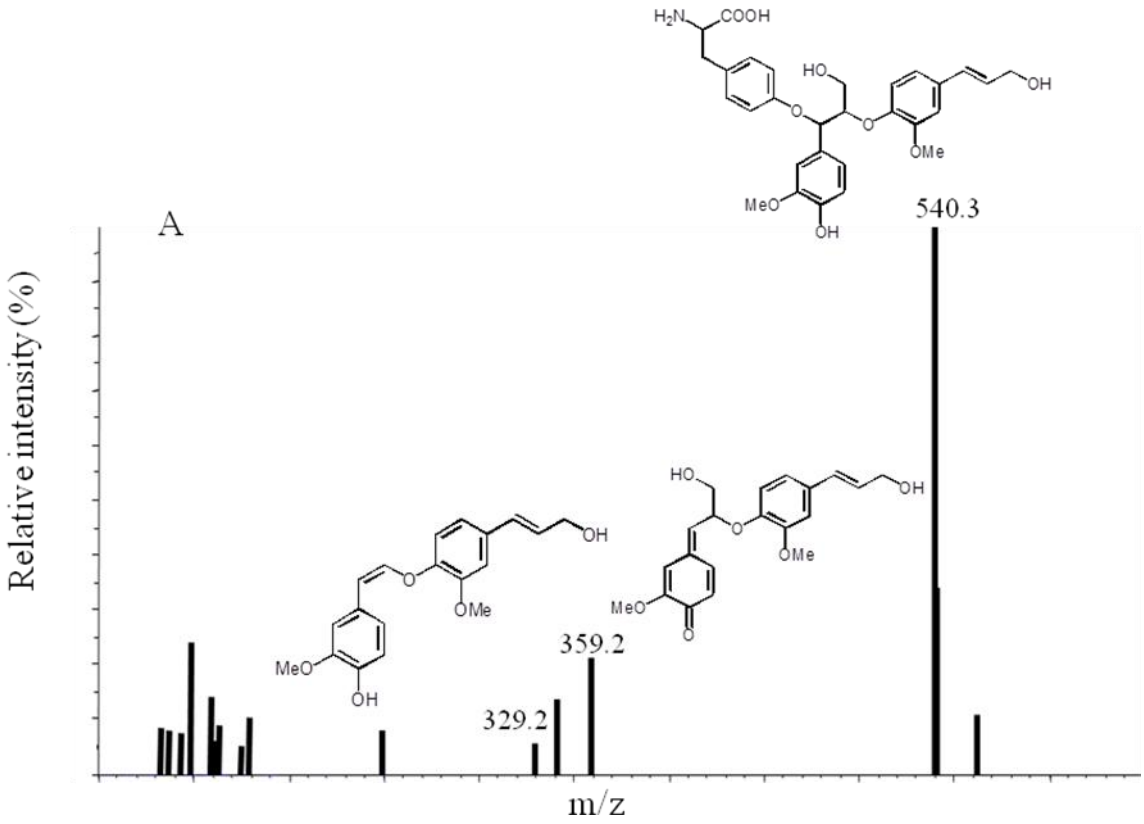
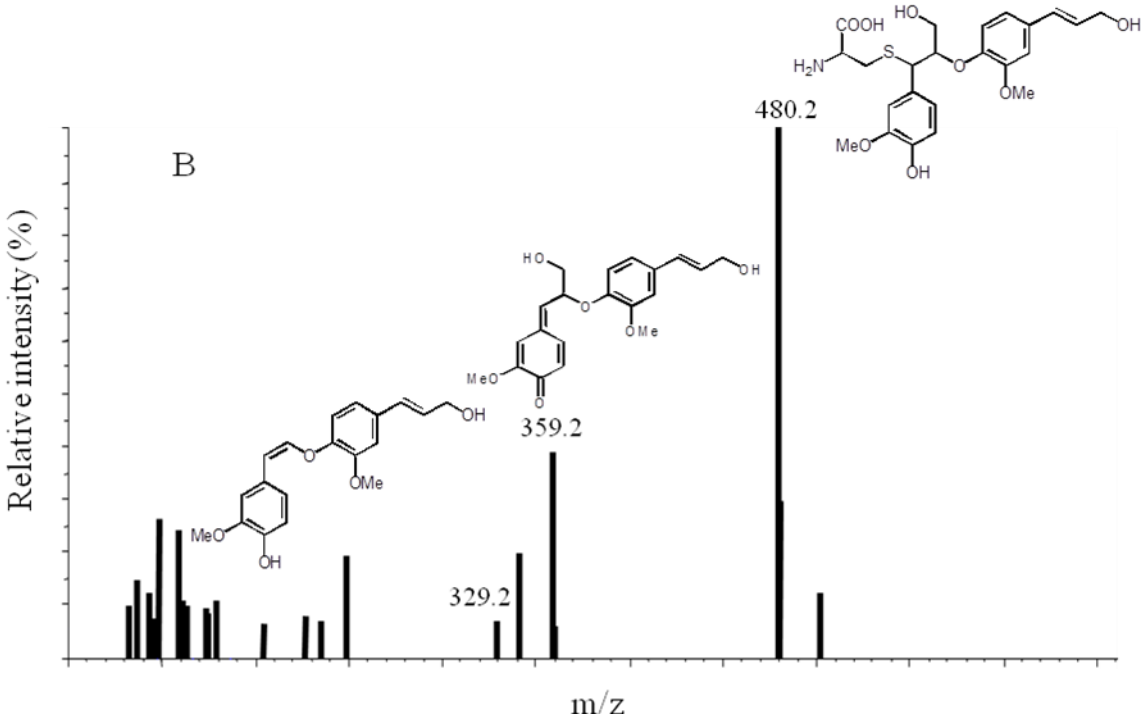


Figure 3-1. HPLC chromatogram of the reaction products of coniferyl alcohol/Tyr (A), coniferyl alcohol/Cys (B), and coniferyl alcohol/Thr (C), shown in black. Each figure also contains the reaction mixture of coniferyl alcohol without amino acid (in red). After incubation as described in Materials and Methods, the reaction mixture was filtered with a membrane with 3,000 M.W. cut-off. The fractions containing the unique reaction product, as indicated by the arrow were collected for LC/MS analysis.


A
Relative intensity (%)
m/z
540.3
359.2
329.2
H2N
COOH
HO
OH
OMe
MeO
OH
O


B
Relative intensity (%)
m/z
480.2
359.2
329.2
COOH
H2N
S
HO
OH
OMe
MeO
OH
O

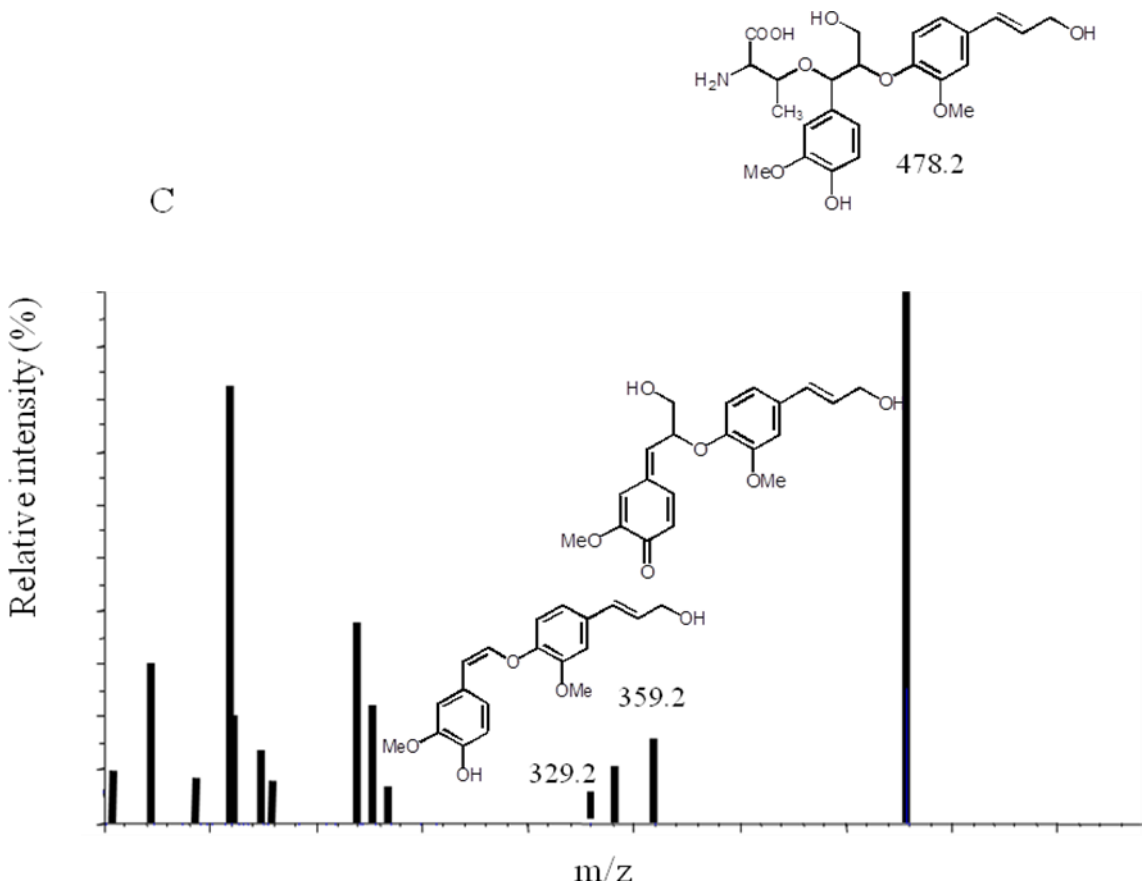


Figure 3-2. Mass spectrum of products isolated from amino acid/coniferyl alcohol cross linking reactions: coniferyl alcohol/Tyr (A), coniferyl alcohol/Cys (B), and coniferyl alcohol/Thr (C). Proposed structures of new products formed from the cross linking reactions and fragmented products are shown on top of the corresponding peaks.

Lack of free radical coupling between Tyr and coniferyl alcohol. We used kinetic studies to determine the chemical rationale behind the lack of free radical coupling between Tyr and coniferyl alcohol. The kinetic studies utilized a stop flow to determine the reactivity of Tyr *versus* lignin or lignin-like precursors. To form a mixed dimer of coniferyl alcohol and Tyr requires formation of a radical on both phenols. Cross linking between the two may be in low yield if one phenol is much more reactive than the other. For example, if coniferyl alcohol is more reactive than Tyr, then mostly coniferyl alcohol self-cross-linked products would be detected. To assess cross linking reactivity, we used horseradish peroxidase and determined the reactivity of selected phenols with the oxidized peroxidase intermediates compound I and compound II. Fig. 3-3 shows the mechanism of oxidation of the enzyme by H_2O_2 to form a two-electron-oxidized intermediated (referred to as compound I) which oxidizes the first phenol molecule by one electron (to form the phenolic radical) and the one-electron-oxidized enzyme intermediate (referred to as compound II). Compound II then oxidizes the second phenolic substrate yielding a second radical product and the resting enzyme. Kinetic studies with the stop flow showed indeed that coniferyl alcohol is exceedingly more reactive than Tyr with compound I and II (Table 3-1). For compound II, the rate with coniferyl alcohol is over 600 times more reactive. Rates with compound I with coniferyl alcohol are too fast to measure by stop flow but are estimated to be at least 100 times faster than with coniferyl alcohol. As such, direct cross linking between the coniferyl alcohol radical and the tyrosyl radical appears unlikely. The likelihood of lignin dimers or oligomers cross linking with Tyr was investigated by examining the reactivity with vanillyl alcohol. Vanillyl alcohol would resemble the monomeric unit of a β-O-4 dimer. Perhaps more importantly, it does not contain the α–β

double bond of coniferyl alcohol. As shown in Table 3-3, the rate with vanillyl alcohol is only 20 times faster than Tyr methyl ester. We thus reason that Tyr cross linking with lignin would only occur in the later stages of lignin formation where the phenols are less reactive than coniferyl alcohol and closer to the reactivity of Tyr.

