AN INVESTIGATION INTO THE USE OF MONOLIGNOL GLUCOSIDES IN POPLAR LIGNIN BIOSYNTHESIS

by

WILLIAM KIM MACDONALD

B.Sc., The University of British Columbia, 1996

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE
in

THE FACULTY OF GRADUATE STUDIES
Department of Plant Science

We accept this thesis as confirming
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA
December 1999

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of PLANT SCIENCE

The University of British Columbia
Vancouver, Canada

Date Jan. 5/00

DE-6 (2/88)

Abstract

The existence of a monolignol glucoside/ β-glucosidase system for the transport and/or storage of lignin precursors in angiosperms is currently speculative. This thesis aimed to identify both active coniferin synthesis in poplar and a coniferin specific β-glucosidase. Radio-tracer feeding studies did not provide evidence of coniferin synthesis in poplar species. However, degenerate polymerase chain reaction was used to isolate partial gene fragments from three family 1 β-glucosidases.

The 1823bp cDNA of one of these genes, POP1, was cloned, sequenced and characterized. POP1 is expressed at moderate levels in young leaves but is not expressed to detectable levels in developing xylem. The recombinant POP1 protein cross-reacts with antibodies raised against a pine coniferin β-glucosidase but does not hydrolyze coniferin in *in vitro* enzyme assays. Based on these findings, it does not appear that POP1 is involved in poplar lignin synthesis. The role of the other family 1 β-glucosidase identified in this study, POP2 and POP3, is unknown.

The work presented here only begins to examine the family 1 β-glucosidases in poplar and in no way precludes the existence of coniferin β-glucosidases in this species. If future work manages to find a β-glucosidase that is involved in lignification it would provide valuable evidence towards the existence of a monolignol glucoside/ β-glucosidase transport system and would identify a potential target for lignin modification through genetic engineering.

Table of Contents

List of Tables

List of Figures

Abbreviations

AMP	ampicillin
bm3	maize brown midrib mutant
C4H	cinnamate 4-hydroxylase
CAD	cinnamyl alcohol dehydrogenase
CCoAOMT	caffeoyl:coenzyme A *O*-methyltransferase
C3H	coumarate 3-hydroxylase
CCoA3H	coumaryl:coenzyme A 3-hydroxylase
CCR	cinnamoyl:coenzyme A reductase
COMT	caffeic acid *O*-methyltransferase
4CL	4-coumarate:coenzyme A ligase
DFR	dihydroflavanol-4-reductase
DTT	dithiothreitol
EST	expressed sequence tag
F5H	ferulate 5-hydroxylase
G	guaicyl lignin
HCN	hydrogen cyanide
HPLC	high pressure liquid chromatography
LB	Luria broth
4-NPG	4-nitrophenyl β-glucoside
PAL	phenylalanine ammonia-lyase
PCR	polymerase chain reaction
PEG	polyethylene glycol
PVDF	polyvinylidenefluoride
PVPP	polyvinylpolypyrrolidone
RT PCR	reversion transcription polymerase chain reaction
S	syringyl lignin
SDS-Page	sodium dodecyl sulphate – polyacrylamide gel electrophoresis
TAE	tris-acetate buffer
TAL	tyrosine ammonia-lyase
TBST	tris-buffered saline-tween 20
TLC	thin layer chromatography
UDPG	uridine-5-diphosphoglucose
VRA-G	5[4-(β-D-glucopyranosyloxy)-3-methoxyphenylmethylene]-2-thioxothiazolidine-4-one-3-ethanoic acid
X-gal	5-bromo-4-chloro-3-indoyl-β-D-galactoside

Acknowledgments

I would sincerely like to thank Brian Ellis for his kind and patient guidance over this thesis work. The balance you achieve between offering guidance and nurturing independence is an art form. Thank you. I would also like to extend my gratitude to Carl Douglas who was always extremely helpful during my moments of need.

To the members of the Ellis lab, what can I say? We laughed, we cried, we drank beer and basically, good fun was had by all. Cheers. I would especially like to thank Stef and Jason who's friendships go far beyond the scope of a thesis.

To my parents, whose continued unconditional love and support gives me the strength to endure and strive beyond, thank you (again).

Chapter 1
INTRODUCTION

1.1 Lignin Structure and Function

Lignin is a complex, three-dimensional polymer of aromatic subunits derived from the phenylpropanoid pathway. During plant development, lignin is deposited in the secondary wall thickenings of specific plant cells such as tracheids, vessel elements, xylem and phloem fibers, and schlerids (Dharmawardhana *et al.* 1992). As an important cell wall component of vascular plants, it adds structural support, contributes to pathogen resistance and provides a hydrophobic surface necessary for longditudinal water transport. It is postulated that because of these attributes, evolution of the ability to produce lignin may have been important for the adaptation of aquatic plant species to terrestrial life.

The primary subunits of the lignin polymer are p-coumaryl alcohol, coniferyl alcohol and sinapyl alcohol (Figure 1.1). Differing only in the extent of methoxylation

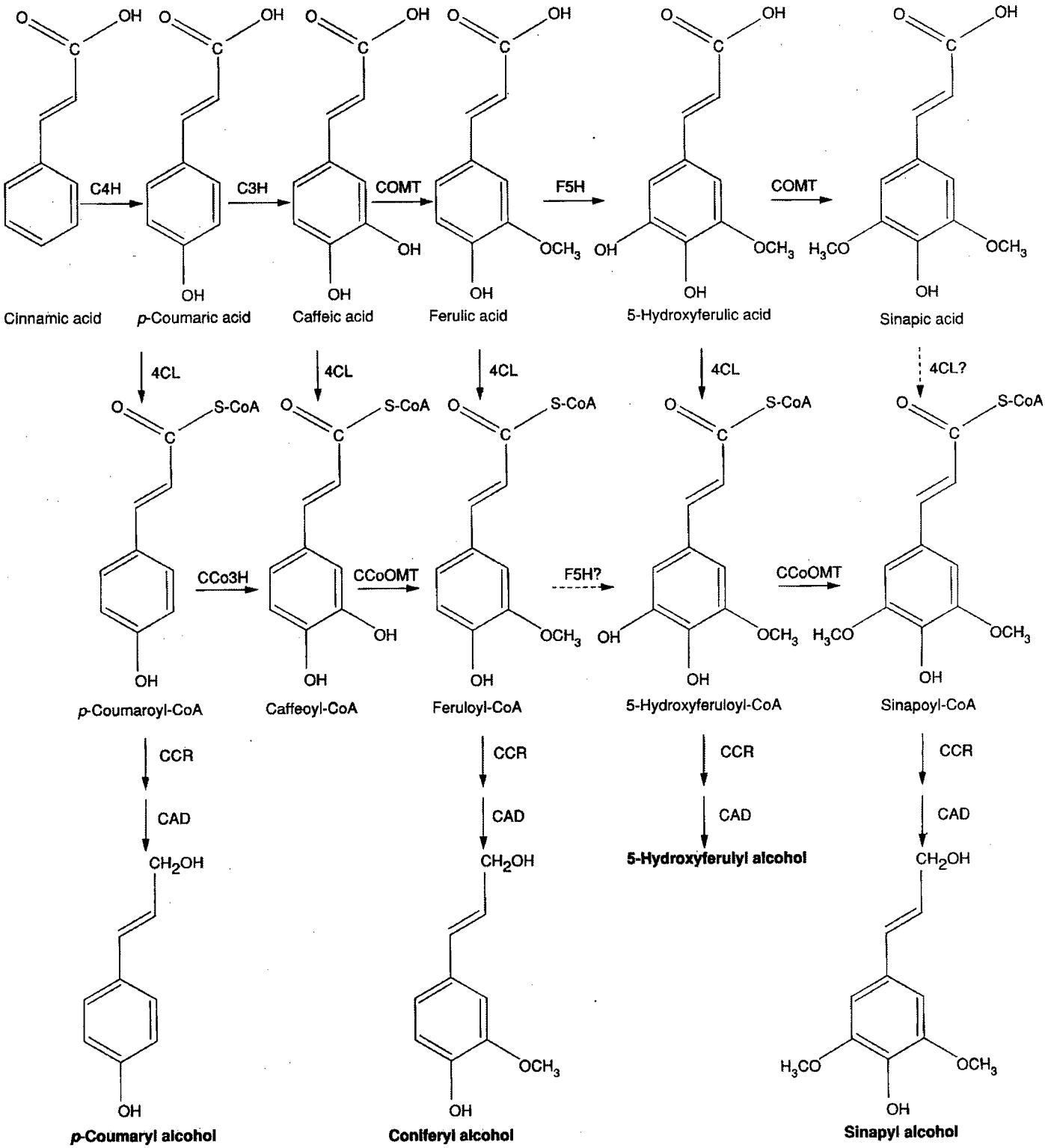


Figure 1.1 Monolignol Synthesis Pathway

about their phenylpropane ring, these monomers, or more specifically their relative proportions, are important factors in determining the properties of the lignin polymer. Within the lignin matrix, hydroxyl groups (of either the aromatic ring or the side chain) form bonds with each other, with aldehyde groups or with ketone groups. Generally, increased methoxylation reduces a monomer's ability to crosslink free ring positions with other molecules in the lignin matrix. Increased methoxylation also leads to greater steric hindrance within the polymer. Together, these two phenomena make a high methoxyl-content lignin less condensed and thus more accessible to enzymatic or chemical reagents.

1.2 Industrial Problems with Lignin

Lignin and cellulose, found intimately associated in the cell wall of vascular plants, are the two most abundant organic compounds on earth. Cellulose is economically valuable as a feed source for ruminants and as the primary constituent of paper. Lignin restricts access to the cellulose microfibrils and hence limits the economic value of plants used in these industries.

During the kraft pulping of woody plants for paper production, lignin is chemically removed as a processing byproduct (www.chem.vt.edu). The pulping begins with plant material (usually trees) being chipped into small pieces. The chips are then 'cooked' in a sodium hydroxide / sodium hydrosulfite solution for 0.5 to 3 hours at 180° C to delignify the fiber. The kappa number, which is an estimate of the lignin content in the plant material being pulped, is used to determine the amount of chemicals to be used

and length of digestion time. When extraction is complete, the solubilized lignin fraction, known as black liquor, is separated from the pulp (brownstock) and sent for chemical recovery. Black liquor treatment and waste disposal are costly procedures that unfavorably impact the environment. With the hope of reducing the quantity of expensive and environmentally harmful chemicals used during lignin extraction, considerable efforts are being made to produce trees for the forest industry with more easily extracted lignin. The potential financial gains from growing plants with reduced levels of lignin, or modified types of lignin, has stimulated research into understanding how plants make lignin and how lignin might be genetically modified to suit industry needs.

Domestic ruminants provide 70% of the total animal protein eaten, 80% of the milk consumed and 10% of the natural fiber used by humans (Fahey 1997). Ruminants digest forage crops, their major feed resource, with the help of symbiotic rumen microorganisms that are capable of digesting cell wall polysaccharides not amenable to mammalian enzymatic hydrolysis (Jung 1997). However, it has long been realized that lignin limits the digestion of cell wall polysaccharides, likely by physically shielding the cellulose from hydrolytic enzymes (Buxton and Redfearn 1997). This has led researchers to propose that forage crops containing less lignin could provide greater nutritional value to ruminants. For instance, the maize (*Zea mays*) brown midrib mutant (bm3) contains less total lignin than wild type maize plants and cellulose from bm3 mutants is more completely digested (Colenbrander *et al.* 1971, Barnes *et al.* 1972). Since ruminants use forage crops as their primary feed source, improvements in nutrient and energy uptake may be realized by reducing the lignin content in forage species.

The initial approaches to genetically modifying lignin content and / or composition have aimed to alter the activity of phenylpropanoid enzymes that contribute to lignin synthesis. Current models predict that lignification requires the involvement of ten to twenty different enzymes. Since the phenylpropanoid pathway leads to the synthesis of a variety of aromatic molecules (including lignin, flavonoids, coumarins, stillbenes, phenolic esters and benzoic acid derivatives) not all of the enzymes involved in lignin synthesis are necessarily specific to lignification. Suppression of the activity of enzymes acting early in the phenylpropanoid pathway may, therefore, have the potential to create pleiotropic effects on other metabolic endproducts. The logical target enzymes for genetic modification of lignin formation would be those that are used exclusively for lignin synthesis since this would avoid disrupting the production of other essential metabolites.

1.3 Core Phenylpropanoid Pathway Enzymes

1.3.1 Phenylalanine Ammonia-Lyase (PAL)

Phenylalanine ammonia-lyase catalyzes the non-oxidative deamination of L-phenylalanine to *trans*-cinnamic acid. In some species, mainly grasses, PAL may also convert L-tyrosine to *trans*-p-coumaric acid (4-hydroxycinnamic acid) (Neish 1961). This activity against L-tyrosine has been described as tyrosine ammonia-lyase (TAL) and claimed to represent a separate enzyme. However, PAL and TAL activities frequently copurify, and the heterologously expressed maize PAL was recently shown to have both

PAL and TAL activity *in vitro* (Rosler *et al.* 1997). PAL is a homotetrameric enzyme, but it is not known whether PAL can also exist as a heterotetramer, i.e. a holoenzyme containing subunits derived from different *PAL* genes in one cell.

PAL is encoded by multigene families in angiosperm and gymnosperm species. In the few cases examined in detail, each *PAL* gene encodes a polypeptide with similar catalytic properties, but PAL family members appear to be differentially expressed in response to developmental and environmental cues. Of the three *PAL* genes found in bean (*Phaseolus vulgaris* L.), only PAL2 was expressed in lignifying tissue while PAL3 was expressed in response to wounding (Liang *et al.* 1989). Furthermore, transgenic experiments with bean *PAL* promoters directing β-glucuronidase gene expression demonstrated that the different bean *PAL* promoters direct unique patterns of tissue-specific GUS staining and direct differential GUS staining in response to wounding and elicitation (Shufflebottom *et al.* 1993). These studies suggest that the timing of expression, tissue location, and responses to stimuli such as UV light, wounding or pathogen attack are uniquely regulated for each *PAL* gene within a species.

Tobacco (*Nicotiana tabacum* L. cv. Xanthi-nc) plants with reduced PAL activity have provided valuable information about the role of this enzyme in lignin synthesis (Bate *et al.* 1994). Reductions in PAL activity were observed in several transgenic lines overexpressing the *PAL2* gene from bean (a phenomenon known as sense suppression), and this resulted in a reduction in total lignin content. Lignin content was not strongly affected until PAL activity fell below 25% of control levels. The greatest reduction in

lignin content was observed in lines with only trace PAL activity (0.2% of controls) where stem lignin levels dropped to 10% that of wild type plants.

Unexpectedly, the lignin in PAL-suppressed tobacco plants had an increased S/G ratio. It appears that, in these transgenic plants, carbon flow through the lignin biosynthetic pathway is altered to increase the production of sinapyl alcohol relative to coniferyl alcohol (as compared to wild type plants) (Sewalt *et al.* 1997). One possible explanation for this observation is that monolignol biosynthesis from L-phenylalanine occurs by more than one route, even at the early stages of core phenylpropanoid metabolism. Sewalt et al. (1997) propose that phenylalanine may be channeled through multienzyme complexes specifically dedicated to the synthesis of either coniferyl alcohol or sinapyl alcohol precursors. Under these circumstances, if each enzyme complex contains a unique PAL isoform, and endogenous *PAL* genes are differentially suppressed by the bean *PAL* transgene, a shift in the S/G ratio could result. Creation of a series of transgenic tobacco lines, each suppressed for a single *PAL* gene family member, will be needed to test this hypothesis.

In addition to reduced lignin content, other biochemical and physiological changes were observed in PAL-suppressed plants. In particular, rutin (a flavonoid glucoside) and chlorogenic acid (a caffeic acid ester) accumulated to lower levels, presumably a pattern that reflects the responses of many of the phenylpropanoid metabolites produced in these transgenic plants (Bate *et al.* 1994). Furthermore, PAL-suppressed lines exhibited more rapid and extensive lesion development upon pathogen attack (Haher *et al.* 1994), and failed to establish systemic acquired resistance when infected with tobacco mosaic virus (Pallas *et al.* 1996). Therefore, although suppression

of PAL was able to markedly reduce lignin content, plant health was also compromised, possibly because PAL mediates the production of many important secondary metabolites essential to plant functions.

1.3.2 Cinnamate 4-Hydroxylase (C4H)

In the next step of the core phenylpropanoid pathway, cinnamate 4-hydroxylase, a cytochrome P450 monoxygenase, catalyses the hydroxylation of cinnamic acid at position 4 on the ring to form p-coumaric acid. *In planta* feeding experiments in which ^{14}C-L-phenylalanine and ^{3}H-cinnamic acid were simultaneously supplied to plants showed that C4H preferentially utilizes cinnamic acid produced by PAL, over exogenously supplied cinnamic acid (Hrazdina and Wagner 1985, Rasmussen and Dixon 1999). This has been interpreted as reflecting an association between these enzymes. C4H cDNAs have been cloned from a number of plants species, including alfalfa, Arabidopsis, Jerusalem artichoke and poplar. The artichoke C4H expressed in yeast is highly specific for cinnamic acid and does not catalyze the hydroxylation of p-coumaric acid, ferulic acid or benzoic acid (Teutsch *et al.* 1993). Unlike PAL, C4H does not appear to exist as a specific gene family although plants contain a remarkable number of related cytochrome P450 genes.

Transgenic tobacco plants containing the full length alfalfa C4H cDNA either in sense or antisense orientation were found to display changes in both lignin content and composition (Sewalt *et al.* 1997). Lines with 25-40% residual C4H activity accumulated less lignin than wild type plants and, in contrast to PAL-suppressed plants, the lignin in

C4H-suppressed lines had a decreased S/G ratio. By the same logic used to rationalize the S/G ratio changes observed in PAL-suppressed plants, Sewalt et *al.* have suggested that these results further support the hypothesis of metabolic channeling between PAL, C4H and other phenylpropanoid enzymes.

1.4 The Synthesis of Lignin Monomers

The canonical pathway for the synthesis of lignin monomers from p-coumaric acid has been defined using radiotracer and enzyme studies (Higuchi and Brown 1963, Brown 1965). In this view, p-coumaric acid is consecutively hydroxylated and O-methylated at the 3' ring position (producing ferulic acid, the coniferyl alcohol precursor) and then at the 5' ring position (producing sinapic acid, the sinapyl alcohol precursor). The free acids are then converted to CoA esters before being twice reduced, first to their respective aldehydes and then to p-coumaryl alcohol, coniferyl alcohol and sinapyl alcohol, the lignin monomers.

1.4.1 Ferulate 5-Hydroxylase (F5H)

Ferulate 5-hydroxylase's role in monolignol synthesis has been characterized in Arabidopsis fah1 mutants that are apparently unable to convert ferulic acid to 5-hydroxyferulic acid (Chapple *et al.* 1992). Plants with a mutation at the fah1 locus cannot methoxylate the 5' ring position of monolignol precursors and fail to synthesize sinapyl alcohol. As a result, the lignin in fah1 plants is entirely composed of guaiacyl subunits.

Cloning and heterologous expression of the functional fah1 cDNA has confirmed that the gene disrupted in these mutants encodes F5H. *In vitro*, F5H, a cytochrome P450 monoxygenase, catalyzes the hydroxylation of ferulic acid and coniferyl aldehyde to 5-hydroxyferulic acid and 5-hydroxyconiferaldehyde, respectively. However, F5H shows much higher activity against coniferaldehye, suggesting that this is the branchpoint in the pathway at which 5' methylation occurs (C. Chapple, personal communication). Furthermore, Arabidopsis stem protein extracts, in the presence of S-adenosyl methionine, have been shown to convert 5-hydroxyconiferaldehyde to sinapaldehyde, thereby completing the 5' methoxylation step necessary for syringyl lignin precursor synthesis. This suggests that during lignin biosynthesis 5' hydroxylation may occur predominantly at the level of the aldehydes or the alcohols.

Transgenic Arabidopsis plants overexpressing F5H in lignifying tissues produce almost entirely syringyl lignin (Meyer *et al.* 1998). Elevated levels of F5H presumably cause a depletion of the intracellular ferulic acid pool, thus impairing coniferyl alcohol synthesis. No marked changes in total lignin content have been observed in either fah1 mutants or in Arabidopsis plants overexpressing F5H.

1.4.2 O-Methyltransferases (COMT/CCoAOMT)

Caffeic acid O-methyltransferase uses S-adenosyl methionine as a methyl group donor to catalyze the conversion of caffeic acid and 5-hydoxyferulic acid to ferulic acid and sinapic acid, respectively. In angiosperms, this enzyme is considered a bifunctional O-methyltransferase since the same polypeptide methylates both caffeic acid and 5-hydroxyferulic acid *in vitro* (Meng and Campbell 1998, Inoue *et al.* 1998). COMT

generally prefers free hydroxycinnamic acid substrates rather than their CoA thioesters. One exception to this pattern is the recently described loblolly pine xylem AEOMT (Li *et al.* 1997). This enzyme catalyzes the hydroxylation of caffeic acid, caffeoyl-CoA, 5-hydroxferulic acid and 5-hydroxyferuloyl-CoA, displaying about the same relative activity against each of these substrates *in vitro*. Analysis of the AEOMT cDNA sequence indicates that AEOMT shares several sequence motifs with other members of the *COMT* gene family.

The canonical pathway clearly implicates COMT's in the synthesis of monolignols and this has been supported by evidence from both transgenic plants and naturally occurring mutants. The *brown midrib* mutation (bm3) in maize results in plants that exhibit a reddish-brown pigmentation in the leaf midrib. This phenotype is known to be associated with a reduction in lignin content and recently it was demonstrated that a retrotransposon insertion in the maize *COMT* gene is responsible for the bm3 mutant phenotype (Vignols *et al.* 1995). This knockout results in plants that are impaired in their ability to synthesize syringyl (S) type lignin but instead produce a novel lignin component derived from 5-hydroxyconiferyl alcohol.

Transgenic plants with down-regulated COMT activity have a lignin phenotype similar to bm mutants. In tobacco, antisense suppression of the OMT1 class of transferases (the class specifically associated with lignification) was successful in reducing endogenous OMT activity to less than 5% of wild-type levels in some lines (Atanassova *et al.* 1995). The syringyl lignin content in these plants was reduced and the lignin was shown to contain novel 5-hydroxyconiferyl units. Transgenic plants with low OMT activity were visually indistinguishable from control plants at all stages of

development and, unlike maize, no coloration was observed in the lignifying tissue. Similar results have been observed in transgenic poplar with downregulated OMT activity. Poplar plants with only 2% residual OMT activity produced markedly less syringyl lignin and they contained 5-hydroxyconiferyl alcohol units (Van Doorsselaere *et al.* 1995). Overall decreases in lignin content were not observed in either transgenic tobacco or poplar with reduced OMT.

The lignin composition in plants with compromised OMT activity provides valuable information about the sequence(s) of reactions in the phenylpropanoid pathway leading to monolignol synthesis. The marked reduction of syringyl lignin in these plants indicates that COMT plays an active role in methylating the 5' ring position of sinapyl alcohol precursors. Furthermore, since alfalfa and poplar *in vitro* enzymatic data shows that these COMT enzymes do not readily methylate 5-hydroxyferuloyl CoA, the 5' methylation reaction likely does not occur at the level of the CoA thioester. Taken together with the recombinant F5H enzymatic data, the results observed in COMT transgenics and mutants would be consistent with a monolignol synthesis pathway in which ferulic acid is converted directly to sinapic acid via 5-hydroxyferulic acid, or in which the 5' methoxylation reactions are carried out on the aldehyde However, experiments examining COMT activity against 5-hydroxyconiferaldehyde or 5-hydroxyconiferalcohol are scarce, which makes it difficult to resolve this question at this time.

An interesting paradox surrounds the 3' methylation of caffeic acid. If COMT is a bispecific enzyme as suggested by *in vitro* studies, how are plants with compromised COMT activity still capable of methylating coniferyl alcohol precursors? One possible

explanation is that the methylation of caffeic acid is normally catalyzed by an independent O-methyltransferase i.e. COMT does not function as a bispecific enzyme *in vivo*. If the gene encoding this putative independent OMT enzyme is not downregulated in antisense-suppressed plants, coniferyl alcohol synthesis would still proceed. A possible test of this hypothesis would be to feed transgenic/mutant plants radio-labeled caffeic acid and observe whether they produce labeled ferulic acid as efficiently as wild type plants do.