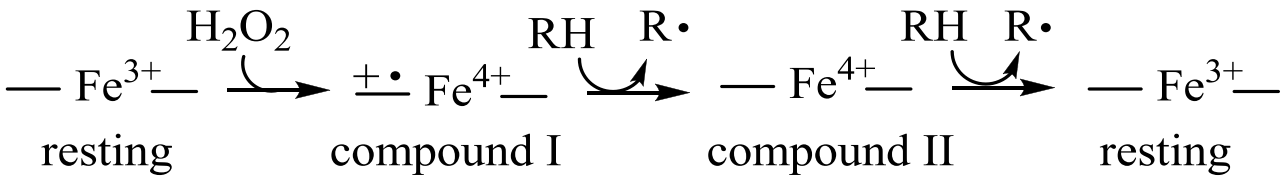


Figure 3-3. Mechanism of oxidation of horseradish peroxidase by H_2O_2 to form a two-electron-oxidized intermediated (referred to as compound I) which oxidizes the first phenol molecule by one electron (to form the phenolic radical) and the one-electron-oxidized enzyme intermediate (referred to as compound II).

Substrate	Compound I to II	Compound II to ferric
Coniferyl alcohol	$> 10^{8}\ M^{-1}\ s^{-1}$	$1.0 \times 10^{7}\ M^{-1}\ s^{-1}$
Tyr methyl ester	$2.4 \times 10^{6}\ M^{-1}\ s^{-1}$	$1.6 \times 10^{4}\ M^{-1}\ s^{-1}$
Vanillyl alcohol	$> 10^{8}\ M^{-1}\ s^{-1}$	$3.2 \times 10^{5}\ M^{-1}\ s^{-1}$

Table 3-3. Rate constants of horseradish peroxidase compound I to compound II and compound II to the resting enzyme with coniferyl alcohol, tyrosine methyl ester, and vanillyl alcohol.

Crosslinking of Cys and Tyr into the DHP polymer. We then aimed at determining whether Tyr, Cys and Thr can be incorporated into the higher polymerized lignin structure. DHP was synthesized as described previously using ^{3}H-coniferyl alcohol (synthesized from sodium borohydride reduction of coniferyl aldehyde) (Lu et al. 1998) in the presence of ^{14}C-Cys. Also, non-radioactive DHP was synthesized in the presence of ^{3}H-Tyr. ^{14}C-Thr is not commercially available and thus was not further characterized. The ability of Gly to cross link with the synthesizing DHP was used as a (negative) control. Incorporation of the amino acid into DHP was assessed by gel permeation chromatography on Sephacryl S-100 with 0.1 M lithium chloride in DMF as the eluent (Lewis et al. 1987). After the polymerization reaction, DHPs-containing pellet was washed and dissolved in DMF. Radioactivities in eluent were determined by scintillation counting (Fig. 3-4) and the molarity of coniferyl alcohol and amino acids in eluent was calculated with their specific activities (Fig. 3-5). As shown in Fig. 3-4A, horseradish peroxidase-catalyzed polymerization of coniferyl alcohol yielded DHP with a broad range of molecular weight. Fig. 3-4 also shows the polystyrene standard molecular weight markers used to calibrate the column. A large proportion of the radioactivity is detected above 2,000 Daltons, which corresponds to DHPs containing at least 10 coniferyl alcohols. For the Cys-containing reaction mixtures, ^{14}C was co-eluted with the ^{3}H-DHP thus showing incorporation of Cys (1 Eq) into the DHP polymer of coniferyl alcohol (4,000 Eq), as shown in Fig. 3-4B and Fig. 3-5B. For the Tyr-containing reaction mixtures, ^{3}H was eluted from 17 ml to 32 ml where DHP was also eluted, thus showing incorporation of Tyr (1 Eq) into the DHP polymer of coniferyl alcohol (2,700 Eq) (see Fig. 3-4C and Fig. 3-5C), consistent with

covalent interaction between Tyr and the DHP. No ^{14}C co-eluted with the DHP when ^{14}C-Gly was added to the reaction mixture.

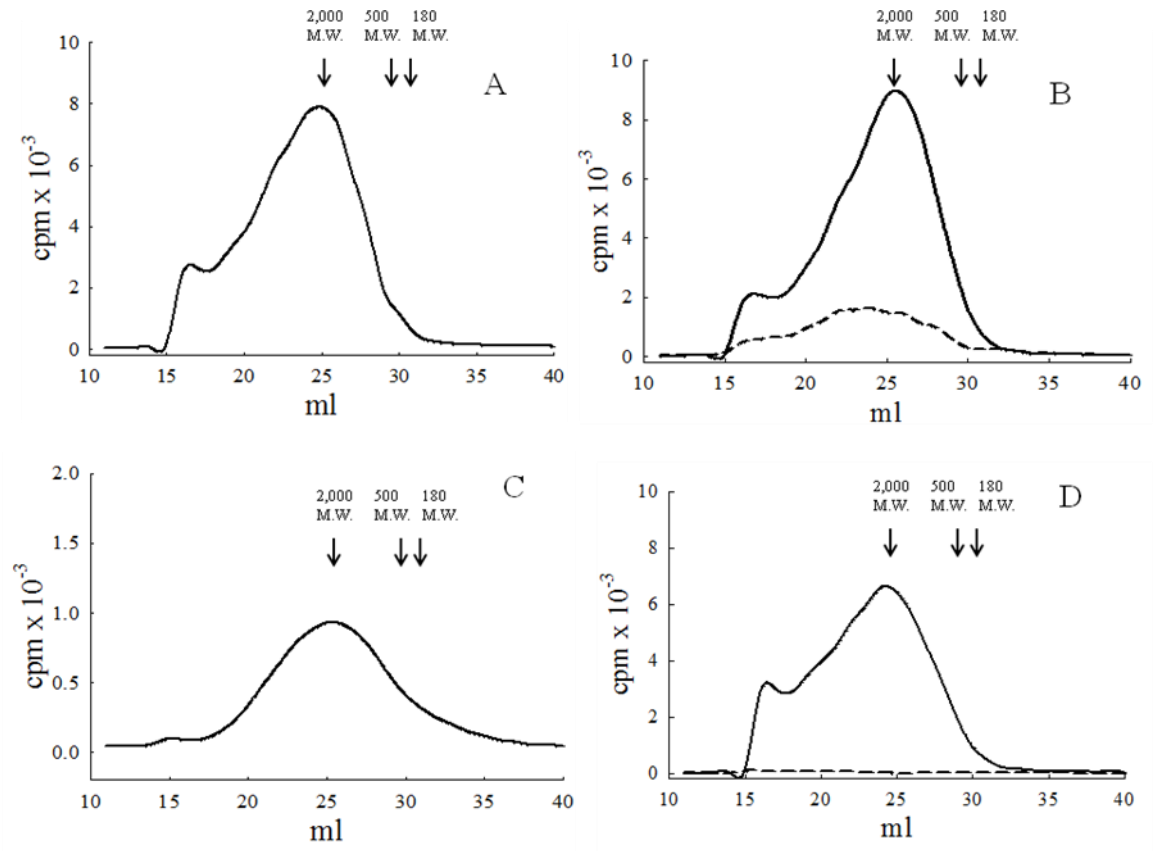


Fig. 3-4. Radioactivity in eluent from Sephacryl S-100 column of cross linking reaction mixtures. Panel A: incubation with only ^{3}H-coniferyl alcohol (—). Panel B: incubation with ^{3}H-coniferyl alcohol(—) and ^{14}C-Cys (----). Panel C: incubation with ^{3}H-Tyr(—) and non-radioactive coniferyl alcohol. Panel D: incubation with ^{3}H-coniferyl alcohol (—) and ^{14}C-Gly (----).

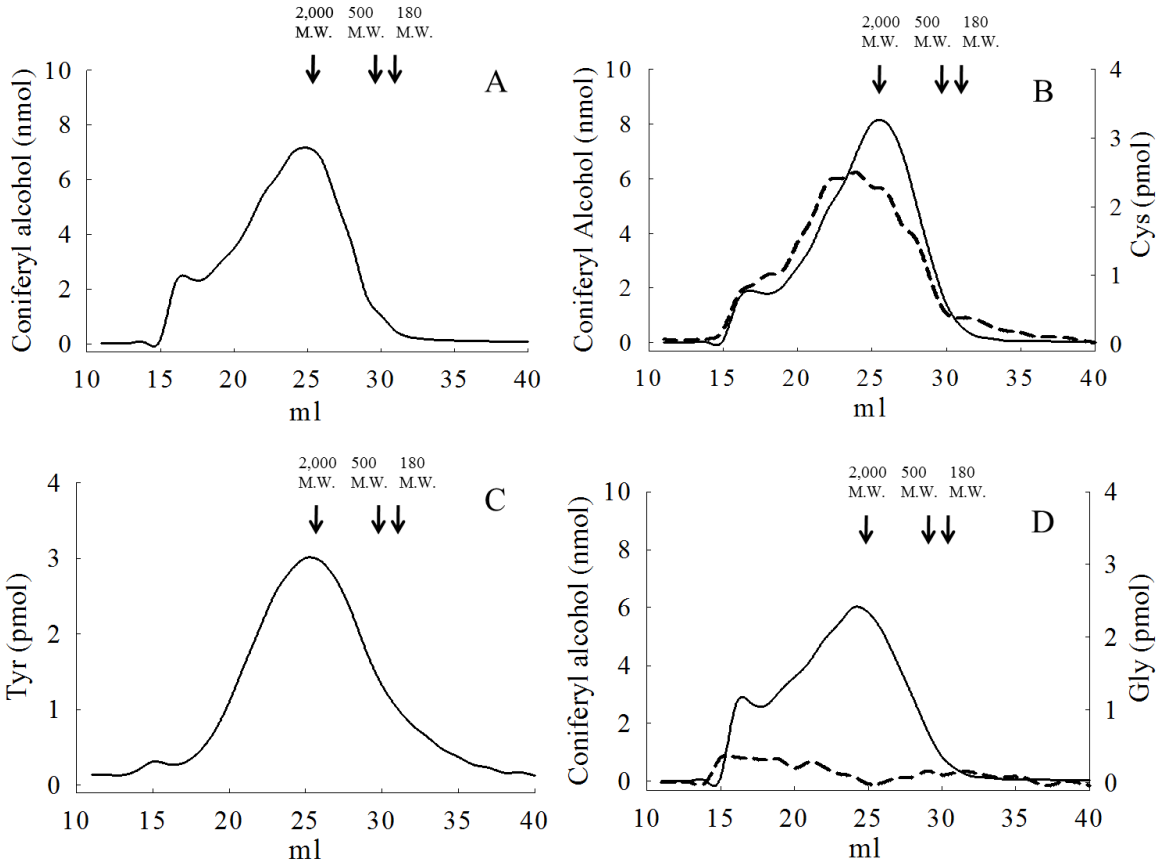


Fig. 3-5. Molarity of coniferyl alcohol and amino acids in eluent from Sephacryl S-100 column of cross linking reaction mixtures. Panel A: incubation with only ^{3}H-coniferyl alcohol (—). Panel B: incubation with ^{3}H-coniferyl alcohol (—) and ^{14}C-Cys (----). Panel C: incubation with ^{3}H-Tyr (—) and non-radioactive coniferyl alcohol. Panel D: incubation with ^{3}H-coniferyl alcohol (—) and ^{14}C-Gly (----).