We must also consider whether an alternate pathway for 3' methylation might be utilized during monolignol biosynthesis. A caffeoyl-CoenzymeA 3-O-methyltransferase (CCoAOMT) has been described in poplar, alfalfa, tobacco and zinnia (Ye *et al.* 1994, Meng and Campbell 1998, Inoue *et al.* 1998, Martz *et al.* 1998). CCoAOMT is constitutively expressed in lignifying tissue toward the basal part of stems and catalyzes the 3' methylation of caffeoyl-CoA and the 5' methylation of 5-hydroxyferuloyl-CoA, but it displays little activity against caffeate or 5-hydroxyferulate *in vitro* (Ye and Varner 1995, Ye 1997).

However, it appears that CCoAOMT is inefficient at 5' methylation *in vivo*. This hypothesis is drawn from the observation that transgenic plants with down-regulated COMT1 activity produce less S lignin. This suggests that CCoAOMT is not capable of compensating for the block in 5' methlation in these plants either because 5-hydroxylferuloyl-CoA is not efficiently methylated *in planta* or is unavailable for methylation.

Transgenic tobacco with down-regulated CCoAOMT activity were shown to produce less lignin and their lignin contained an increased proportion of syringyl subunits

(Zhong *et al.* 1998). Although these plants had a decrease in both G and S lignin, the more dramatic reduction in G subunit production suggests that CCoAOMT is predominantly involved in the methylation of guiacyl subunits. In addition, transgenic plants with down-regulated COMT and CCoAOMT activity had an even greater decrease in lignin content associated with decreased syringyl subunit synthesis (compared to CCoAOMT down-regulated plants). These results indicate that both COMT and CCoAOMT are involved in lignin synthesis and that CCoAOMT may be more involved in G subunit synthesis while COMT may be more involved in S subunit synthesis. As the 3' methoxylation of CoA thioesters appears to occur during monolignol synthesis, the activity of CCoAOMT provides an explanation as to why plants lacking COMT activity are still capable of producing guaiacyl lignin.

1.4.3 Para-Hydroxylation (C3H)/ (CCoA3H)

A further complication in our understanding of how 3' methoxylation of p-coumaric acid occurs in plants is the failure to identify genes for any enzyme(s) responsible for hydroxylating the 3' ring position of p-coumaric acid or p-coumaryl CoA. Some researchers have suggested that 3' hydroxylation of p-coumaric acid is catalyzed by a phenolase which is capable of oxidizing a wide range of substrates (Whetton and Sederoff 1995). Coumaryl:Coenzyme A 3-Hydroxylase (CCoA3H), a zinc-dependent enzyme first identified in elicitor-treated parsley cells, catalyzes hydroxylation of the 3' ring position of both p-coumaric acid and p-coumaryl-CoA (Kneusel *et al.* 1989). Analogous enzyme activity has since been characterized in various plant cell culture systems. Inactive at normal cytoplasmic pH, parsley CCoA3H becomes functional when

the pH of the cytosol in elicitor-treated cells shifts toward this enzyme's optimum acidic pH. There is currently no data linking phenolase or CCoA3H to lignification and, as such, this step in monolignol synthesis remains the most poorly understood.

1.4.4 4-Coumarate:CoenzymeA Ligase (4CL)

4-coumarate:coenzymeA ligase catalyses the formation of CoA thioesters of cinnamic acid and hydroxycinnamic acids as part of the biosynthesis of a wide range of phenolic molecules including benzoic acid, salicylic acid, phenolic esters, flavonoids and lignin precursors. Most angiosperm 4CL enzymes, purified or heterologously expressed, accept p-coumaric acid, caffeic acid, ferulic acid and 5-hydroxferulic acid as substrates but their activity against sinapic acid is low or non-detectable. However, in a few species distinct isoforms of 4CL have been reported to possess relatively high activity against sinapic acid.

Analysis of transgenic Arabidopsis expressing a homologous *4CL* gene in antisense orientation have confirmed that this enzyme is involved in monolignol biosynthesis (Lee *et al.* 1997). The most severely suppressed plants (8% residual 4CL activity) contained less lignin (50% of wild type) and their lignin had a reduced S/G ratio resulting from decreased coniferyl alcohol synthesis. This suggests, in keeping with the *in vitro* enzymatic data, that 4CL is not involved in the activation of sinapic acid since suppression of 4CL did not reduce sinapyl alcohol synthesis. Transgenic tobacco plants with suppressed 4CL activity were found to have lignin phenotypes consistent with the results in Arabidopsis (Kajita *et al.* 1997). 4CL-suppressed lines with only 1% residual

4CL activity produced only 65-78% as much lignin as control plants. Furthermore, in several antisense 4CL tobacco plants, decreases in the S/G ratio were observed.

The data from F5H and COMT studies indicates that 5' methoxylation may occur either at the level of the free acid or at the level of the aldehyde. But, if 4CL does not esterify sinapic acid, how is this acid activated during sinapyl alcohol synthesis? This conflicting data may suggest that 5' methoxylation occurs only at the level of the aldehyde. If this proposed pathway is correct, it would not be necessary for 4CL to esterify sinapic acid en route to sinapyl alcohol synthesis, thus explaining 4CL's poor *in vitro* activity towards sinapic acid.

However, a hypothesis proposing that 5' methoxylation occurs at the level of the aldehyde still leaves an unexplained observation in transgenic plants. If coniferaldehyde is the branch point for sinapyl alcohol synthesis, why do 4CL down-regulated plants have reduced levels of syringyl lignin when the 4CL enzyme should be responsible for the synthesis of a feruloyl-CoA pool from which coniferaldehyde, and ultimately both coniferyl and sinapyl alcohol, are derived? One plausible explanation is that divergent 4CL isoforms, not suppressed by the 4CL cDNA's used in these studies, are active in distinct pathways leading to either coniferyl alcohol synthesis or syringyl alcohol synthesis (i.e. metabolite channeling). This hypothesis seems reasonable in those plant species (Arabidopsis, tobacco) which have been shown to contain multiple *4CL* genes encoding divergent 4CL isoforms.

1.4.5 Cinnamoyl:Coenzyme A Reductase (CCR)

Cinnamoyl:coenzymeA reductase catalyzes the conversion of hydroxycinnamoyl-CoA esters to their corresponding aldehydes. CCR has been purified and partially characterized from eucalyptus, soybean cultures, spruce cambial sap and poplar xylem. The purified protein has activity against its proposed natural substrates (p-coumaryl-CoA, feruloyl-CoA and sinapyl-CoA) and shows no particular preference among these *in vitro*. The incorporation of 5-hydroxyconiferyl alcohol-based subunits into the lignin of COMT mutant and transgenic plants indicates that CCR may also reduce 5-hydroxyferuloyl CoA *in vivo*.

CCR has been purified to apparent homogeneity from differentiating xylem of *Eucalyptus gunnii* and a reverse genetics approach was used to clone the corresponding full length cDNA (Goffner *et al.* 1994). Sequence analysis of the 1008bp open reading frame indicates that CCR is closely related to dihydroflavanol-4-reductase (DFR), the first enzyme in the anthocyanin biosynthesis branch pathway. Common intron locations in the two genes supports the hypothesis that both CCR and DFR have arisen from a common ancestor. Only one copy of the *CCR* gene appears to be present in the eucalypt genome and this gene is highly expressed in lignifying tissue.

It has recently been show in transgenic tobacco that reduction of CCR activity leads to a reduction in lignin deposition (Piquemal *et al.* 1998). Antisense suppression with a homologous CCR cDNA produced plants with 25% residual CCR activity and 25% less lignin than control plants. Furthermore, CCR-suppressed plants displayed an orange-brown coloration of the xylem cell walls, had an increased S/G ratio in their lignin and incorporated unusual phenolics into their cell wall. The lines with the most

severely suppressed CCR activity (30% residual activity) were impaired in development (stunting, abnormal leaf morphology and collapsed vessels). Overall, the observations made of transgenic CCR-suppressed tobacco support the role assigned to CCR in the canonical monolignol synthesis pathway.

1.4.6 Cinnamyl Alcohol Dehydrogenase (CAD)

The last step of the monolignol synthesis pathway, the reduction of cinnamaldehydes to cinnamyl alcohols, is catalyzed by cinnamyl alcohol dehydrogenase (CAD). CAD cDNAs have been isolated from several plant species including tobacco, eucalypt, poplar, alfalfa, pine and spruce. Extensive homology exists among those CAD cDNAs associated with lignification (80-90% at the amino acid level). It is not clear, however, whether a single CAD enzyme, or several, are normally involved in monolignol synthesis. In gymnosperms, a single CAD isoform has been identified. This isoform has activity against p-coumaraldehyde and coniferaldehyde but not sinapaldehyde, consistent with the predominance of guiacyl lignin in gymnospems. In several angiosperm species, however, multiple CAD isoforms exist that differ in their substrate specificities, molecular weight and amino acid sequence. This suggests that, in these species, CAD may be a control point at which lignin composition is modulated although data confirming each isoform's participation in lignification has yet to be reported.

Sequence analysis of CAD shows that it belongs to the zinc containing long-chain alcohol dehydrogenase family. Furthermore, molecular modeling, based upon the crystallographic coordinates of other enzymes in this family, has made it possible to tentatively identify functionally important residues in the CAD enzyme that might serve

as appropriate sites for engineering changes in substrate specificity. Interestingly, a unique CAD cDNA recently cloned from eucalyptus has been shown to have extensive sequence similarities to *CCR*, suggesting that this form of CAD and CCR may have arisen from a common ancestor (Goffner *et al.* 1998).

There also exists a subset of 'CAD-like' sequences identified in plants that display approximately 50% similarity to the CADs associated with lignification. The most interesting gene among these is *ELI3* that is produced by plants in response to pathogen attack. In arabidopsis, *ELI3* transcripts rapidly accumulate around fungal infection sites (Somssich *et al.* 1996). The expression of *ELI3* is also up-regulated in cell cultures treated with fungal elicitors and this transcriptional response requires the presence of the *RPM1* resistance locus. Expression of the arabidopsis ELI3 cDNA in bacteria indicates that, *in vitro*, this enzyme prefers benzaldehyde substrates and has only low activity towards of cinnamaldehydes. ELI3 may therefore participate in metabolite synthesis for cell wall reinforcement, a known defense response. Convincing evidence associating this enzyme with classical lignification has not yet been presented. Analysis of transgenic plants with suppressed ELI3 activity should help clarify this enzyme's role in plant phenolic metabolism.

The creation of genetically modified plants has confirmed the role of lignification-specific CADs in monolignol synthesis. A 93% reduction in CAD activity was achieved in tobacco by antisense suppression, using a homologous CAD cDNA (Halpin *et al.* 1994). The lignin content in these plants was unchanged, but there was a dramatic increase in aldehyde-based subunits and a modest switch towards increased syringyl subunit production. Similarly, in transgenic poplar with 20% residual CAD

activity, the amount of lignin produced was not affected but an increase in aldehyde-based subunits was observed (Baucher *et al.* 1996). In this case, changes in the S/G ratio were not detected. These data support a role for CAD in the reduction of cinammaldehydes to cinnamyl alcohols and also highlight the metabolic plasticity of lignin biosynthesis.

Naturally-occurring CAD mutants in pine and maize have provided additional information about the contribution of CAD to lignification. Pine mutants with altered lignin phenotypes have been shown to have severely reduced CAD activity resulting from a mutation in the *cad* gene (MacKay *et al.* 1997, Ralph *et al.* 1997). The mutant *CAD* allele results in severely decreased *CAD* gene expression and, in homozygotes, modified lignin content and composition. Similar to antisense CAD plants, the pine mutant was reported to have an increase in aldehyde-based subunits, from 7% in control plants to 15% in mutant plants. However, the pine mutant, unlike transgenic CAD-suppressed plants, also showed a decrease in total lignin content. In maize, another of the brown midrib mutants, *bm1*, has been mapped to the maize CAD locus (Halpin *et al.* 1998). Plants homozygous for the *bm1* mutation have increased aldehyde subunits in their lignin and accumulate less lignin. Reductions in lignin content may reflect the severely reduced CAD activity in these mutant plants (1% residual activity), a phenotype that has not yet been obtained with antisense suppression. Alternatively, there may be underlying differences between lignin biosynthesis in these species (maize and pine) and dicot species.