Evidence for cross linking of lignin with proteins. Having established that free amino acids are covalently linked into DHP, we then determined whether whole proteins can also participate in the cross linking reaction. We used the glycine-rich protein (GRP) from maize (GenBankACG24587). We also mutated the glycine-rich protein to increase the Cys and Thr content from 7.3% Cys and none Thr, to 11% Cys and 3% Thr, referred as MGRP. Both GRP and MGRP were heterologously expressed in *E. coli* with a His-tag. The purified protein of GRP or MGRP was mixed with ^{3}H-coniferyl alcohol in the horseradish peroxidase reaction mixture for cross-linking. After a four-hour incubation, Ni-NTA agarose was added to separate the protein from the unreacted coniferyl alcohol and the agarose pellet was then washed with wash buffer (20 mM Tris-HCl, pH 8.0, 100mM NaCl, and 20 mM imidazole). When no coniferyl alcohol could be detected in the wash solution as monitored by scintillation counting, elution buffer (20 mM Tris-HCl, pH 8.0, 100 mM NaCl, and 250 mM imidazole) was added to the agarose beads. In the first three rounds of washing, radioactivity eluted was due to the free ^{3}H-coniferyl alcohol. In the forth and fifth washing, radioactivity was not detected. The radioactivity was detected upon addition of the elution buffer (Fig. 3-6). A total of 849 cpm (0.77 nmol coniferyl alcohol) was eluted with the GRP incubation. The results are consistent with coniferyl alcohol cross-linking with the proteins of interest. Higher radioactivity (1,719 cpm, 1.56 nmol coniferyl alcohol) was detected in the elution of MGRP which contains more Cys and Thr than the glycine-rich protein. The reaction incubation without the purified protein was used as a negative control.

To confirm that cross-linking between the protein and coniferyl alcohol was via Cys and Tyr residues, iodoacetamide was used to block these two functional groups.

Iodoacetamide binds covalently to the thiol group of Cys and the hydroxyl group of Tyr (Smythe 1936; Cotner et al. 1973). The results with the treated proteins showed no radioactivity in cross linking reaction mixtures containing the radioactive amino acids with either the glycine-rich protein or the modified glycine-rich protein.

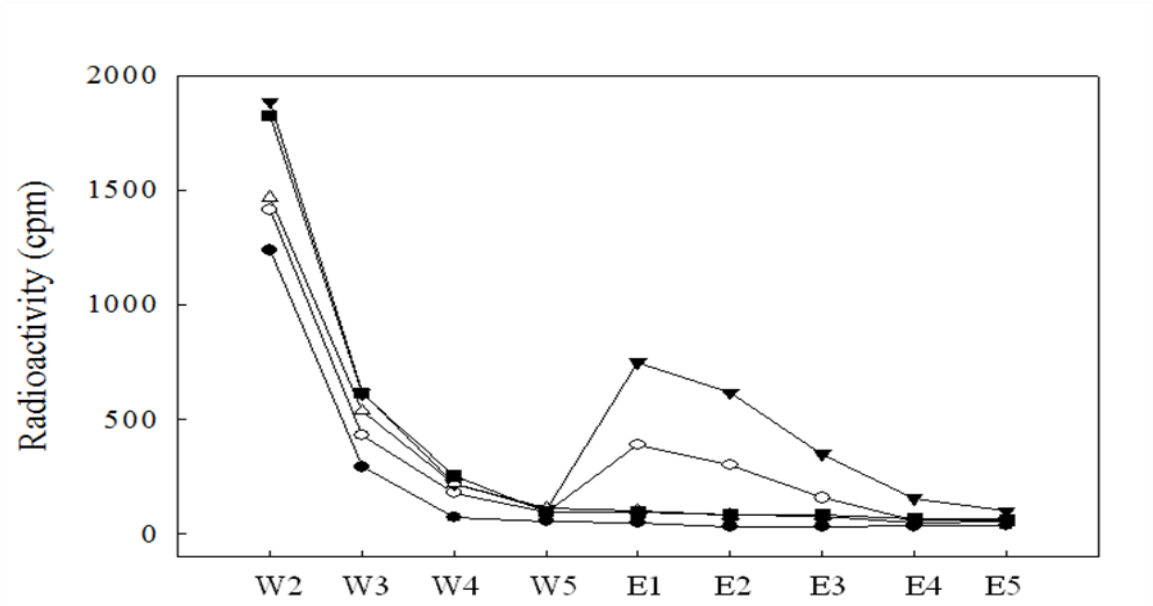


Fig. 3-6. Cross linking of coniferyl alcohol with either GRP or MGRP. His-tag containing GRP or MGRP were incubated with ^{3}H-coniferyl alcohol, HRP and H_2O_2 as described in Materials and Methods. The reaction mixture was then added into a Ni-NTA agarose column and washed with 5 aliquots (100 μl each) of wash buffer and then with 4 aliquots of elution buffer. Coniferyl alcohol-derived counts were then measured by scintillation spectrometry from the following reaction mixtures: (—o—) GRP/coniferyl alcohol; (—▼—) MGRP/coniferyl alcohol; (—Δ—) GRP treated with iodoacetamide/coniferyl alcohol; (—■—) MGRP treated with iodoacetamide/coniferyl alcohol; (—●—) coniferyl alcohol alone with no added amino acids.

Autoradiogram of SDS-PAGE of cross-linking products of protein and coniferyl alcohol. To visualize the cross-linking products of GRP and coniferyl alcohol, we loaded the reaction products in Elution 1 on SDS-PAGE and obtained the autoradiogram image of that (Lane 1 in Fig. 3-7). A band was clearly detected on the film, indicating that after cross-linking reaction, eluted GRP was cross-linked with ^{3}H-coniferyl alcohol. Furthermore, compared with unreacted GRP protein (shown in Lane 2 in Fig. 3-7), the size of reacted GRP enlarged by 7 kD, due to the addition of lignin oligomers.

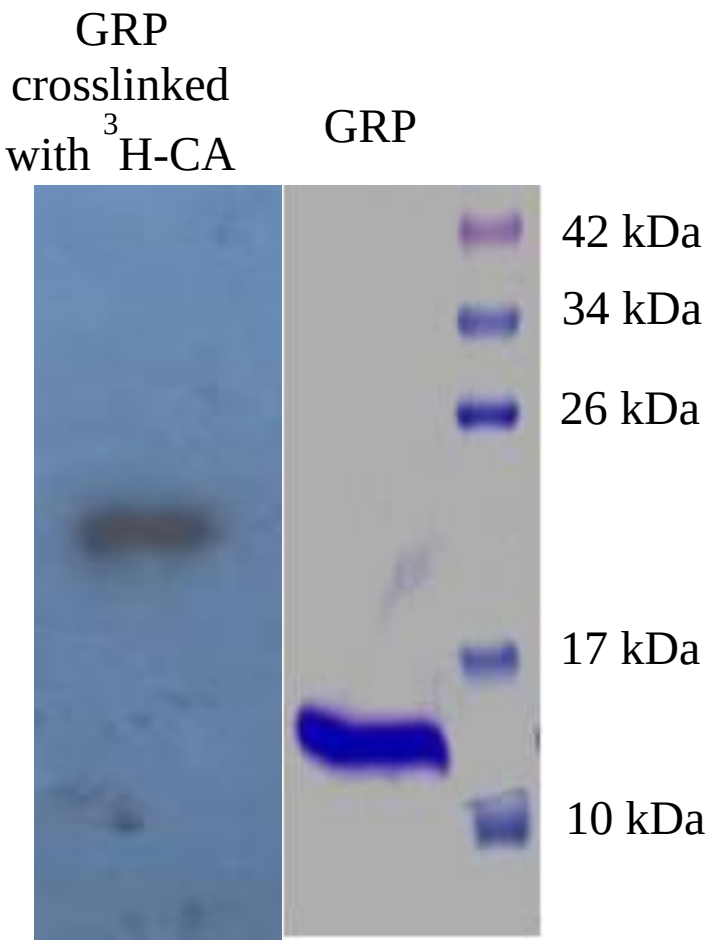


Figure 3-7. Cross linking of GRP with coniferyl alcohol as determined by SDS-PAGE and autoradiography. Lane 1 is Coomassie blue-stained SDS-PAGE of GRP (no incubation). Lane 2 is autoradiogram of SDS-PAGE of GRP incubated with ^{3}H-coniferyl alcohol as described in Material and Methods. Molecular weights markers are shown to the right of the gel.

DISCUSSION

Our work here is not the first to suggest interactions between proteins and lignin. Whitmore (Whitmore 1978; Whitmore 1982) provided evidence for cell wall proteins binding to lignin precursors during the lignification process. In the first study by Whitmore in 1978, data indicated that proteins rich in hydroxyproline were cross linked with lignin in incubations. They added the protein in the DHP-synthesis reaction system containing ^{14}C-coniferyl alcohol, HRP, and H_2O_2. They found that the protein bound to the DHP was resistant to acid hydrolysis. Comparison of the radioactivity per gram of DHP before and after acid hydrolysis showed no loss of the protein. They proposed that this was consistent with covalent crosslinking between the protein and lignin.

A subsequent study by Whitmore in 1982 (Whitmore 1982) involved isolating cell wall tissue and adding this tissue to DHP synthesis reaction mixtures. The results showed that the proteins embedding in the cell wall tissue from the *Pinus* callus are able to crosslink with coniferyl alcohol in incubations containing H_2O_2 and HRP. They measured the relative proportion of the cell wall proteins *versus* lignin, and found the proportion did not decrease by incorporating new coniferyl alcohol in the cell wall tissue. This could only be explained by formation of new linkages between the newly synthesized lignin and proteins. They also labeled cell wall protein by incubating *Pinus* callus with ^{14}C-proline, and then radioactivity was detected in the lignin fraction, which also indicates proteins are able to bind to lignin polymer in the cell wall. While these previous studies are consistent with the cross linking of proteins to lignin, they provide little or no information on the

nature of the interactions. The present paper characterizes the nature of the protein to lignin linkages and the reactivity of the amino acids that can participate in the interaction.

We previously worked on cross linking Tyr-containing proteins with lignin (Liang et al. 2008) with the rationale that Tyr contains the same reactive phenolic group as that of the lignin precursors. Transgenic poplars were generated where a protein with high Tyr content was secreted into the growing, lignifying cell wall (Liang et al. 2008). The resultant lignocellulosic material from these transgenic plants exhibited increased digestibility with no observable compromise in fitness (Liang et al. 2008). The present work focuses on the mechanism of how the amino acid is able to cross link with lignin. We first aimed at identifying the nature of the cross link between Tyr and lignin using an *in vitro* reaction system. We fully expected cross linking to occur through the phenol ring of Tyr, similar to that found in diTyr (Malencik et al. 1996). The reaction products from the HRP-catalzyed reaction of coniferyl alcohol with Tyr were analyzed by HPLC and then by LC/MS. Surprisingly, the mass spec data indicated that Tyr was linked through the phenolic hydroxyl to the α-carbon of coniferyl alcohol. LC/MS results showed that, for the new product of coniferyl alcohol/Tyr reaction, mass to charge ratio (*m/z*) of the major ion is 540, which corresponds to the mass of coniferyl alcohol dimer (358) and Tyr (181), plus 1 from the addition of proton during ionization. The 359.2 peak corresponds to loss of Tyr. These results can only be explained by a nucleophilic attack of the quinone methide by the phenolic hydroxyl (Fig. 3-2A).

To examine the reason behind a lack of free radical coupling products between Tyr and coniferyl alcohol, we used stop flow kinetics. The rate constants between Tyr with

HRP compound I and II were determined. Similar experiments were performed with coniferyl alcohol. The reactivity of coniferyl alcohol was at least 600 times greater than that of Tyr. However, the reactivity of vanillyl alcohol, a proxy for more cross linked lignin, is more similar to that of Tyr. This would then indicate that Tyr would cross link with lignin only at the later stages of polymerization.

The product analysis from Tyr indicating that amino acids can be cross linked with lignin through nucleophilic attack of the quinone methide resulted in us examining other amino acids with nucleophilic side chains. As such, Ser, Thr and Cys were also tested. Mass spectral analysis of unique products as analyzed by HPLC indicated that Thr and Cys also made adducts with the lignin dimeric product. The proposed structures of new products of coniferyl alcohol/amino acids reaction were shown in Fig. 3-2A, Fig. 3-2B and Fig. 3-2C.

Again, the mechanism is consistent via reaction with the quinone methide intermediate. The quinone methide is the intermediate from radical coupling of two phenylpropanoid free radicals. When the radical on the β-position of a monolignal is involved in a coupling reaction, the quinone methides is produced (Freudenberg 1959). The quinone methide is a relatively unstable structure and susceptible to nucleophilic attack. The predominant nucleophile is water yielding a α-hydroxyl group. However, phenols and alcohol groups from either the propyl side chain or from carbohydrates have been proposed to form adducts through nucleophilic attack (Freudenberg 1959). The addition of carbohydrate to lignin quinone methide is considered as the main way by which lignin and polyssacharides are cross linked in the plant cell wall (Morrison 1974).

While the mass spectral analysis showed the products from the dimeric quinone methide with the amino acid, further experiments showed that the cross linking can occur at the protein level and also at the level of DHP. This was shown in our work with the glycine rich protein and the modified glycine rich protein which contain different Cys content. Both of our heterologously-expressed proteins GRP and MGRP, when added to coniferyl alcohol cross linking reactions, were modified with lignin-like compounds as determined by scintillation counting of the washed protein (after Ni-agarose affinity purification). Also, SDS-PAGE of the protein followed by autoradiography showed that the proteins contained ^{3}H-coniferyl alcohol-derived products. Treatment of the protein with iodoacetamide eliminated their ability to cross link with lignin.

In conclusion, our work provides the chemical details of how proteins can cross link with lignin. Our original strategy for cross linking proteins with lignin involved the use of Tyr in the hope that the phenolic substituent would cross link via a free radical mechanism. However, the present work clearly shows that the preferred route of cross linking of Tyr is through the quinone methide. We also showed that other amino acids with nucleophilic side chains, Thr and Cys also can covalently attach to lignin through a quinone methide intermediate. This now expands our ability to modify the lignin structure by directing secretion into the apoplast of proteins with high Tyr, Thr or Cys content.

REFERENCES

Boerjan, W., K. Morreel, O. Dima, H. Kim, F. C. Lu, C. Niculaes, R. Vanholme, R. Dauwe, G. Goeminne, D. Inze, E. Messens and J. Ralph (2010). "Mass Spectrometry-Based Sequencing of Lignin Oligomers." Plant Physiology**153**(4): 1464-1478.

Boerjan, W., J. Ralph and M. Baucher (2003). "Lignin biosynthesis." Annual Review of Plant Biology**54**: 519-546.

Bonilla, I., C. Mergold-Villasenor, M. E. Campos, N. Sanchez, H. Perez, L. Lopez, L. Castrejon, F. Sanchez and G. I. Cassab (1997). "The aberrant cell walls of boron-deficient bean root nodules have no covalently bound hydroxyproline/proline-rich proteins." Plant Physiology**115**(4): 1329-1340.

Bull, S. R. (1991). "The U.S. Deparment of Energy Biofuels Research Program." Energy Resource**13**: 10.

Cassab, G. I. (1998). "Plant cell wall proteins." Annual Review of Plant Physiology and Plant Molecular Biology**49**: 281-309.

Cathala, B., B. Saake, O. Faix and B. Monties (1998). "Evaluation of the reproducibility of the synthesis of dehydrogenation polymer models of lignin." Polymer Degradation and Stability**59**(1-3): 65-69.

Cheung, A. Y., H. Wang and H. M. Wu (1995). "A Floral Transmitting Tissue-Specific Glycoprotein Attracts Pollen Tubes and Stimulates Their Growth." Cell**82**(3): 383-393.

Chundawat, S. P., G. Bellesia, N. Uppugundla, L. da Costa Sousa, D. Gao, A. M. Cheh, U. P. Agarwal, C. M. Bianchetti, G. N. Phillips, P. Langan, V. Balan, S. Gnanakaran and B. E. Dale (2011). "Restructuring the crystalline cellulose hydrogen bond network enhances its depolymerization rate." J Am Chem Soc**133**(29): 11163-11174.

Cotner, R. C. and C. O. Clagett (1973). "O-Carboxamidomethyl Tyrosine as a Reaction-Product of Alkylation of Proteins with Lodoacetamide." Analytical Biochemistry**54**(1): 170-177.

Freudenberg, K. (1959). "Biosynthesis and Constitution of Lignin." Nature**183**(4669): 1152-1155.

Helm, R. F. (1998). "Lignin-polysaccharide interactions in woody plants." Abstracts of Papers of the American Chemical Society**215**: U127-U127.

Iiyama, K., Z. F. Jin, K. S. Katsumata and T. B. T. Lam (2006). "Covalent linkages between cellulose and lignin in cell walls of coniferous and nonconiferous woods." Biopolymers**83**(2): 103-110.

Jamet, E., H. Canut, G. Boudart and R. F. Pont-Lezica (2006). "Cell wall proteins: a new insight through proteomics." Trends in Plant Science**11**(1): 33-39.

Keller, B., M. D. Templeton and C. J. Lamb (1989). "Specific Localization of a Plant-Cell Wall Glycine-Rich Protein in Protoxylem Cells of the Vascular System." Proceedings of the National Academy of Sciences of the United States of America**86**(5): 1529-1533.

Knox, J. P., P. J. Linstead, J. Peart, C. Cooper and K. Roberts (1991). "Developmentally Regulated Epitopes of Cell-Surface Arabinogalactan Proteins and Their Relation to Root-Tissue Pattern-Formation." Plant Journal**1**(3): 317-326.

Kreuger, M. and G. J. Vanholst (1993). "Arabinogalactan Proteins Are Essential in Somatic Embryogenesis of Daucus-Carota L." Planta**189**(2): 243-248.

Kuan, I. C., K. A. Johnson and M. Tien (1993). "Kinetic-Analysis of Manganese Peroxidase - the Reaction with Manganese Complexes." Journal of Biological Chemistry**268**(27): 20064-20070.

Laemmli, U. K. (1970). "Cleavage of structural proteins during the assembly of the head of bacteriophage T4." Nature**227**(5259): 680-685.

Laidlaw, R. A. and G. A. Smith (1965). "Proteins of Timber of Scots Pine (Pinus Sylvestris)." Holzforschung**19**(5): 129-&.

Leary, G. J. (1980). "Quinone Methides and the Structure of Lignin." Wood Science and Technology**14**(1): 21-34.

Lewis, N. G., R. A. Razal and E. Yamamoto (1987). "Lignin Degradation by Peroxidase in Organic Media - a Reassessment." Proceedings of the National Academy of Sciences of the United States of America**84**(22): 7925-7927.

Liang, H. Y., C. J. Frost, X. P. Wei, N. R. Brown, J. E. Carlson and M. Tien (2008). "Improved sugar release from lignocellulosic material by introducing a tyrosine-rich cell wall peptide gene in poplar." Clean-Soil Air Water**36**(8): 662-668.

Lu, F. C. and J. Ralph (1998). "Highly selective syntheses of coniferyl and sinapyl alcohols." Journal of Agricultural and Food Chemistry**46**(5): 1794-1796.

Ludley, F. H. and J. Ralph (1996). "Improved preparation of coniferyl and sinapyl alcohols." Journal of Agricultural and Food Chemistry**44**(10): 2942-2943.

Malencik, D. A., J. F. Sprouse, C. A. Swanson and S. R. Anderson (1996). "Dityrosine: Preparation, isolation, and analysis." Analytical Biochemistry**242**(2): 202-213.

Martius, C. (1992). "Density, Humidity, and Nitrogen-Content of Dominant Wood Species of Floodplain Forests (Varzea) in Amazonia." Holz Als Roh-Und Werkstoff**50**(7-8): 300-303.

McDougall, G. J., D. Stewart and I. M. Morrison (1996). "Tyrosine residues enhance cross-linking of synthetic proteins into lignin-like dehydrogenation products." Phytochemistry**41**(1): 43-47.

Merewether, J. W. T. (1957). "A lignin-carbohydrate complex in wood." Holzforschung**11**: 16.

Morreel, K., H. Kim, F. C. Lu, O. Dima, T. Akiyama, R. Vanholme, C. Niculaes, G. Goeminne, D. Inze, E. Messens, J. Ralph and W. Boerjan (2010). "Mass Spectrometry-Based Fragmentation as an Identification Tool in Lignomics." Analytical Chemistry**82**(19): 8095-8105.

Morrison, I. M. (1974). "Lignin-Carbohydrate Complexes from Lolium-Perenne." Phytochemistry**13**(7): 1161-1165.

Morrison, I. M. (1974). "Structural Investigations on Lignin-Carbohydrate Complexes of Lolium-Perenne." Biochemical Journal**139**(1): 197-204.

Neilson, M. J. and G. N. Richards (1982). "Chemical Structures in a Lignin Carbohydrate Complex Isolated from the Bovine Rumen." Carbohydrate Research**104**(1): 121-138.

Rubinstein, A. L., J. Marquez, M. SuarezCervera and P. A. Bedinger (1995). "Extensin-like glycoproteins in the maize pollen tube wall." Plant Cell**7**(12): 2211-2225.

Scheller, H. V. and P. Ulvskov (2010). "Hemicelluloses." Annual Review of Plant Biology, Vol 61**61**: 263-289.

Smythe, C. V. (1936). "The reaction of iodoacetate and of iodoacetamide with various sulfhydryl groups, with urease, and with yeast preparations." Journal of Biological Chemistry**114**(3): 601-612.

Taylor, N. G. (2008). "Cellulose biosynthesis and deposition in higher plants." New Phytologist**178**(2): 239-252.

Whitmore, F. W. (1978). "Lignin-Protein Complex Catalyzed by Peroxidase." Plant Science Letters**13**(3): 241-245.

Whitmore, F. W. (1982). "Lignin-Protein Complex in Cell-Walls of Pinus-Elliottii - Amino-Acid Constituents." Phytochemistry**21**(2): 315-318.

Wyman, C. E. (1994). "Ethanol from Lignocellulosic Biomass - Technology, Economics, and Opportunities." Bioresource Technology**50**(1): 3-16.

CHAPTER 4

Improved Digestibility of Poplar by Expression of a Tyr/Cys Rich Peptide in the Cell Wall

INTRODUCTION

Lignocellulose is the most abundant renewable biomass on Earth. In recent years, lignocellulosic material has gained much attention as a high-volume, low-cost resource for sustainable production of liquid fuels to substitute fossil fuels (Wyman 1994). The conversion of lignocellulosic biomass into ethanol involves an initial process referred to as pretreatment which is followed by enzymatic hydrolysis of the cellulose and hemicellulose to fermentable sugars, and then fermentation of the monosaccharides to ethanol (Pandey et al. 2010). The chemical properties make lignin difficult to degrade, chemically or biochemically. Lignin is thus a formidable barrier for enzymes and as such it limits the accessibility of hydrolytic enzymes to cellulose and hemicellulose. Lignin is therefore an impediment toward the conversion of polysaccharides to ethanol (Zaldivar et al., 2001).