Accompanying altered lignin phenotypes in transgenic and mutant CAD-reduced plants is a reddish-brown coloration of the xylem tissue. This coloration may result from

increased coniferaldehyde content in the wood, since peroxidases convert coniferaldehyde to a red dehydropolymerisate (DHP) *in vitro*.

Despite the increased level of aldehyde subunits in the lignin of severely CAD-suppressed plants, cinnamyl alcohols are still the primary constituents of lignin in these plants. This suggests that other CAD enzymes, not suppressed in these transgenic or mutant plants, must be continuing to contribute to lignin synthesis. In support of this hypothesis, poplar trees grown in the presence of a chemical CAD enzyme inhibitor have been shown to contain 45% less lignin in their woody tissue than control plants (Wallace and Fry 1994).

Data from transgenic plants and genetic mutants has drawn into question the accuracy of the canonical monolignol synthesis pathway. The traditional version of the pathway has been suggested to include parallel pathways of hydroxylation and O-methylation that operate at the level of the CoA thioesters and at the level of the aldehydes. However, there exist results which even such a partially revised model is incapable of explaining, suggesting the need to consider novel routes in lignin monomer synthesis.

1.5 Monolignol Storage and Transport

The enzymes of the phenylpropanoid pathway appear to be located within the cytoplasm of the cell while lignin synthesis occurs at the cell wall. Since the monolignols (cinnamyl alcohols) are highly reactive and toxic to the cell, it is believed that they are transported to the site of lignin polymerization as monolignol glucosides (cinnamyl

alcohol-β-D-glucosides). The most convincing evidence supporting this theory comes from radiotracer studies in which it was shown that the aglycone of labeled monolignol glucosides is incorporated at the appropriate time and place into newly synthesized lignin (Terashima 1989). This was observed in several angiosperm and gymnosperm species of the genera *Pinus*, *Ginkgo*, *Magnolia*, *Syringa*, *Populus* and *Oryza*. The cinnamyl alcohols may also be stored as monolignol glucosides. Gymnosperms species, particularly the conifers, and some angiosperm species within the Magnoleaceae and Oleaceae families, accumulate large quantities of monolignol glucosides in their cambial sap during the early spring, prior to lignin deposition (Terazawa *et al.* 1984).

Glucosylation of the cinnamyl alcohols requires the activity of a glucosyl transferase. Since first being identified in cell cultures of 'Paul's Scarlet' rose, enzyme activity catalyzing the addition of a β-linked glucose to the *para*-hydroxyl group of the cinnamoyl alcohols has been reported in several angiosperm and gymnosperm species (Ibrahim and Grisebach 1976). Partial purification from rose cells and *Forsythia* stems has allowed further characterization of this enzyme activity (Ibrahim 1977). The 52kd glucosyl transferase from these species has a pH optimum of 7.5-8.0, requires UDPG as a glucose donor and efficiently glucosylates coniferyl alcohol and sinapyl alcohol to form coniferin and syringin (monolignol glucosides). A 50kd enzyme from spruce (*Picea abies)*, purified to apparent homogeneity, has similar properties to the angiosperm glucosyl transferases, but prefers coniferyl alcohol as a substrate over sinapyl alcohol (Schmid and Grisebach 1982). In addition, analogous enzyme activity in lignifying tissue of *Pinus banksiana* and *P. strobus* has recently been reported (Savidge RA,

personal communication). The pine enzyme has a strong substrate preference for coniferyl and sinapyl alcohols, is developmentally regulated and peaks in activity during lignin synthesis. Cloning of the genes encoding this apparently widely occurring enzyme could provide valuable tools for investigating the potential role of UDPG:glucosyl transferases in lignification.

The subsequent release of cinnamyl alcohols from their glucosides requires the activity of a monolignol glucoside-accommodating β-glucosidase. The first report of an enzyme from angiosperms with high activity for the hydrolysis of coniferin came from crude cell wall preparations of chick pea (*Cicer arientinum* L.) cell suspension cultures (Hosel *et al.* 1978). The purified enzyme had a molecular weight of 110kD (two subunits, 43kD and 63kD) and was most active at pH 5. This enzyme preferred coniferin among accepted β-glucoside substrates and has since been identified in the stems and roots of chick pea seedlings, tissues known to undergo developmentally-regulated lignification (Surholt and Hosel,1981).

Another angiosperm β-glucosidase capable of hydrolyzing coniferin and syringin has been isolated from cell cultures, roots and hypocotyls of soybean (*Glycine max*) (Hosel and Todenhagen 1980). This protein was most active at pH 4.6-6 and had a predicted molecular mass of 45kD. Glucono-1,5-lactone was an effective inhibitor of this soybean enzyme, but was a relatively poor inhibitor of the chick pea β-glucosidase. Furthermore, polyclonal antibodies raised against the chickpea β-glucosidase did not cross-react with the purified soybean protein. Differences in size, substrate preference and epitope display suggest that the two β-glucosidases isolated from chick pea and

soybean are structurally unrelated. Which enzyme, if either, is involved in lignification is unknown.

Lignification-associated β-glucosidases have also been purified from gymnosperm species. Spruce seedlings and the differentiating xylem of lodgepole pine (*Pinus contorta*) both contain a 60kd enzyme that, at its pH optimum of 5.5, readily hydrolyzes syringin and coniferin, with a marked preference for the latter (Marcinowski and Grisebach 1978, Dharmawardhana *et al.* 1995). The main reported difference between these two enzymes is that the spruce enzyme appeared to be cell wall-bound while the pine enzyme was cytosolic. The purification of a distinct coniferin β-glucosidase from the cambial sap of *Pinus banksiana* has also recently been reported (Leinhos *et al.* 1994). This enzyme had a much larger molecular mass than the other conifer β-glucosidases and its N-terminal amino acid sequence showed no homology to known glucosidases. Common characteristics (molecular weight, pH optimum, glucono-1,5-lactone inhibition, pI values) among the purified β-glucosidases from lodgepole pine, spruce and soybean suggest these proteins may be encoded by homologous genes, or at least by genes within the same family.

Using N-terminal amino acid sequence information, the lodgepole pine β-glucosidase cDNA corresponding to the purified enzyme has been isolated (Dharmawardhana *et al.*, 1999). Expression of this cDNA in *E. coli* yielded a recombinant enzyme with an identical substrate usage profile as the native protein. Furthermore, sequence analysis indicates extensive identity between the pine cDNA and other, previously identified, plant β-glucosidases. Experiments with transgenic trees

suppressed for coniferin β-glucosidase activity are currently underway to evaluate this enzyme's role in lignification.

1.6 Lignin Polymerization

Models of lignin synthesis generally depict cell wall-bound enzymes oxidizing secreted monolignols, with the resulting phenoxy free radicals polymerizing into the growing lignin matrix. Peroxidases, laccases and catechol oxidase have all been implicated in the oxidation of lignin precursors (Savidge *et al.* 1998, Dean *et al.* 1998, Barcelo *et al.* 1998). These enzymes are found, in part, in the cell wall, they are often temporally and spatially associated with lignification, and they catalyze the synthesis of dehydropolymerisates from monolignols *in vitro*. Peroxidases, H_2O_2-dependent hemoproteins, have long been implicated in lignin synthesis. The overexpression of a specific acidic peroxidase in transgenic tobacco increased lignin deposition, but suppression of the same gene had no effect on lignification (Whetten *et al.* 1998). Laccases, O_2-dependent copper proteins, had been discounted as participants in lignification until observations were made that plants grown in a copper-deficient environment contained less lignin (Downes *et al.* 1991). However, transgenic poplar with 10% residual laccase activity have been created but these showed no changes in lignification (A. Boudet, personal communication). Catechol oxidases, also O_2-dependent copper proteins, are able to oxidize lignin substrates but differ from laccases in catalytic activity, amino acid composition, protein size and antigenic display. All three

enzymes have been implicated in monolignol polymerization however, currently, their modes of action remain speculative.

Lignin polymerization has been proposed to occur via random phenoxyl radical polymerization after it was observed that plant peroxidases could form dihydropolymerisates (DHP's) with coniferyl alcohol and H_2O_2 *in vitro* (Lewis and Davin, 1998). However, the recent discovery of a 'dirigent' protein in *Forsythia sp.* which directs the stereospecific synthesis of lignans suggests that the structure of lignin could potentially be the result of stereospecific, protein-mediated coupling (Davin LB *et al.*, 1997). Lignans are a structurally diverse class of natural products that arise from the oxidative coupling of monolignols. They are widely distributed throughout the plant kingdom and are present in various plant tissues, including, roots, rhizomes, wood, stems, leaves, fruits and seeds. As single catalysts, oxidative enzymes (peroxidases, laccases, etc.) produce coniferyl alcohol radicals that randomly couple *in vitro* to form (±) pinoresinols and (±) erythro/threo guaiacyl-glycerol-8-O-4'-coniferyl alcohol ethers. However, when the 78kDa *Forsythia.* dirigent protein, which has no enzymatic activity itself, is combined with an oxidase, coniferyl alcohol is stereospecifically dimerized to form only (+) pinoresinol. A similar dirigent protein from *Linum usitatissimum* guides coniferyl alcohol coupling towards the production of only (-) pinoresinol (Lewis and Davin, 1998). These dirigent proteins apparently bind specific substrates (in this case only coniferyl alcohol radical, not p-coumaryl or sinapyl alcohols) and spatially orientate them in a precise manner. The dirigent protein is encoded by a multigene family and putative homologues have been identified in several plant species, including western red

cedar, western hemlock and loblolly pine. This work shows that stereospecific coupling of plant phenols is enzymatically catalyzed and the discovery of dirigent protein mediatating lignan synthesis may have implications for lignin polymerization.

1.7 Cell Wall Synthesis

Cell wall synthesis proceeds in distinct, highly organized stages (Lewis and Davin, 1998). Following cell division, golgi vesicles fuse at the plasma membrane to form a cell plate comprised primarily of pectin. Next, deposition of the primary wall commences with hemicellulose and cellulose microfibrils being randomly deposited around the cell plate, usually accompanying cell growth. This layer also contains non-cellulosic polysaccharides and proteins of unknown function (e.g. extensin). Synthesis of the following layers (S1, S2 and S3) involves further cellulose deposition; however these microfibrils are now strictly arranged into a twisted honeycomb structure with 15°-30° angles of deviation. The mechanisms underlying cell wall synthesis and deposition are unclear. It is believed that cellulose microfibrils are synthesized in the cell wall by membrane embedded, multisubunit cellulose synthase complexes. Microtubules are thought to direct these complexes two-dimensionally through the plasma membrane, their patterns of movement determining the orientation of cellulose microfibril deposition. Other cell wall polysaccharides seem to be synthesized in the cytoplasm and transported to the outer membrane in vesicles derived from the golgi and/ or endoplasmic reticulum.

Lignin synthesis, which always follows cellulose deposition, starts with p-coumaryl alcohol-based lignin being laid down in the cell corners and middle lamella

region (Jung and Allen 1995). In gymnosperm species, subsequent lignin deposition in the primary and secondary cell walls proceeds with an increase in the incorporation of guiacyl lignin. This process appears to be similar in angiosperm species except that significant syringyl lignin deposition also occurs in the secondary cell wall. This non-uniform, yet predictable, pattern of lignin deposition could be directed by molecules or glycoproteins (maybe dirigent proteins) embedded in the cell wall.

Although this basic model highlights the general processes of lignin deposition in cell wall synthesis, subtle variations in the monomer composition of lignin cannot be ignored. Angiosperm lignin consists of both coniferyl and sinapyl alcohol subunits whereas gymnosperm lignin contains almost entirely coniferyl alcohol subunits. Furthermore, grasses incorporate the three types of monomers into their lignin in approximately equal quantities. Variation in monomer composition exists not only between species, but also between tissues within a species. For example, in Arabidopsis, G-subunits predominate in the cells of the vascular bundle, while adjacent, heavily lignified sclerenchyma cells contain mostly S subunits (Chapple *et al.* 1996). The evolution of distinct lignin types suggests that the monomer composition influences lignin function and that underlying biochemical mechanisms are likely responsible for orchestrating complex patterns of lignin deposition.

Monolignol glucosides may be transported to the site of lignin synthesis in membrane-bound vesicles before being deglycosylated and polymerized into the growing lignin matrix. This hypothesis is supported by the observation that a steady stream of radioactive vesicle traffic is observed in the lignifying tracheids of some angiosperm species fed radio-labeled L-phenylalanine or trans-cinnamic acid. Conifers, which appear

to store monolignol glucosides in their vacuole, may also utilize an intermediate transport step in which the glucosides are targeted for the vacuole, stored, and then rerouted to the plasma membrane when needed for lignin synthesis. Angiosperms, which generally don't accumulate monolignol glucosides, may have developed an alternate mechanism for managing the supply of lignin substrates which does not require glucoside storage.

Upon arrival at the cell wall, the monolignol glucosides must be hydrolyzed by a β-glucosidase, since the oxidative polymerization enzymes do not appear to accept monolignol glucosides as substrates. The coniferin β-glucosidase from pine contains a 23 amino acid membrane target sequence that suggests this enzyme may be released into the cell wall space (by fusion with the plasma membrane) where it could liberate the monolignols for lignin synthesis (Dharmawardhana *et al.* 1995). Alternately, vesicles containing β-glucosidases could fuse with monolignol-loaded vesicles before the latter reached the cell wall.

1.8 How Close Are We To Solving Industry's Problems?

As mentioned, the goal of lignin modification is to increase the amount of accessible cellulose within plant material. This can be achieved in two ways; 1) by reducing the quantity of lignin in the cell wall or 2) by altering the properties of lignin in the cell wall. Although still in the early research stages, transgenic and mutant plants potentially capable of addressing the needs of industry have already been developed.

Reducing the total lignin content in a plant presents the greatest challenge at this time. Transgenic plants suppressed for PAL, C4H or 4CL activity generally produce less

lignin (10-50% of wild type levels). However, changes in lignin content are accompanied by pleitropic effects that can compromise plant health. Since these three enzymes are involved in the synthesis of many phenylpropanoid metabolites in addition to lignin, they may be inappropriate targets for lignin modification. Downregulation of CCR also results in plants with reduced lignin levels. CCR is believed to be dedicated to the synthesis of monolignols and, in tobacco (B3H line), a 70% reduction in CCR activity results in a 40% drop in lignin levels. However, B3H plants are reduced in size, have abnormal leaf morphology and collapsed vessels. This is in sharp contrast to PAL suppressed tobacco with 10% residual lignin levels that show none of the phenotypic changes observed in B3H plants. Monocot brown midrib mutants (bm) with reduced lignin content, apparently healthy, have been used for many years as a feed resource for ruminants. Thus, aside from natural bm mutants, an effective strategy for reducing lignin levels has still to be devised.

Modulating lignin composition, particularly the S/G ratio, has been more successful. Modifying F5H activity in Arabidopsis has produced plants with entirely guiacyl lignin, by suppression, or entirely syringyl lignin, by overexpression. Although untested, it would appear that control of F5H activity with appropriate promoters / transgene combinations could be used to produce plants with any S/G ratio that one desires. However, preliminary trials using mutant (*fah1*) and wild type Arabidopsis as a feed resource for ruminants reveal no appreciable differences in digestibility between these plants (Jung and Allen, 1995). It remains to be tested whether the results obtained in arabidopsis hold true for other, agronomically important, species.

Plants with unique lignin phenotypes can also be valuable for industry. Trees with reduced CAD activity incorporate a greater proportion of aldehyde-based subunits into their lignin than do wild type plants. In poplar, kraft pulping data from three year old transgenic plants indicates that aldehyde-based lignin is easier to extract than wild type lignin (Baucher *et al.* 1996). Trees 7-10 years of age, the normal age for harvesting, remain to be tested. Based on these data, naturally occurring *cad* mutants in pine and maize (bm1) could also make valuable contributions to their respective industries. Although in its early stages, the goal of modifying lignin for industrial uses is being realized with more valuable contributions likely still to come.

1.9 Research Objectives and Approaches

In this study, biochemical and molecular approaches were used to investigate how angiosperms might store and/or transport monolignols, a mechanism about which little is known. Currently, we understand that 1) poplar incorporate radio-labeled monolignol glucosides into newly synthesized lignin at the appropriate time and place and 2) that VRA-G, a chromogenic substrate analogue of coniferin, is hydrolyzed by a β-glucosidase(s) in the differentiating xylem of poplar (Dharmarwardhana, Ph.D. Thesis). These limited data suggest monolignol glucosides may be used in the transport/storage process, as has been proposed for gymnosperm species, but controversial evidence also exists. For instance, 1) neither coniferin nor syringin have been identified in *Populus*, 2) neither a glucosyl transferase nor a β-glucosidase have been purified from *Populus* and 3) *Populus* genes encoding these putative enzymes have not been detected. This study

attempted to address some of these questions using tools generated from the cloning of a coniferin β-glucosidase from pine. In particular, family 1 β-glucosidase sequence data was used to screen the poplar genome for homologous β-glucosidase's by degenerate PCR. Genes identified in this fashion were then assessed to determine if their expression pattern and enzymatic activity were consistent with a role in monolignol glucoside usage during lignin synthesis. In the future, genes encoding proteins for the transport and storage of monolignols in poplar will hopefully be effective targets for modifying lignin content.

Poplar is quickly emerging as a model woody plant system. Poplar has a relatively small genome size (680MB), can easily be sexually and vegetatively propagated, hybrids reach harvesting maturity in as little as 7-10 years, and it is amenable to Agrobacterium transformation. F1 hybrids derived from a cross between *Populus trichocarpa* and *P. deltoides* are of commercial interest (Bradshaw and Stettler 1995), and a three-generation pedigree is now available for some lines. In addition, the poplar genome is being mapped using a combination of RFLP, sequence tagged sites, RAPD and AFLP markers (Bradshaw et al. 1994). A collection of poplar EST clones from wood-forming tissue is also available (Sterky et al. 1998). These resources and attributes make poplar an appropriate system for studying lignification.

Chapter 2
Materials and Methods

2.1 Plant Growth Conditions and Tissue Sampling

Poplar tissues used in these studies were sampled from two different sources. Young leaves (harvested just prior to flushing) and young stems (approximately 30 cm in length, 1cm in diameter, with light green bark and two leaves of similar surface area) were harvested from a clonal population of two-year-old H-11 hybrid poplar (*Populus deltoides* x *Populus trichocarpa*). These had been propagated vegetatively from cut branches and maintained under greenhouse conditions [23°C, 8 h dark/16 h light (~120μ $Es^{-1}m^{-1}$)regime]. Young leaves were used for DNA and RNA isolation and young stems were used for radio-labeled L-phenylalanine feeding experiments. Differentiating xylem was harvested in the early spring (middle May 1995 and 1996) by peeling the bark and scraping the most recently developed xylem (0.2-0.5mm) from 5 year-old H-11 trees grown on the University of British Columbia campus. This tissue was used for protein

extractions and RNA isolation. All harvested tissue was immediately frozen in liquid nitrogen and stored at -80°C until needed.

2.2 Radiotracer Feeding Study

Cut stems were allowed to absorb a radio-labeled phenylalanine feeding solution (0.5uM L-phenylalanine, 5uCi L-[U-^{14}C]-phenylalanine) by transpiration for either 30 min. or 2 hours, in the presence and absence of 5mM glucono-1,5-lactone. Following feeding and removal of the leaves, the stems were cut into small pieces and extracted in boiling 70% ethanol for 30min. Boiled stem samples were stored overnight at room temperature in the 70% ethanol solution.

After soaking overnight, the samples were filtered, the residual ethanol was evaporated, and 5mg coniferin was added to each sample (Figure 2.1). This crude extract was partitioned three times with equal volumes of hexane. Following hexane extraction, the aqueous layer was further extracted three times with equal volumes of ethyl acetate. The resulting aqueous layer was then concentrated for preparative thin layer chromatography (TLC). During liquid–liquid extractions, the presence of coniferin was monitored by silica gel TLC, using a developing solvent of acetone, ethyl acetate, and water (10:10:1).

2.2.1 Thin Layer Chromatography (TLC)

Two TLC separations were used to purify the recovered coniferin. In the first round of TLC, 20cm X 20cm fluorescent silica gel plates (Whatman Inc.) were developed

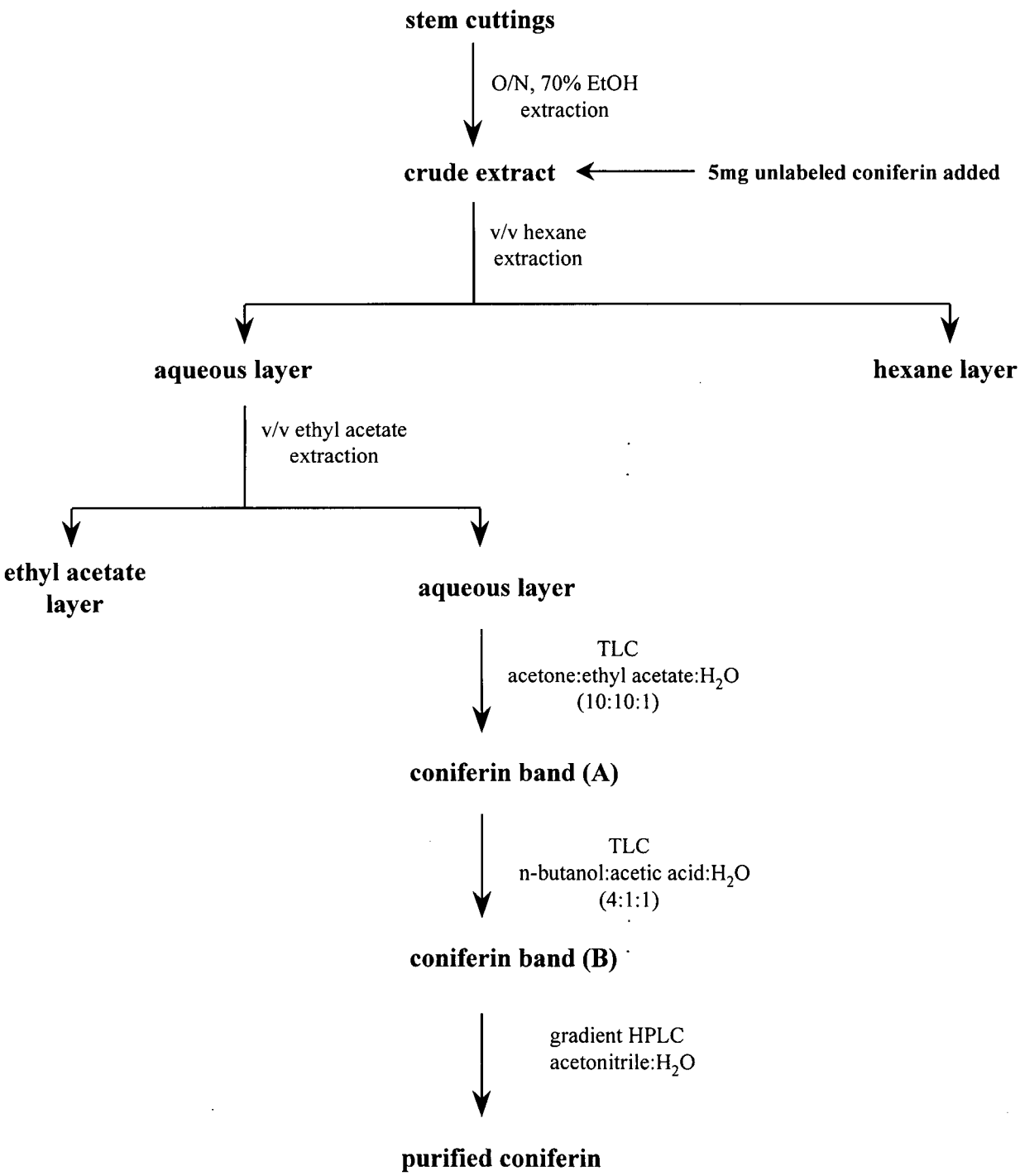


Figure 2.1 Isolation procedure for the purification of coniferin from *Populus* stems fed radio-labelled L-[U-^{14}C]-phenylalanine.

in acetone, ethyl acetate and water (10:10:1). The coniferin band in the sample was visualized relative to a coniferin standard under short wave UV light and the appropriate zone of the silica plate was scraped and eluted overnight in 95% ethanol. The eluate was concentrated and subjected to a second round of TLC in a developing solvent of n-butanol, acetic acid, and water (4:1:1). Again, the coniferin band was visualized under short wave UV light and the appropriate zone of the silica gel was scraped and eluted overnight in 95% ethanol. The ethanol eluate was then evaporated to dryness and the residue was resuspended in distilled water in preparation for high-pressure liquid chromatography (HPLC).