Pretreatment is a process involving chemical, physical and thermal means which results in disrupting the lignin seal and also disrupts the crystalline structure of cellulose (Ladisch et al. 2005). By altering the structure of lignocellulosic biomass, pretreatment can make polysaccharides more accessible to hydrolytic enzymes and in turn, increase the yield of fermentable sugars (Yang and Wyman 2008). However, pretreatment processes are expensive, energy-intensive and produce environmental toxins (Yang and Wyman 2008). Green technologies to circumvent the lignin barrier by reducing lignin content in biomass

have evolved in recent years to substitute physical/chemical pretreatment and/or decrease the severity of pretreatment (Dixon and Chen 2007; Dixon et al. 2011).

A strategy used in recent years to circumvent the lignin barrier is by modifying the synthesis of lignin to decrease total lignin content (for review see (Chapple et al. 2008)). Most if not all of the steps involved in the biosynthesis of monolignols have been identified (Boerjan et al. 2003), which provides many potential genetic targets to alter lignin content. In the last decades, by taking advantage of plant transformation technologies, researchers have successfully changed lignin content in a variety of plant species (Chiang et al. 1999; Chapple et al. 2008) and therefore facilitate the downstream applications. For example, work with alfalfa by Reddy et al. (Reddy et al. 2005) achieved up to 40% reduction in lignin content by antisense down-regulation of *p*-coumarate 3-hydroxylase (C3H), and resulted in an increase in forage digestibility by 60% (Reddy et al. 2005). While decreasing lignin content can result in an increase in lignocellulose digestibility (Chapple et al. 2008), there is concern with reduced agronomic fitness of such plants. Fitness issues involve growth rate, productivity, and resistance to stresses (Pedersen et al. 2005; Li et al. 2008).

To circumvent the lignin barrier, we have developed a strategy where we have incorporated peptide linkages within the lignin polymer (Liang et al.). We have replaced a portion of lignin-lignin linkages with lignin-peptide linkages. Such a strategy does not decrease the lignin content but does increase the digestibility of the lignocellulosic material (Liang et al. 2008). Our previous work using this strategy utilized a Tyr-rich peptide secreted into the lignifying cell wall of poplar (Liang et al. 2008). It resulted in increased sugar release after enzymatic hydrolysis.

In the present study, we have made new constructs in attempt to further increase the digestibility of transgenic plants. We have designed two constructs to transform into poplar. One construct encodes a protein with an increased Tyr content (17%) and the other encodes a protein with high Cys content (11%). Our unpublished work has shown that Cys side chain can also cross link with lignin.

MATERIALS AND METHODS

Binary vector construction. The cDNA (234 bp, the last 78 amino acids of the C terminus) encoding hydroxyproline-rich glycoprotein gene from *Petroselinum crispum* (parsley) (Kawalleck et al. 1995) was cloned into binary vector PBI101 by replacing the β-glucuronidase (GUS) gene. This truncated hydroxyproline-rich glycoprotein contains 17% Tyr. The sequence encoding the signal peptide of a lodgepole pine CβG (Samuels et al. 2002) was fused with the transgene to direct the secretion of the protein to the extracellular space. The poplar phenylalanine ammonia lyase gene (PAL2) promoter (Gray-Mitsumune et al. 1999) was chosen for expression in lignifying tissues. The resulting vector was named as NTRG2.

Another construct encodes the modified glycine-rich protein (MGRP) is derived from maize (GeneBank ACG24587). The Cys content of the protein was increased to 11% Cys by the following changes: E8C, E67C and E70C. MGRP was also inserted into PBI101. The sequence encoding the signal peptide of a poplar CβG (unpublished data) was fused before MGRP. The double CaMV 35S promoter (Omirulleh et al. 1993) was used for high expression. The resulting vector was named as MGRP.

Poplar transformation. Binary vectors NTRG2 and MGRP were transformed into *Agrobacterium tumefaciens* strain EHA105 for hybrid poplar *Populus deltoides*×*nigra* (Ogy) transformation. Poplar transformation and regeneration were conducted according to Liang et al. (Liang et al. 2001). When plants were transplanted from tissue culture in soil, genomic DNA was extracted from leaves of transgenic and wildtype plants, according to the manual of Plant DNAzol® Reagent (Invitrogen, USA) and used as the templates of

PCR amplification to check the presence of the transgene. For transgenic poplars of NTRG2, primers were designed to amplify the DNA sequence of pine CβG signal peptide and NTRG2: forward: 5'ATGGAGGTGTCTGTGTTGATGTGGGTACTG3' and reverse: 5'GCGTGGGGACTTATAGTAATAAGGAGTCG3'. Similarly, for transgenic poplars of MGRP, PCR primers were forward: 5'ATGGGAATTTCTTCACTTTGTAAAGC3' and reverse: 5'GTTATGCACGCACACCAAGCAAGGTC3'.

Transcription of the transgene. The total RNA was extracted from leaves of transgenic and wildtype plants, according to the protocol of RNeasy Plant Mini Kit (QIAGEN Sciences, MA, USA) and treated with DNAse (Ambion, TX, USA). RNA was reverse-transcribed and PCR amplified, (SuperScript® III One-Step RT-PCR System with Platinum® Taq DNA Polymerase,Invitrogen, CA, USA), with transgene-specific primers as used for PCR of genomic DNA. Ribosomal 18S primers (forward: 5'AAACGGCTACCACATCCAAG3'; reverse: 5'CCCAACCCAAAGTCCAACTA3') were included in this experiment for confirming that the same amount of RNA was used for RT-PCR among different plant samples.

UV detection and histochemical staining of lignin. Stems from internodes between leaf plastochron index 8 and 9 were transversely sectioned. These sections of transgenic and wildtype plants were observed under UV light to determine lignin distribution by auto-fluorescence of lignin. Also, the sections were stained with phloroglucinol-HCl according to Liang et al. (Liang et al. 2008), which reflects the total lignin content in the stems.

Compositional analysis of the cell wall. Stems of transgenic poplars of NTRG2 from internodes between leaf plastochron index 10-19 were grounded by Wiley Mill (Thomas Scientific, NJ, USA), and screened through Mesh 20. This procedure was according to NREL Laboratory Analytical Procedures for standard biomass analysis (Sluiter 2008). To determine the dry weight, 0.25 g of air-dried sample were further dried in the oven at 105 ℃ for 24 h. Water and ethanol extractives were removed prior to structural carbohydrate measurements by loading 0.5 g of air-dry sample into the ASE 350 system (DIONEX, CA, USA). The extractive-free biomass samples were subjected to hydrolysis with 72% (w/w) H_2SO_4 at 30 ℃ for 1 h and subsequently with 4% (w/w) H_2SO_4 at 121 ℃ for 1 h in autoclave. The weight of the remaining insoluble portion is measured as Klason lignin (acid-insoluble lignin) with ash. In order to determine ash weight and calculate the Klason lignin weight, the remaining insoluble portion was placed in furnace (105 ℃, for 12 min, 250 ℃ for 30 min, 575 ℃ for 3 h, and 105 ℃ for overnight). The acid-soluble lignin was measured by spectrophotometer at 240 nm with the extinction coefficient of 25 $M^{-1}cm^{-1}$. Glucose content was measured with DIONEX ICS3000 system coupled with anion-exchange CarboPac PA20 analytical column and pulsed amperometric detector (Dionex, USA).

In addition to the procedure described above, lignin content of MGRP transgenic and wildtype plants was determined by molecular-beam mass-spectrometry, according to Agblevor et al. (Agblevor et al. 1994). Stems of transgenic poplars of MGRP from internodes between leaf plastochron index 1-10 were grounded by Wiley Mill, screened through Mesh 20 and used for lignin analysis.

Digestibility assays of steam-exploded pretreated samples. For poplars transformed with the NTRG2 construct, stem samples were harvested from internodes between leaf plastochron index 10-19, grounded by Wiley Mill, and screened through Mesh 20. Samples were analyzed using the standard Bioenergy Sciences Center high throughput recalcitrance screening pipeline method developed at NREL with a few modifications (Decker et al. 2009). Briefly, 5.0 mg samples were dispensed in triplicate into a custom 96-well reactor plate using the Symyx Powdernium MTM robot. Water (200 μL) was added to each sample, and the reactor plate was sealed with a Teflon film tape with a silicone-based adhesive. Holes were punched in the film for the steam ports and the plate was clamped into the reactor. Samples were heated for 40 min at 180 ℃, cooled with water, and removed from the steam chamber. Then 70 mg SpezymeCP (Genencor International, Inc., NY, USA) and 3 mg Novozymes188 (Novozymes, USA) per g biomass was added to each sample in 50 μL of 1.0 M sodium citrate, pH 5.0. The plates were sealed, mixed, incubated for 72 h, and then analyzed for glucose release via the glucose oxidase/peroxidase enzyme based assays (Ionescus.F 1974).

Digestibility assays of protease and hot water pretreated samples. The stems of transgenic poplar NTRG2 (between leaf plastochron index 10-19) and MGRP (between leaf plastochron index 1-10) were harvested, grounded by Wiley Mill, and screened through Mesh 20. Ground tissues were extracted with water, 80% ethanol and acetone, and then air-dried. Approximately 30 mg of ground tissues were weighed and 500 μl of 50 mM Tris-Cl, 1 mM $CaCl_2$, pH 8.0 were added to each sample. After the pretreatment at 121 ℃ in an autoclave for 20 min, the sample was cooled and then 15 μl of proteinase K (Sigma-Aldrich, USA) (stock solution of 100 mg/ml 50 mM Tris-Cl, 1 mM $CaCl_2$, pH 8.0) or as a

control, just buffer, was added to the samples. Samples were digested overnight at 37 ℃ with shaking. For cellulase hydrolysis, 2.5 ml of 100 mM sodium citrate, pH 5.0 and 10 μL of cellulase (C2730, 840 units/mL) (Sigma-Aldrich, USA) were added to the samples. Samples were incubated at 50 ℃ for 3 days with shaking. The undigested stem residue were collected by centrifugation (13,000 ×g, 5min), washed with 100 mM sodium citrate, pH 5.0, water, and acetone, and dried for overnight at room temperature before weighing.

RESULTS

Expression of transgene. *A. tumefaciens* harboring the vectors NTRG2 or MGRP were used to transform "Ogy" leaf disks. Regenerated explants for each vector were transferred to the new medium for induction of shoots and roots growth. After shoots and roots growth were initiated and elongated, three transgenic plants of each vector were transferred to soil and then grown in a greenhouse under 16 h/8 h light dark condition. To confirm the presence of transgenes in the genome of the transgenic poplars, PCR was conducted with genomic DNA and primers of transgenes (see Materials and Methods for primers). For the three different NTRG2 transformants (NTRG2-1, NTRG2-7, and NTRG2-11), PCR amplicons of 408 bp were detected (Fig. 4-1). For plants of MGRP-1, MGRP-2, and MGRP-8, PCR amplicons of 306 bp were detected (Fig. 4-1). No bands were detected in the wildtype.