2.2.2 High Pressure Liquid Chromatography (HPLC)

The water soluble extracts were fractionated by gradient HPLC using a Nucleosil C-18 reversed-phase column (5μm, Alltech). Solvent A: 2% (v/v) acetic acid; Solvent B: 2% (v/v) acetic acid in acetonitrile; gradient conditions: 8.5% B for 0.56 min., 8.5-11.5% B over 30 min., 11.5-80% B over 0.01 min, 80% B over 5 min., 80%-8.5% B over 0.01 min., 8.5% B over 5 min.; solvent flow:1ml/min. The eluate was monitored at 258nm (the absorbance maximum for coniferin) and the peak corresponding to coniferin was collected. Pooled coniferin fractions from repeated sample injections were then concentrated by evaporation, resuspended in distilled water and re-analyzed by HPLC to check for coniferin purity. The amount of incorporated radioactivity (DPM) in the coniferin fractions was determined by liquid-scintillation counting.

2.3 PCR Amplification of β-glucosidase Sequences from *Populus* with Degenerate Primers BGL1 and BGL2

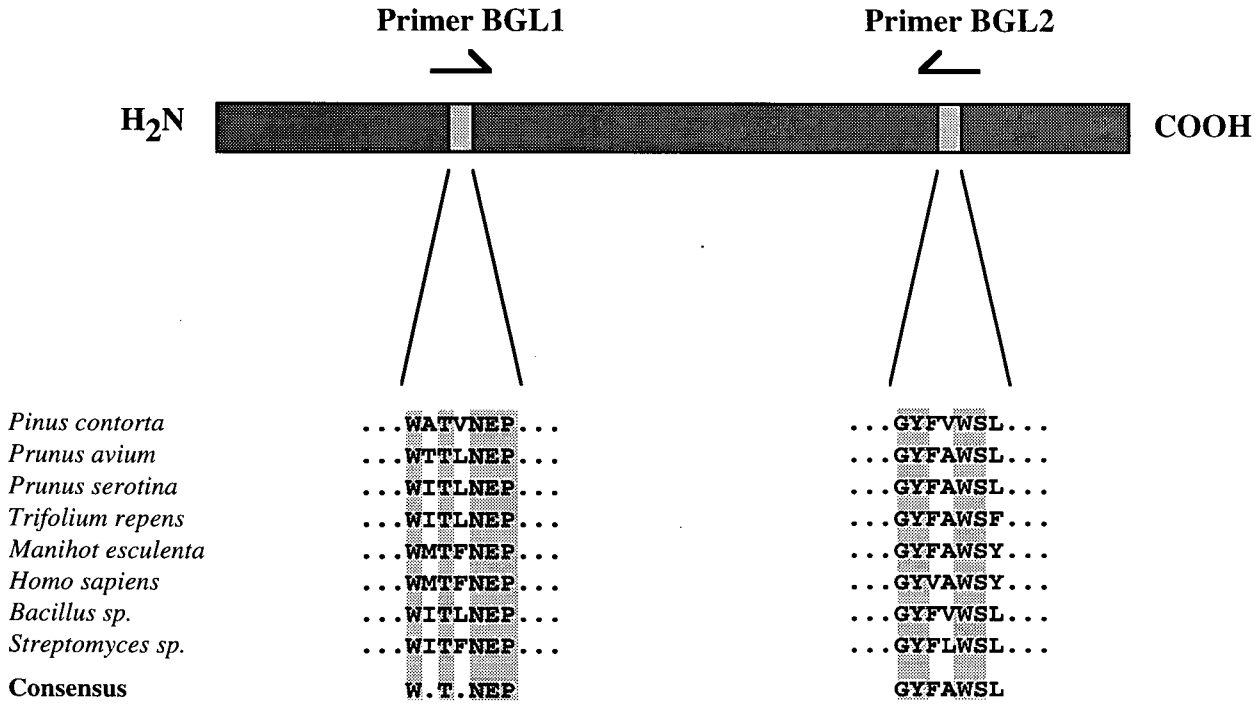
Computer-assisted amino acid sequence alignments were used to design degenerate oligonucleotide primers targeted to conserved motifs among *family 1* β-glucosidases of plant, mammal and bacterial origin (Figure 2.2). Primers BGL1 and BGL2 contained the following sequences: primer BGL1, 5'-TGG A(C/T)I ACI TTI AA(C/T) GA(A/G) CC-3'; primer BGL2, 5'-AG IGA CCA I(T/G)C (A/G)AA (A/G)TA IAT-3'. Inosine was incorporated into nucleotide positions of complete degeneracy.

Genomic DNA for PCR amplification was isolated from young leaves based on the CTAB method of Doyle and Doyle (1990). Control PCR reactions contained 5ng of the pine coniferin β-glucosidase cDNA (Dharmawardhana *et al.*, 1999) instead of poplar genomic DNA.

PCR was carried out in 100μl volumes containing the following reaction components: 1X PCR buffer (Appligene), 1.5mM $MgCl_2$, 25μM each dNTP, 200pmols BGL1, 200pmols BGL2, 1μg poplar genomic DNA template or 5μl of young leaf library template and 3 units of *Taq* DNA polymerase (Appligene). The thermal cycling regime was as follows: 94°C for 3min.; 30 cycles (94°C/1min., 60°C/1min., 72°C/2min.); 72° C/10min. final extension.

2.4 PCR Screen of λ-ZAP *Populus* H-11 Young Leaf Library

Two methods were used to screen a poplar H-11 λ-ZAP young leaf library (kindly donated by Carl Douglas) for the presence of POP1, POP2 and POP3 cDNA's. First,



Primer BGL1 - W(I/T/M)T(F/L)NEP **Primer BGL2 - GYF(A/V)WSL**

Figure 2.2 Degenerate primer design strategy. Schematic representation of *family 1* β-glucosidases highlighting the conserved amino acid motifs to which primers BGL1 and BGL2 were targeted. Grey boxes indicate perfectly conserved residues.

gene-specific primers POP1a (5'-TTC ACC GGT GAT GGC TAC GA-3'), POP2a (5'-CTG ATC GAT AGG ATT TAA TAG- 3') and POP3a (5'-GCA GCA ATT CAC GGT TAC AGA-3') were designed to a variable exonic region among POP1, POP2 and POP3. Each primer was used in combination with vector primer T7 (5'-GTA ATA CGA CTC ACT ATA GGG C-3') in an attempt to amplify either POP1, POP2 or POP3 from the library. PCR amplification was carried out in 100µl volumes containing the following reaction components: 1X PCR buffer (Appligene), 1.5mM $MgCl_2$, 25µM each dNTP, 200pmols POP1a primer, 200pmols T7 primer, 5µl young leaf library template and 3 units *Taq* DNA polymerase (Appligene). The young leaf library template was heated to 70°C for 5min. before the remaining reaction components were added. The thermal cycling regime was as follows: 94°C for 3min.; 30 cycles (94°C/1min., 60°C/1min., 72° C/2min.); 72°C/10min. final extension.

Degenerate primers BGL1 and BGL2 were also used to screen for *family 1* β-glucosidases in the poplar H-11 λ-ZAP young leaf library. All PCR conditions were identical to those used for amplification of genomic DNA with primers BGL1 and BGL2 except that 5µl of library was used as a template instead of poplar genomic DNA.

PCR products from amplification of the young leaf library with primers BGL1and BGL2 were cloned (see cloning section of Materials and Methods) into T-tailed Bluescript S/K+. Transformed colonies were screened by PCR with primers BGL1 and BGL2 for the presence of the desired insert and eighteen positive colonies were chosen for southern analysis. The chosen amplification products were size fractionated in a 1% TAE-agarose gel and blotted onto a Zeta-Probe GT Genomic Membrane. The blots were

hybridized overnight at 65° in 0.5M Na_2HPO_4, pH 7.2 and 7% SDS with ^{32}P-labelled POP1 cDNA (complete coding region) prepared by nick translation (Sambrook *et al.* 1989). The labeled probe was estimated to have a specific activity of $6x10^7$ CPM/μg. Following hydridization, blots were washed at high stringency (0.2X SSC, 0.1% SDS, 65 °C) and radioactive bands were detected using a phosphor imager (Molecular Dynamics).

2.5 PCR Amplification of the 5' and 3' Ends of the POP1 cDNA

Internal *POP1* gene-specific primers (POP1a and POP1b) were designed for amplification of the 5' and 3' ends of the POP1 cDNA from a λ-ZAP poplar H-11 young leaf library (Figure 2.3).

The 3' end of the POP1 cDNA was amplified using primer POP1a (5'-TTC ACC GGT GAT GGC TAC GA-3') and vector primer T7 (5'-GTA ATA CGA CTC ACT ATA GGG C-3'). PCR amplification was carried out in 100μl volumes containing the following reaction components: 1X PCR buffer (Appligene), 1.5mM $MgCl_2$, 25μM each dNTP, 200pmols POP1a primer, 200pmols T7 primer, 5μl young leaf library template and 3 units *Taq* DNA polymerase (Appligene). The young leaf library template was heated to 70°C for 5min. before the remaining reaction components were added. The thermal cycling regime was as follows: 94°C for 3min.; 30 cycles (94°C/1min., 60° C/1min., 72°C/2min.); 72°C/10min. final extension.

The 5' end of the POP1 cDNA was amplified using primer POP1b (5'-GTA GCC ATC ACC GGT GAA-3') and vector primer T3 (5'-AAT TAA CCC TCA CTA AAG GG-3'). PCR reaction conditions were the same as those used for 3' POP1 amplification

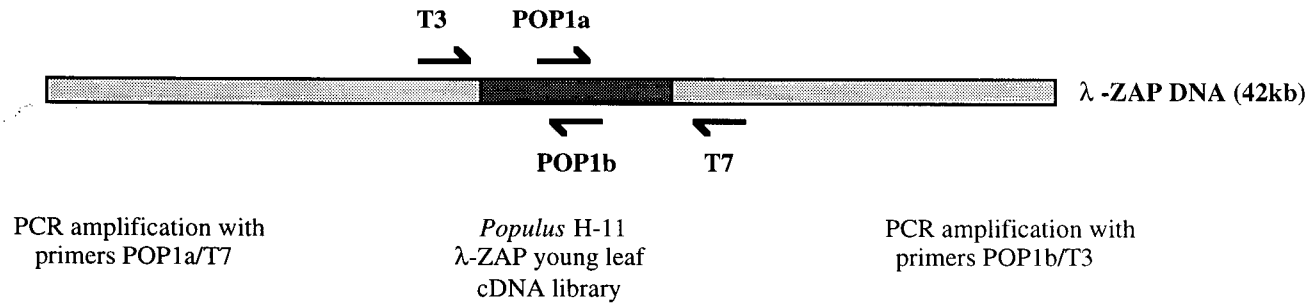


Figure 2.3 POP1 cDNA amplification strategy. The strategy used for PCR amplification of the 5' and 3' ends of the POP1 cDNA from a *Populus* young leaf cDNA library.

except that an annealing temperature of 54°C was used instead of 60°C.