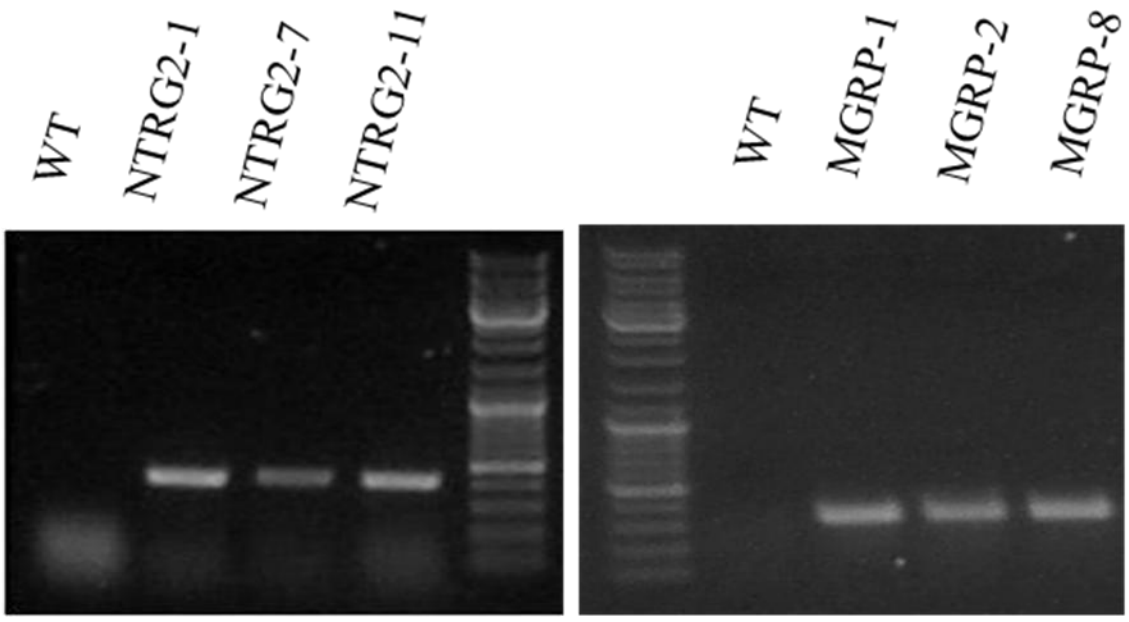


Figure 4-1. PCR of genomic DNA from wildtype and NTRG2 and MGRP. Genomic DNA was prepared from leaves as described in Materials and Methods.

RT-PCR was also used to examine the transcription of the transgene in different transformants with the same primers used for PCR of genomic DNA. For the three transformants from NTRG2 and MGRP, DNA fragments of expected size were detected after 30 cycles of amplification and again were not detected in wildtype (Fig. 4-2). Ribosome 18S primers were used as standard to confirm that the same amounts of RNA were used among samples (Fig. 4-2).

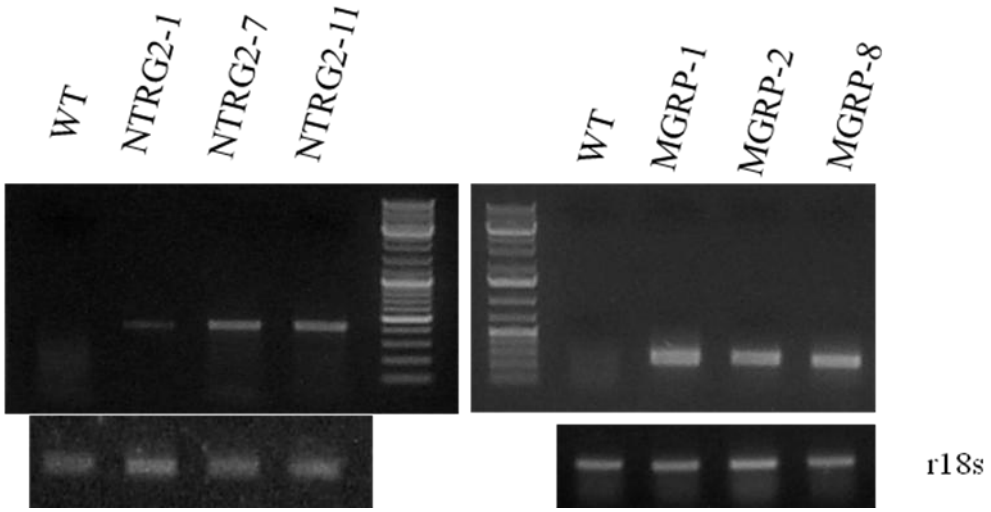


Figure 4-2. RT-PCR detection of transgene transcription. Total RNA was isolated from leaves as described in Materials and Methods. RT-PCR was carried out and poplar ribosomal 18S was used as a RNA loading control.

Phenotype of transgenic poplar. Compared to wildtype *OGY*, the NTRG2 transgenic poplars grew without any apparent aberrant phenotype under greenhouse conditions. However, MGRP grew much slower with a dwarf phenotype, narrow leaf, and curly stem (Fig. 4-3). Shown in Fig. 4-3 are wildtype and NTRG2-1, grown from a stem cutting for 3 months and MGRP-1 which was transferred from medium and grown in soil for 11 months.

Figure 4-3. Phenotype of transgenic and wildtype plants grown under greenhouse condition. Plants shown left to right are wildtype (grown from a stem cutting for 3 months), transgenic NTRG2-1 (grown from a stem cutting for 3 months), MGRP-1 (transferred from medium and grown in soil for 11 months).

To examine the effect of NTRG2 and MGRP transgenes on lignin deposition in the plant stems, we examined the stem sections by UV light microscopy. As shown in Fig. 4-4, the UV fluorescence pattern and intensity in transgenic plants of NTRG2-1 and wildtype are similar. There is no obvious abnormal cellular morphology observed with the NTRG2-1 transgenic poplars. In contrast, transgenic poplars of MGRP-1 showed smaller stem radius, curly stem edges, and thinner xylem width. Plant stem sections were also stained with phloroglucinol-HCl to visualize total lignin (Lewis and Yamamoto 1990). As with the UV visualization, the staining patterns with wildtype and transgenic plants of NTRG2-1 were similar, indicating that there were no great changes in total lignin content and deposition. MGRP-1 transgenic poplars showed vessel cells are smaller and lignification is impaired, suggesting that transgene of MGRP impact lignin synthesis and formation.

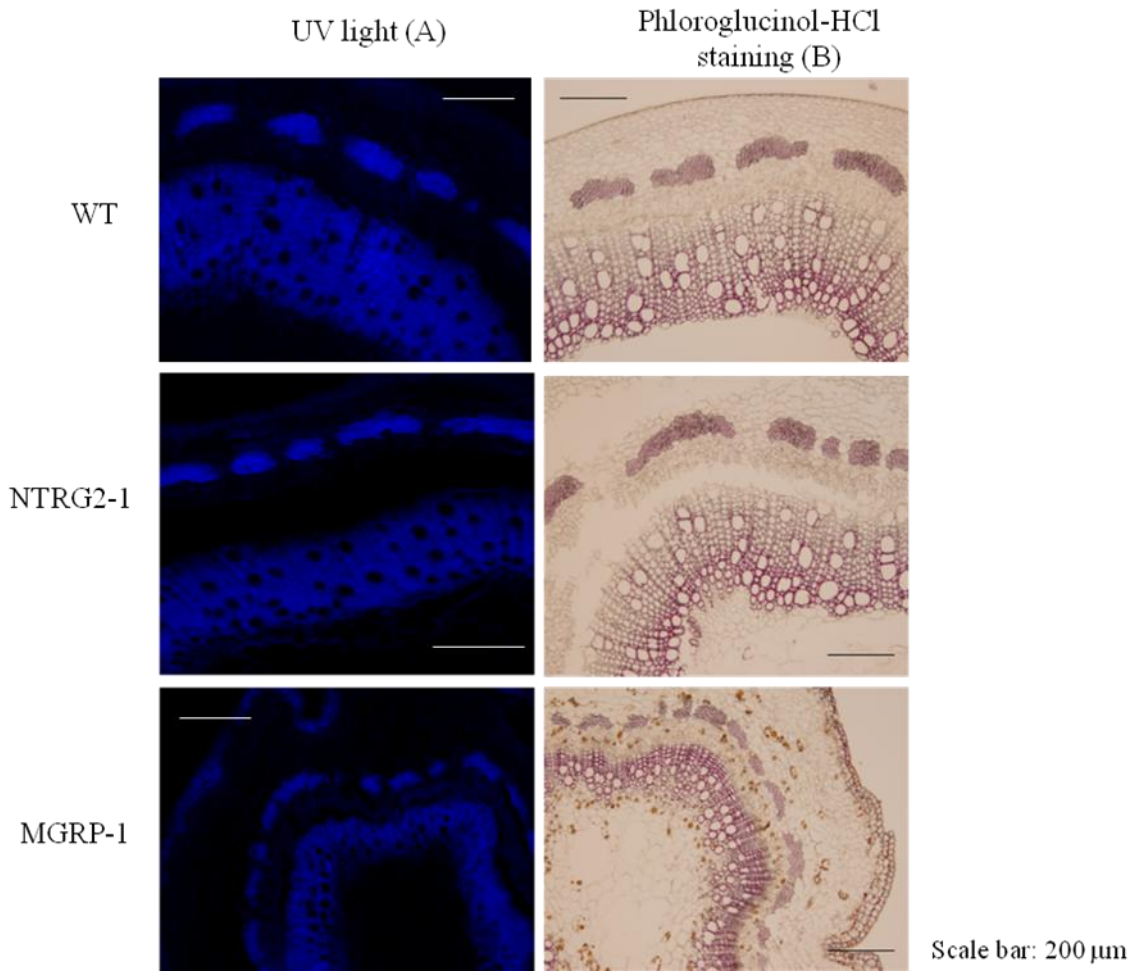


Figure 4-4. Visualization of lignin in transverse stem sections of wildtype, NTRG2, and MGRP. A. Stem tissues observed under UV light. B. Stem tissues stained with phloroglucinol-HCl.

Cell wall composition of transgenic poplar. To determine whether the transgene affects the cell wall composition of transgenic poplars, we analyzed acid-soluble lignin, acid-insoluble lignin, glucan, and xylan content in stems from internodes between leaf plastochron index 10-19 of NTRG2 transgenic poplars and wildtype. The results showed no significant differences in lignin, glucan or xylan content among wildtype and NTRG2 transgenic lines, suggesting that introduction of NTRG2 did not impair cell wall composition of transgenic poplars (Table 4-1).

Table 4-1. Acid-soluble lignin, acid-insoluble lignin, glucan, and xylan content of wildtype, NTRG2-1, NTRG2-7, and NTRG2-11.

	WT	NTRG2-1	NTRG2-7	NTRG2-11
Insoluble lignin %	$28.0^{x} \pm 2.9$	$27.2^{x} \pm 0.2$	$27.6^{x} \pm 1.7$	$28.3^{x} \pm 0.2$
Soluble lignin %	$6.5^{x} \pm 0.1$	$6.6^{x} \pm 0.3$	$6.5^{x} \pm 0.4$	$6.8^{x} \pm 0.1$
Glucan %	$28.4^{x} \pm 3.4$	$24.6^{x} \pm 2.4$	$28.7^{x} \pm 1.3$	$25.4^{x} \pm 0.4$
Xylan %	$6.8^{x} \pm 0.5$	$5.6^{x} \pm 0.4$	$9.6^{x} \pm 1.5$	$7.3^{x} \pm 0.4$

Lignin samples are from either acid soluble or acid insoluble treatments. Standard deviation was determined from two replicates. Dissimilar superscripts (x) indicate significant ($P < 0.05$) treatment differences

Due to the low biomass yield of MGRP transgenic poplars, analysis for lignin and for glucans could not be performed as for NTRG2. As such, a microscale method using molecular-beam mass-spectrometry (MBMS) was adapted to determine the lignin content. All three MGRP transgenic poplars showed major reduction in lignin content compared to wildtype (Table 4-2). This indicates that lignin synthesis, polymerization and/or deposition in the stem was affected by the MGRP transgene.