2.6 Cloning of PCR Products

All PCR products generated during these studies were analyzed by TAE (tris-acetate) agarose gel electrophoresis (Sambrook *et al.* 1989). Desired products were cloned directly into *Eco*RV digested T-tail Bluescript II SK+ vectors according to the T/A cloning protocol of Holton and Graham (Holton and Graham 1991). Ligation mixes were transformed by heat shock into chemically competent DH5α *E.coli* cells (Sambrook *et al.* 1989) and transformants were selected on LB-AMP (100mg/ml) agar plates containing X-gal (5-bromo-4-chloro-3-indoyl-β-D-galactoside, 25μl aliquots of a 40mg/ml solution spread on top of the agar layer) and grown at 37°. White colonies, chosen from a blue/white β-galactosidase clone insertion assay (Sambrook *et al.* 1989), were grown in LB broth overnight at 37° and plasmids were isolated by the miniprep protocol of Zhou et *al.* (Zhou *et al.* 1990). For each clone generated, the isolated plasmids were digested with restriction enzymes that were appropriate to confirm that the desired PCR products had been ligated into the vector in the correct orientation.

2.7 Sequencing and Data Analysis

Plasmid DNA was prepared from clones chosen for sequencing, using the mini-alkaline/PEG precipitation method (Sambrook *et al.* 1989). Primer-directed sequencing of the cloned PCR products was carried out using ABI Amplitaq dye termination cycle sequencing chemistry according to the manufacturer's recommendations and analyzed on

an ABI 373 DNA sequencer (Nucleic Acid and Protein Services Unit, Biotechnology Lab., U.B.C.).

BLAST sequence searches were conducted in the EMBL, PDB, SWISS-PROT, PIR and Genbank databases (Altschul *et al.* 1990). Sequence analysis to detect predicted intron/exon locations was carried out with the ASLP program (prediction of splice sites in plant DNA sequences) (Solovyev *et al.* 1994). Signal peptide predictions were carried out with Signal P V1.1 (Neilson *et al.* 1997). Sequence alignments were performed with CLUSTAL W (Thompson *et al.* 1994).

2.8 Northern Blot

Total RNA was isolated from young leaves, the developing xylem of trees harvested in May of 1996 and the developing xylem of trees harvested in May of 1995, using Trizol (Gibco, BRL) and quantified spectrophotometrically. Total RNA (10μg) was size-fractionated on 1% MOPS-agarose denaturing formaldehyde gels and blotted onto Zeta-Probe GT Genomic Membranes. The blots were hybridized overnight at 65° in 0.5M Na_2HPO_4, pH 7.2 and 7% SDS with ^{32}P-labelled POP1 cDNA (complete coding region) prepared by nick translation (Sambrook *et al.* 1989). The labeled probe was estimated to have a specific activity of $1.8x10^8$ CPM/μg. Following hydridization, blots were washed at high stringency (0.2X SSC, 0.1% SDS, 65°C) and radioactive bands were detected using a phosphor imager (Molecular Dynamics).

2.9 RT PCR

SuperscriptII reverse transcriptase (Gibco, BRL) was used to prepare single-stranded DNA templates for PCR amplification from total RNA extracts of poplar xylem, according to the manufacturer's recommendations. Reverse transcription was carried out using 5μg of RNA template and a poly $T_{(18)}$ primer. A 1μl aliquot of the first strand synthesis reaction was used as a template for PCR amplification of poplar *family 1* β-glucosidase fragments. The reaction conditions and thermal cycling regime were identical to those used for amplification of the *POP1*, *POP2* and *POP3* gene fragments from poplar genomic DNA except that reactions were carried out in 25μl volumes instead of 100μl volumes.

2.10 Recombinant POP1 Expression in Baculovirus

The POP1 coding region (without predicted signal peptide) was amplified by PCR from a λ-ZAP young leaf cDNA library using primers CODf (5'-CAG CTC GAG ATG CCA AGA GGT A-3') and CODr (5'-TCG GAT CCA TCA AGC TAT ATT CAA TCG-3'). *Pst*I and *Bam*HI restriction sites were introduced into primers CODf and CODr to facilitate directional cloning of the PCR product into the pVL1392 baculovirus expression vector. PCR was carried out in 100μl volumes containing the following reaction components: 1X PCR buffer (Appligene), 1.5mM $MgCl_2$, 25μM each dNTP, 200pmols CODf, 200pmols CODr, 5μl young leaf library template and 3 units *Taq* DNA polymerase (Appligene). The aliquot of young leaf library was heated to 70°C for 5min. before the remaining reaction components were added. The thermal cycling regime was as follows: 94°C for 3min.; 30 cycles (94°C/1min., 55°C/1min., 72°C/2min.); 72°

C/10min. final extension.

Following amplification, the *Pst*I/*Bam*HI-digested, gel-purified (Geneclean protocol, BIO 101), PCR product was ligated into the *Pst*I/*Bam*HI digested pVL1392 baculovirus expression vector (Pharmingen). Ligation mixes were transformed by heat shock into chemically competent DH5α *E.coli* cells (Sambrook *et al.* 1989) and transformants were selected on LB-AMP (100mg/ml) agar plates containing X-gal (25µl aliquots of a 40mg/ml solution spread on top of the agar layer) and grown at 37°. Selected colonies were grown in LB broth overnight at 37° and plasmids were isolated by the mini-alkaline/PEG precipitation method (Sambrook *et al.* 1989). Screening of the plasmids by digestion with *Xho*I and *Bam*HI was used to confirm that the desired PCR product had been cloned.

Recombinant POP1 protein was kindly expressed in Sf-9 insect cells by Jennifer Norton using the pVL1392:POP1 construct prepared as described above, and crude cell extracts were supplied to me for SDS-Page and Western blot analysis.

2.11 SDS-Page and Western Blot

Proteins were extracted from developing xylem scrapings by first grinding the tissue with liquid nitrogen in a mortar and pestle. The resulting powder was stirred in cold buffer (50mM MES, pH 6.0 with polyvinylpolypyrrolidone at a ratio of 0.1g PVPP/ g tissue) for 30 min. Two mls buffer was used per gram tissue (fresh weight). The crude extract was then filtered through two layers of Miracloth and clarified by centrifugation at

10,000x*g* for 30 min. The final supernatant was used for SDS-PAGE and western blot analysis.

Recombinant baculovirus-infected insect cells ($1x10^8$) were harvested 48hrs. post-infection and pelleted at 1000x*g* for 10 min. at 4°C. The cells were twice washed in 1X PBS and re-centrifuged as above. This pellet was resuspended in extraction buffer (50mM MES, pH5.8, 3mM DTT, 10% glycerol) and homogenized in a glass Dounce homogenizer on ice, using 20 strokes. This crude extract was centrifuged at 10,000x*g* for 30 min. at 4°C and the supernatant was used for western blot analysis.

Protein concentrations in all crude extracts were calculated by the method of Bradford (Bradford 1976) using the Bio-Rad dye-binding reagent. Bovine serum albumin was used as a protein standard.

SDS-Page gel and western blot analysis was performed as described by Sambrook *et al.* (Sambrook *et al.*1989). Plant (25μg) or insect cell (5μg) proteins were electrophoresed in 10% SDS-polyacrylamide separating gels and either stained with Coomassie Brilliant Blue R 250, or blotted on PVDF membranes (Westran) according to Bio-Rad electrophoresis transfer protocols. Following blocking for 1 h with 5% (w/v) non-fat powdered milk in TBST (tris-buffered saline-tween 20), the blots were reacted with a 1:3300 dilution of a rabbit antiserum raised against the denatured, recombinant pine coniferin β-glucosidase for 1 h at RT. After washing in TBST, the membrane was reacted with a 1:3300 dilution of goat anti-rabbit IgG conjugated to alkaline phosphatase and immuno-reactive proteins were visualized using the BCIP/NBT detection system.

2.12 Recombinant POP1 Enzyme Assays

Assay mixtures containing the crude enzyme preparation (10-50μl) and substrate (1-4mM, 4-nitrophenyl β-glucoside or coniferin) in 0.2M MES, pH 5.5 (final volume 150 μl) were incubated at 37° for 30min. to 1h. The reaction was stopped by basification of the assay mixture with an equal volume of 0.5M CAPS, pH10.5 and the activity was determined by measuring the absorbance of the released aglycone. For quantitative calculations, the following wavelengths and extinction coefficients (mM^{-1} x cm^{-1}) were used: coniferyl alcohol, 325nm, ε=7.0; 4-nitrophenol, 400nm, ε=19.3.

Chapter 3
Radiotracer Feeding Study

3.1 Introduction

In recent years, research on the phenylpropanoid pathway has given us a fundamental understanding of the biochemistry underlying the conversion of L-phenylalanine to coniferyl, sinapyl and p-coumaryl alcohols (Whetten and Sederoff, 1995). However, subsequent steps in lignin synthesis, particularly the modification of the monolignols for transport or storage, require clarification. In several conifers, it has been proposed that the lignin precursors are stored or transported in glucosylated form until their release at the site of lignification (Figure 3.1) (Terazawa *et al.,* 1984; Freudenberg and Harkin, 1963). However, it is not clear whether the use of monolignol glucosides in lignification is ubiquitous in the plant kingdom.

The occurrence of monolignol glucosides within lignifying tissue has provided preliminary evidence that several plant species may utilize monolignol glucosides during lignin synthesis. In many gymnosperm families, for instance the Taxaceae and Pinaceae, coniferin, syringin and p-coumaryl 4-O-β-D-glucoside are readily detectable by chromatographic analysis of differentiating xylem tissue (Freudenberg and Harkin, 1963). However, in contrast, very few non-coniferous species have been shown to contain monolignol glucosides. Of nineteen dicot tree species studied by Terazawa, none, except members of the Oleacea and Magnoleacea, contained 4-O-glucosylated forms of the

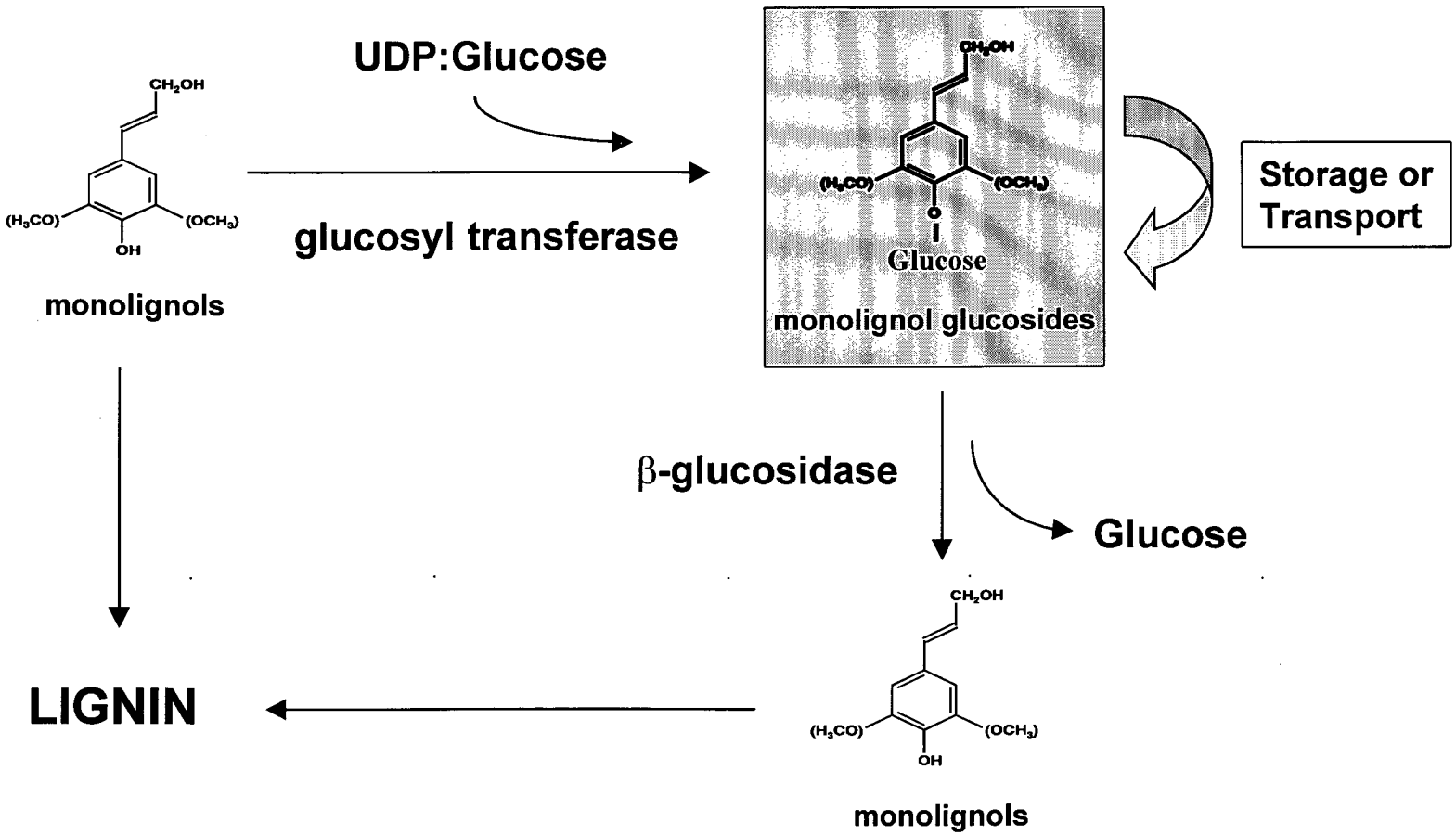


Fig. 3.1 Proposed alternate pathway for lignin synthesis occuring via monolignol glucosides.

lignin precursors (Terazawa *et al.,* 1984). These results could lead to criticism of proposed mechanisms of angiosperm wood lignification involving the monolignol glucosides.