Table 4-2. Lignin content of wildtype, MGRP-1, MGRP-2, and MGRP-8.

	WT	MGRP-1	MGRP-2	MGRP-8
Lignin	24.6	13.3	14	10.6

Lignin content determined by molecular-beam mass-spectrometry. One sample for each plant was measured.

Digestibility and pretreatment of transgenic poplar. The transgenic poplars were then analyzed for digestibility. Wildtype and the transgenic NTRG2 and MGRP plants (three of each) were first subjected to hot water pretreatment after which they were treated overnight with proteinase K. Another set was incubated without proteinase K as a control. All samples were then incubated with the commercial cellulase mixture for 72 hours (see Material and Methods) and then washed and dried. The undigested matter was then weighed. In this measurement, digestibility was determined by the percent of undigested material relative to the dry weight of the starting stem material. As shown in Table 4-3, in wildtype, the addition of protease K had not effect on digestibility of the cell wall. Meanwhile, most transgenic poplars of NTRG2 and MGRP showed significant differences in stem digestibility between with and without protease treatment. Notably, most transgenic plants even without protease treatment exhibited increased digestibility compared to wildtype.

Table 4-3. Percent residue remaining after digestion of stem tissues of wildtype, NTRG2 and MGRP.

Addition	WT	NTRG2-1*	NTRG2-7*	NTRG2-11	MGRP-1*	MGRP-2*	MGRP-8*
Protease	$64.1^{a} \pm 2.2$	$55.1^{bc} \pm 2.2$	$57.4^{bc} \pm 0.5$	$59.5^{ab} \pm 3.5$	$47.0^{d} \pm 0.5$	$50.6^{cd} \pm 3.1$	$47.7^{d} \pm 1.6$
None	$61.7^{w} \pm 1.5$	$50.0^{y} \pm 0.1$	$51.0^{xy} \pm 2.8$	$58.8^{wx} \pm 3.9$	$42.6^{z} \pm 3.1$	$43.4^{yz} \pm 2.3$	$38.2^{z} \pm 2.1$

Results are expressed as percent of starting stem weight remaining after digestion as described in Materials and Methods. Samples were pretreated with hot water and plus or minus proteinase as described in Materials and Methods. Plant stem tissues of three transformants of NTRG2, three transformants of MGRP, and wildtype were subjected to sequential treatment of hot water, plus or minus protease K and cellulase. Standard deviation was determined from three replicates. Star (*) indicates significant difference ($P<0.05$) between the treatment of protease and the treatment of none. Dissimilar superscripts (a-d and w-z) indicate significant ($P < 0.05$) treatment differences.

To further evaluate the effect of transgene on plant digestibility, the release of glucose from stems after steam-explosion pretreatment and saccharification was determined. The NTRG2 transgenic poplars and wildtype were pretreated under mild conditions. We reasoned that harsher pretreatment may mask the differences between the wildtype and NTRG2. The samples were moderately pretreated by steam explosion and then hydrolyzed with cellulases. NTRG2-1 and NTRG2-7 released 19.3 g and 19.1 g glucose from 100g biomass respectively, displaying 30% and 29% increases compared with wild type (Table 4-4). NTRG2-11 showed minor change in glucose release compared with wild type.

Table 4-4. Percent glucose released from wildtype and NTRG2 after steam-explosion and enzymatic hydrolysis.

	WT	NTRG2-1	NTRG2-7	NTRG2-11
Released glucose %	$14.8^{a} \pm 2.7$	$19.3^{b} \pm 1.4$	$19.1^{b} \pm 1.0$	$15.4^{a} \pm 0.8$

Results are given as grams of glucose released divided by dry weight of the stem times 100. Ground stem tissues of three transformants of wildtype and NTRG2 were subjected to steam explosion pretreatment and cellulase hydrolysis. Glucose content was determined with the glucose oxidase/peroxidase enzyme based assays as described in Materials and Methods. Standard deviation was determined from three replicates. Dissimilar superscripts (a and b) indicate significant ($P < 0.1$) treatment differences.

DISCUSSION

Circumventing the lignin barrier is central toward the hydrolytic enzymes gaining access to the cellulose polymer for bioethanol production (Chapple et al. 2008). Researchers have sought to minimize pretreatment costs, which can be 30% to 50% of the total cost of bioethanol production (Ladisch et al. 2005) and its negative down stream impact (Lynd 1996). Green technology alternatives to chemical pretreatment have received much attention in recent years (for review, see (Chapple et al. 2008)). One such strategy is use or generation of plants with lowered lignin content. A group of naturally occurring mutants of maize (bm) has been identified phenotypically by a reddish-brown color that occurs in vascular tissues of leaves and stems (Marita et al. 2003). The bm mutants exhibit up to 40% decrease in lignin content and displays a moderate to strong (15-46%) increase in digestibility (Barriere et al. 2004) making them attractive as bioenergy feedstocks (Vermerris et al. 2007). Plants with lowered lignin content have also been obtained by modern genetic methods (Boerjan et al. 2003). Manipulation of lignin biosynthesis via genetic modification to improve biomass digestibility have been accomplished in Arabidopsis (Franke et al. 2002), poplar (Lapierre et al. 1999), switchgrass (Dixon et al. 2011), and alfalfa (Reddy et al. 2005). A study with switchgrass showed that 15% reduction in lignin content by suppressing caffeic acid O-methyl transferase (COMT) results in 35% increase in ethanol production (Dixon et al. 2011).

The concern with lowered lignin content is the potential of reduced agronomic fitness (Pedersen et al. 2005). Plants with decreased lignin have exhibited reduced growth rate (Reddy et al. 2005), weakened pollen viability (Preston et al. 2004), collapsed vessels (Jones et al. 2001), impaired pathogen resistance (Nicholson et al. 1976), and decreased

biomass yield (Lee and Brewbaker 1984). The strategy that we have developed, that of splicing in peptide linkages with the lignin linkages (Liang et al. 2008) does not involve decreasing the lignin content. Our previous work with this strategy successfully increased the poplar digestibility by introducing Tyr rich peptide (12% Tyr) to the plant cell wall (Liang et al. 2008). To further test this strategy and to seek plants with even more enhanced digestibility, in the present study, we tested two constructs in poplar. One construct (NTRG2) is based on our previous work with Tyr except the present vector encodes a protein with even higher percentage of Tyr (17%). It contains the same promoter and signal peptide as previously described (Liang et al. 2008). The second vector (MGRP) tests the ability of a Cys-rich peptide to crosslink with lignin. Our unpublished data has shown that Cys can also cross link with lignin. While this vector used signal peptide of poplar coniferin-β-glucosidase, it differs in using the stronger double 35S promoter (Omirulleh et al. 1993).

Our results here with the two different constructs have yielded mixed results. The NTRG2 transformants may be potentially useful in bioethanol production. They showed a normal phenotype, cellular morphology, and cell wall composition. Furthermore, the digestibility studies, performed by two methods with NTRG2 transformants are promising. With digestibility tested with glucose release, NTRG2 exhibited up to 30% increase. When measured as residual residue after extensive digestion, NTRG2 lines contained less than that of wildtype. These results are obtained even without treatment with proteinase K.

In contrast to NTRG2, the transgenic poplars of MGRP displays slow growth, dwarf, narrow leaf, and curly stem. The phenotype observed with transgenic poplars of MGRP is consistent with the decreased lignin content of MGRP plants. As observed with

NTRG2, MGRP is apparently also easier to digest and containing less residue after hydrolysis. However, it is apparent that MGRP has fitness issues with the dwarf phenotype. The dramatic decrease in lignin content is probably due to that normal polymerization of lignin was impaired by a large extent of the formation of linkages between the Cys-rich protein and lignin. The dwarf phenotype has been observed in the naturally occurring (Lam and Iiyama 2000) or transgenic plants (Grima-Pettenati et al. 1998; Turner et al. 2001; Chapple et al. 2002; Reddy et al. 2005) with reduced lignin content. It these studies, dwarf phenotype were resulted in collapsed vessels in xylem caused by reduced lignin which impairs the water transportation. Therefore, we reasoned that the dwarf phenotype of MGRP transgenic poplars is also attributed to the collapsed vessels caused by reduced lignin content.

This study successfully created transgenic poplars carrying cell-wall-localized Tyr/Cys rich peptides, and the digestibility was increased in most of the surveyed transgenic poplars. Our results support the feasibility of the novel strategy we proposed before (Liang et al. 2008). The transgene with moderate promoter, PAL2, are able to increase plant digestibility without compromising plant fitness, while the transgene driven by strong promoter, double 35S promoter, increased the digestibility but also impaired plant fitness, which suggests normal lignification can be affected by large amount of cell wall proteins. Therefore, this strategy can reach a point of diminishing returns (compromising fitness). In the future study, moderate promoters will be adapted to limit the amount of cell-wall-localized peptides.

REFERENCES

Agblevor, F. A., R. J. Evans and K. D. Johnson (1994). "Molecular-Beam Mass-Spectrometric Analysis of Lignocellulosic Materials .1. Herbaceous Biomass." Journal of Analytical and Applied Pyrolysis **30**(2): 125-144.

Barriere, Y., J. Ralph, V. Mechin, S. Guillaumie, J. H. Grabber, O. Argillier, B. Chabbert and C. Lapierre (2004). "Genetic and molecular basis of grass cell wall biosynthesis and degradability. II. Lessons from brown-midrib mutants." Comptes Rendus Biologies **327**(9-10): 847-860.

Boerjan, W., J. Ralph and M. Baucher (2003). "Lignin biosynthesis." Annual Review of Plant Biology **54**: 519-546.

Boerjan, W., J. Ralph and M. Baucher (2003). "Lignin Biosynthesis." Annual Review of Plant Biology **54**(1): 519-546.

Chapple, C., R. Franke, M. R. Hemm, J. W. Denault, M. O. Ruegger and J. M. Humphreys (2002). "Changes in secondary metabolism and deposition of an unusual lignin in the ref8 mutant of Arabidopsis." Plant Journal **30**(1): 47-59.

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

Chiang, V. L., W. J. Hu, S. A. Harding, J. Lung, J. L. Popko, J. Ralph, D. D. Stokke and C. J. Tsai (1999). "Repression of lignin biosynthesis promotes cellulose accumulation and growth in transgenic trees." Nature Biotechnology **17**(8): 808-812.

Decker, S. R., R. Brunecky, M. P. Tucker, M. E. Himmel and M. J. Selig (2009). "High-Throughput Screening Techniques for Biomass Conversion." Bioenergy Research **2**(4): 179-192.

Dixon, R. A. and F. Chen (2007). "Lignin modification improves fermentable sugar yields for biofuel production." Nature Biotechnology **25**(7): 759-761.

Dixon, R. A., C. X. Fu, J. R. Mielenz, X. R. Xiao, Y. X. Ge, C. Y. Hamilton, M. Rodriguez, F. Chen, M. Foston, A. Ragauskas, J. Bouton and Z. Y. Wang (2011). "Genetic manipulation of lignin reduces recalcitrance and improves ethanol production from switchgrass." Proceedings of the National Academy of Sciences of the United States of America **108**(9): 3803-3808.

Franke, R., M. R. Hemm, J. W. Denault, M. O. Ruegger, J. M. Humphreys and C. Chapple (2002). "Changes in secondary metabolism and deposition of an unusual lignin in the ref8 mutant of Arabidopsis." Plant Journal **30**(1): 47-59.

Gray-Mitsumune, M., E. K. Molitor, D. Cukovic, J. E. Carlson and C. J. Douglas (1999). "Developmentally regulated patterns of expression directed by poplar PAL promoters in transgenic tobacco and poplar." Plant Molecular Biology **39**(4): 657-669.

Grima-Pettenati, J., J. Piquemal, C. Lapierre, K. Myton, A. O'Connell, W. Schuch and A. M. Boudet (1998). "Down-regulation of cinnamoyl-CoA reductase induces significant changes of lignin profiles in transgenic tobacco plants." Plant Journal **13**(1): 71-83.

Ionescus.F (1974). "Glucose Determination with Para-Amino-Dimethyl-Aniline as Chromogen through Glucoseoxidase-Peroxidase System." Revue Roumaine De Biochimie **11**(1): 25-34.

Jones, L., A. R. Ennos and S. R. Turner (2001). "Cloning and characterization of irregular xylem4 (irx4): a severely lignin-deficient mutant of Arabidopsis." Plant Journal **26**(2): 205-216.

Kawalleck, P., E. Schmelzer, K. Hahlbrock and I. E. Somssich (1995). "Two pathogen-responsive genes in parsley encode a tyrosine-rich hydroxyproline-rich glycoprotein (hrgp) and an anionic peroxidase." Mol Gen Genet **247**(4): 444-452.

Ladisch, M., N. Mosier, C. Wyman, B. Dale, R. Elander, Y. Y. Lee and M. Holtzapple (2005). "Features of promising technologies for pretreatment of lignocellulosic biomass." Bioresource Technology **96**(6): 673-686.

Lam, T. B. T. and K. Iiyama (2000). "Characteristics of senescent straw cell walls of dwarf, semidwarf, and normal strains of rice (Oryza sativa) plants." Journal of Wood Science **46**(5): 376-380.

Lapierre, C., B. Pollet, M. Petit-Conil, G. Toval, J. Romero, G. Pilate, J. C. Leple, W. Boerjan, V. Ferret, V. De Nadai and L. Jouanin (1999). "Structural alterations of lignins in transgenic poplars with depressed cinnamyl alcohol dehydrogenase or caffeic acid O-methyltransferase activity have an opposite impact on the efficiency of industrial kraft pulping." Plant Physiology **119**(1): 153-163.

Lee, M. H. and J. L. Brewbaker (1984). "Effects of Brown Midrib-3 on Yields and Yield Components of Maize." Crop Science **24**(1): 105-108.

Lewis, N. G. and E. Yamamoto (1990). "Lignin - Occurrence, Biogenesis and Biodegradation." Annual Review of Plant Physiology and Plant Molecular Biology **41**: 455-496.

Li, X., J.-K. Weng and C. Chapple (2008). "Improvement of biomass through lignin modification." The Plant Journal **54**(4): 569-581.

Liang, H. Y., C. J. Frost, X. P. Wei, N. R. Brown, J. E. Carlson and M. Tien (2008). "Improved sugar release from lignocellulosic material by introducing a tyrosine-rich cell wall peptide gene in poplar." Clean-Soil Air Water **36**(8): 662-668.

Liang, H. Y., C. A. Maynard, R. D. Allen and W. A. Powell (2001). "Increased Septoria musiva resistance in transgenic hybrid poplar leaves expressing a wheat oxalate oxidase gene." Plant Molecular Biology **45**(6): 619-629.

Lynd, L. R. (1996). "Overview and evaluation of fuel ethanol from cellulosic biomass: Technology, economics, the environment, and policy." Annual Review of Energy and the Environment **21**: 403-465.

Marita, J. M., W. Vermerris, J. Ralph and R. D. Hatfield (2003). "Variations in the cell wall composition of maize brown midrib mutants." Journal of Agricultural and Food Chemistry **51**(5): 1313-1321.

Nicholson, R. L., L. F. Bauman and H. L. Warren (1976). "Association of Fusarium-Moniliforme with Brown Midrib Maize." Plant Disease Reporter **60**(11): 908-910.

Omirulleh, S., M. Abraham, M. Golovkin, I. Stefanov, M. K. Karabaev, L. Mustardy, S. Morocz and D. Dudits (1993). "Activity of a Chimeric Promoter with the Doubled Camv 35s Enhancer Element in Protoplast-Derived Cells and Transgenic Plants in Maize." Plant Molecular Biology **21**(3): 415-428.

Pandey, R. A., S. Banerjee, S. Mudliar, R. Sen, B. Giri, D. Satpute and T. Chakrabarti (2010). "Commercializing lignocellulosic bioethanol: technology bottlenecks and possible remedies." Biofuels Bioproducts & Biorefining-Biofpr **4**(1): 77-93.

Pedersen, J. F., K. P. Vogel and D. L. Funnell (2005). "Impact of reduced lignin on plant fitness." Crop Science **45**(3): 812-819.

Preston, J., J. Wheeler, J. Heazlewood, S. F. Li and R. W. Parish (2004). "AtMYB32 is required for normal pollen development in Arabidopsis thaliana." Plant Journal **40**(6): 979-995.

Reddy, M. S. S., F. Chen, G. Shadle, L. Jackson, H. Aljoe and R. A. Dixon (2005). "Targeted down-regulation of cytochrome P450 enzymes for forage quality improvement in alfalfa (Medicago sativa L.)." Proceedings of the National Academy of Sciences of the United States of America **102**(46): 16573-16578.

Samuels, A. L., K. H. Rensing, C. J. Douglas, S. D. Mansfield, D. P. Dharmawardhana and B. E. Ellis (2002). "Cellular machinery of wood production: differentiation of secondary xylem in Pinus contorta var. latifolia." Planta **216**(1): 72-82.

Sluiter, A., B. Hames, R. Ruiz, C. Scarlata, J. Sluiter, D. Templeton, and Crocker, D. (2008). "Determination of Structural Carbohydrates and Lignin in Biomass." NREL Laboratory Analytical Procedures for standard biomass analysis.

Turner, S. R., L. Jones and A. R. Ennos (2001). "Cloning and characterization of irregular xylem4 (irx4): a severely lignin-deficient mutant of Arabidopsis." Plant Journal **26**(2): 205-216.

Vermerris, W., A. Saballos, G. Ejeta, N. S. Mosier, M. R. Ladisch and N. C. Carpita (2007). "Molecular breeding to enhance ethanol production from corn and sorghum stover." Crop Science **47**: S142-S153.

Wyman, C. E. (1994). "Ethanol from Lignocellulosic Biomass - Technology, Economics, and Opportunities." Bioresource Technology **50**(1): 3-16.

Yang, B. and C. E. Wyman (2008). "Pretreatment: the key to unlocking low-cost cellulosic ethanol." Biofuels Bioproducts & Biorefining-Biofpr **2**(1): 26-40.

CHAPTER 5

FUTURE RESEARCH

1. Subcellular localization of the signal peptide CβG has been characterized in both onion epidermal cells and poplar cells with transient-transformation technique in the present study. The enhanced green fluorescence protein (EGFP) was used as reporter protein. My results showed that signal peptide CβG is able to transport EGFP to the extracellular space. Therefore signal peptide CβG was selected for delivering the Tyr/Cys rich peptides to the poplar cell wall. However, the subcellular localization of Tyr/Cys rich peptides in the cell wall of transgenic poplars is still undemonstrated. High-resolution microscope, such as transmission electron microscope, coupled with immune-labeling technique can be adapted to answer this question. His-tag was already constructed at the C-terminus of Tyr/Cys rich protein. Therefore, anti-his can be directly utilized for immune-labeling process, which saves time and labor for generation of Tyr/Cys rich peptide antibody.

2. In the present study, I have characterized the chemical nature of cross linking between amino acids and proteins with lignin *in vitro*. Our findings provide a mechanism by which proteins and lignin can cross link and reveal the chemical nature of the interactions between protein and lignin. However, it would be of great significance to elucidate the linkages between lignin and proteins occurring in the plant cell wall. Pyrolysis of lignin with tetramethylammonium hydroxide (TMAH) is a technique to break lignin polymer and reveal the difference linkage types in the lignin (Geib et al. 2008). Pyrolyzed products will be analyzed with gas chromatography mass spectrum (GC/MS).

The products containing amino acid residue would be collected and purified, and the crosslinking products of amino acids and monolignol can be used as the proxy to find the amino acid-containing products. Then their structure will be analyzed with nuclear magnetic resonance (NMR).

3. This study successfully increased the digestibility of transgenic poplars carrying cell-wall-localized Tyr/Cys rich peptides, supporting the feasibility of the novel strategy we proposed before. However, when the transgene driven by strong promoter, transgenic poplars showed increased the digestibility but also impaired plant fitness, suggesting regular lignification can be disturbed by large amount of cell wall proteins. Therefore, in the future study, moderate promoters will be tested to limit the amount of cell-wall-localized peptides.

4. The smallest protein produced by the transgenic poplar is 10 kD. While this is relatively small, the cross linking strategy may work best with even smaller peptides. Smaller size proteins would allow for greater access of the Tyr-peptide to regions of the cell wall where the larger proteins may not be able access. To achieve this, an expression system that contains polyprotein will be utilized. The polyprotein will contain several peptide sequences linked together by the peptide ENLYFQ/S, which can be cleaved by nuclear-inclusion-a (NIa) proteinase. Several small poly-Tyr/Cys encoding genes will be linked with this linker sequence and with the NIa proteinase encoding sequence at the C-terminus. When the polyprotein is produced, NIa proteinase can release itself from polyprotein by recognizing and cleaving the linker sequence and continues to perform *cis* or *trans* functions to produced small peptides. This system has been successfully utilized in Arabidopsis and tobacco (Liang et al. 2005 and Ceriani et al. 1998).

Geib SM, Filley TR, Hatcher PG, Hoover K, Carlson JE, Jimenez-Gasco Mdel M, Nakagawa-Izumi A, Sleighter RL, Tien M. (2008) “Lignin degradation in wood-feeding insects”. Proc Natl Acad Sci. **35** 12932-12937

Ceriani, M.F., J.F. Marcos, H.E. Hopp, and R.N. Beachy (1998) Simultaneous accumulation of multiple viral coat proteins from a TEV-NIa based expression vector. Plant Mol. Biol. 36, 239-248.

Liang, H., H. Gao, C.A. Maynard, and W.A. Powell (2005) Expression of a self-processing, pathogen resistance-enhancing gene construct in Arabidopsis. Biotechnol. Lett. 27, 435-442.

VITA

Fang Cong

Education

The Pennsylvania State University — Aug 2006 – Dec 2011

Ph.D. in Plant Biology (Advisor: Dr. Ming Tien)

China Agricultural University — Sep 2003 – July 2006

B.S. Honors Program (Major in Biology)

Publications

- **Fang Cong,** Haiying Liang, John Carlson, and Ming Tien. "Seeking for signal peptides driving heterogeneous proteins to apoplastic space via secretion pathway in poplar", Manuscript in prep
- **Fang Cong,** Nicole Brown, and Ming Tien. "Understanding cross-linking between protein and lignin", Manuscript in prep
- **Fang Cong,** Haiying Liang, Nicole Brown, John Carlson, Steve Decker, and Ming Tien. " Improved digestibility of poplar by introducing cysteine rich peptide gene to facilitate bioethanol production", Manuscript in prep
- Yi Xu, Chin-Fu Chen, Christopher Frost, Abdelali Barakat, **Fang Cong**, John E. Carlson, Ming Tien, Haiying Liang. "Molecular characterization of a hybrid poplar clone transformed with a tyrosine-rich cell wall gene for improved sugar release", Manuscript in prep