In this study I supplied young poplar stems with L-[U-^{14}C]-phenylalanine by transpiration feeding and assayed for labeled coniferin content in soluble stem extracts. In previous studies, the use of radio-labeled phenylpropanoid pathway metabolites in feeding experiments has led to the detection of lignin synthesis intermediates that had been difficult to identify with standard chromatographic techniques (Freudenberg and Harkin, 1963). Here I used a similar strategy to assay lignifying poplar stems for the presence of monolignol glucosides that may previously have been overlooked with less sensitive assay techniques.

3.2 Radiotracer Feeding Study

The accumulation of radio-labeled coniferin was analyzed in young poplar shoots fed L-[U-^{14}C]-phenylalanine by transpiration for either 30 min. or 2 h. Pure coniferin was added to crude poplar stem extracts to facilitate the isolation of any endogenous, radio-labeled coniferin by a series of purification steps which included liquid-liquid extraction, thin layer chromatography and high pressure liquid chromatography. Several techniques were used to confirm that the compound purified from the spiked crude poplar stem extracts was authentic coniferin. During HPLC analysis, the isolated sample had the same retention time (RT=13.54 min.) as a coniferin standard (RT=13.5 min.). Furthermore, the isolated compound had the same mobility profile as a coniferin standard when analyzed by TLC in two different developing solvents. Finally, the isolated

compound had a UV absorption spectrum that was identical to coniferin.

Through the purification sequence, the specific radioactivity of the coniferin-containing fractions continued to decrease until, in apparently purified coniferin samples, no radioactivity counts above background levels were detected. The same outcome was observed when poplar stems were allowed to absorb the L-[U-^{14}C]-phenylalanine feeding solution for 30 min. or 2 hours.

A second set of feeding experiments was conducted in which young poplar shoots were administered a feeding solution containing both L-[U-^{14}C]-phenylalanine and glucono-1,5-lactone, a potent *in vitro* inhibitor of the soybean and pine coniferin β-glucosidase enzymes (Dharmawardhana *et al.* 1995, Hosel and Todenhagen 1980). As in the experiments conducted with a feeding solution containing only L-[U-^{14}C]-phenylalanine, shoots were administered the L-[U-^{14}C]-phenylalanine/glucono-1,5-lactone feeding solution, by transpiration, for either 30 min. or 2 h and the accumulation of radio-labeled coniferin in the stem extracts of these plants was analyzed. Once again, endogenous, radio-labeled coniferin could not be detected in coniferin samples purified from the stem extracts. During all of the above outlined experiments between 30% and 36% of the pure coniferin added to the crude poplar stem extracts was recovered after the HPLC purification step.

3.3 Discussion

Attempts to trap label from L-[U-^{14}C]-phenylalanine in coniferin did not provide any evidence that young poplar stems undergoing lignification produce this monolignol

glucoside. These results are in keeping with the findings of Terazawa et al. who failed to establish the presence of monolignol glucosides in any of seventeen angiosperm families outside the Magnoleaceae and Oleaceae (Terazawa *et al.*1984). That earlier report included the analysis of poplar (*Populus maximowiczii*) cambium extracts by gas chromatography and thin layer chromatography, but neither analytical approach produced evidence that supported the existence of coniferin or syringin in this species. Similarly, wheat plants supplied with ferulic acid-α-^{14}C by transpiration feeding did not accumulate radio-labeled coniferin although ^{14}C label was found in coniferyl alcohol and in lignin isolated from the same plants (Higuchi and Brown 1963). Taken together, these results suggest that monolignol glucosides may only be utilized in the lignin biosynthesis pathway of gymnosperm species, and perhaps, in the two most evolutionarily primitive angiosperm families, the Magnoleaceae and Oleaceae. However, this hypothesis is derived largely from negative results and the controversy surrounding the potential role of monolignol glucosides in the lignification of non-coniferous species will likely continue until a definitive pathway to lignin formation is identified in angiosperm species.

It must be borne in mind that shortcomings in experimental design and limitations associated with the experimental techniques used in this study may have made it difficult to detect coniferin in the lignifying stems of poplar. To detect labeled coniferin in this study, the poplar stems had to be actively synthesizing detectable amounts of radio-labeled coniferin during the time period when the feeding was conducted. Several factors could potentially have affected this process, including the photoperiod, light level or

temperature, the elapsed time between administering the L-[U-^{14}C]-phenylalanine and sampling the stems for coniferin content, or the developmental stage of the stem being used. An attempt was made to increase the coniferin pool in the cut poplar stems by co-feeding L-[U-^{14}C]-phenylalanine with glucono-1,5-lactone, an established inhibitor of the β-glucosidase responsible for the breakdown of coniferin in soybean and pine (Dharmawardhana *et al.* 1995, Hosel and Todenhagen 1980). However, this approach did not result in the recovery of ^{14}C-labelled coniferin. Another potential experimental problem is that L-[U-^{14}C]-phenylalanine absorbed by a plant stem through the transpiration stream may not reach the appropriate cells in the plant, or even the compartment within a cell, to allow it to contribute significantly to coniferin synthesis. However, taken at face value, my data indicates that coniferin was not being synthesized in the lignifying poplar stems.

If this is correct, one possible explanation may be that there has been an evolution of the molecular machinery responsible for lignin synthesis that has, in most angiosperm species, reduced, or even eliminated, the involvement of monolignol glucosides in lignification. As mentioned previously, only gymnosperm species and members of the two most primitive angiosperm families have been shown to accumulate monolignol glucosides. In *Pinus banksiana*, large amounts of coniferin accumulate in the symplast of differentiating xylem cells during the months before lignification begins (Leinhos and Savidge 1993). It has been proposed that these monolignol glucosides could serve as storage forms of the lignin precursors, thus providing a readily available source of monolignols in the early summer during active lignin deposition. However, it is

conceivable that, in an environment with limited 'molecular' resources, a lignification system that utilizes monolignol glucoside as storage molecules may have been replaced by a more efficient mechanism of lignification. Plants that store coniferin invest an enormous amount of 'molecular equity' in the synthesis of monolignol glucosides before they can be used and before the plant's seasonal growth needs are known. If abundant molecular building blocks are available then this synthesis strategy may be appropriate. But in an environment where resources are limited it may be more efficient to synthesize the monolignols needed for lignin synthesis *de novo*, thus eliminating the storage of monolignol glucosides, while freeing valuable 'molecular equity' to sustain other biochemical needs. This could explain why more recently evolved angiosperm species do not appear to contain monolignol glucosides.

It has also been proposed that the monolignols may be transported from the cytoplasm (or vacuole) to the cell wall in the form of monolignol glucosides since the free alcohols are relatively cytotoxic and easily oxidized. In this role, the monolignol glucosides may have such a rapid turnover rate that they never accumulate to appreciable levels within the cell and are therefore undetected with our experimental approaches. In *Linum flavum*, both coniferin β-glucosidase and coniferin can be detected and a reciprocal relationship exists between β-glucosidase activity and coniferin levels (Uden *et al.* 1990). In *L. flavum* suspension cultures, where lignification does not occur, coniferin accumulates to high levels, while levels of coniferyl alcohol and 5-methoxypodophyllotoxin (a cytotoxic lignan thought to be formed from the oxidative dimerization of phenylpropane units) are low and little β-glucosidase activity is

measurable. In the leaves of *L. flavum* plantlets, however, where lignification does occur, the β-glucosidase activity is thirty-fold greater than it is in suspension cultured cells, the level of coniferyl alcohol and 5-methoxypodophyllotoxin are high and coniferin is undetectable. Although these data do not provide direct evidence supporting a role for coniferin in the lignification of *Linum flavum*, it does suggest that the breakdown of coniferin by a β-glucosidase can serve to reduce coniferin levels *in planta*.

It is also possible that, as angiosperms evolved, modifications to the organization of the phenylpropanoid pathway may have made it unnecessary for cells to transport lignin monomers as monolignol glucosides. Hrazdina *et al.* have proposed that in *Hippeastrum* petals, buckwheat hypocotyls and cabbage seedlings, the synthesis of phenylpropanoids and flavonoids takes place via complexes of consecutively assembled enzymes associated with membranes (Wagner and Hrazdina 1984, Hrazdina and Wagner 1985). In gel filtration studies, a significant portion of PAL and C4H activity was shown to be associated with membrane fractions containing the endoplasmic reticulum marker enzyme, NADH-cytochrome c reductase. Furthermore, membrane-associated PAL and C4H activities were not eliminated by trypsin digestion, consistent with a model in which these enzymes were protected within a membrane compartment that was inaccessible to the trypsin. Finally, gently prepared PAL-containing and C4H-containing membrane fractions were shown to preferentially utilize [^{3}H]-phenylalanine over [^{14}C]-cinnamic acid in the production of *p*-coumarate, a characteristic of consecutively assembled enzyme complexes where the product of the first enzyme is released into the vicinity of the active center of the next enzyme in the pathway. These data suggest that phenylpropanoid

enzyme complexes could reside in the endoplasmic reticulum or could be targeted to secreted membrane compartments that originate in the endoplasmic reticulum and pass through the golgi en route to the cell membrane. Synthesis of lignin precursors via enzyme complexes contained within secreted membrane compartments could isolate toxic phenylpropanoid metabolites from the rest of the cell while simultaneously providing a means of transporting the monolignols to the cell wall for lignin synthesis.

The aggregation of phenylpropanoid pathway enzymes into multi-enzyme complexes within more evolved angiosperm species would not be unprecedented. It has been shown that multi-enzyme complexes become more advanced and involve more enzymes in a given pathway when they are compared between evolutionarily distant organisms (Srivastava and Bernhard, 1986). For instance, six individual independent enzymes carry out fatty acid synthesis in prokaryotes. In yeast, several of the enzymes in this pathway form a complex during fatty acid synthesis. Finally, in mammals, all of the enzymes involved in the synthesis of fatty acids have coalesced to form a single enzymatic entity that is elegantly regulated as one enzymatic machine. Through all of these structural modifications the essential functionality of this pathway has remained unchanged. Perhaps, as plants evolved, they began to shuttle phenylpropanoid metabolites through multienzyme complexes to avoid releasing into the cytosol potentially damaging molecules such as the monolignols. In the process, the need for glycosylation as a protective mechanism may have been lost.

In considering how angiosperms may have adapted to synthesize lignin without monolignol glucosides there are many scenarios that can be proposed, but none have been rigorously proven. Mutant analysis and the use of transgenic technology in angiosperm

plant model systems such as Arabidopsis will likely be needed to clarify the true path of lignin monomer processing in this taxon.

Chapter 4
Polymerase Chain Reaction Study of Family 1 β-glucosidases from *Populus*

4.1 Introduction

If monolignol glucosides are used as storage or transport molecules in *Populus* the enzymatic hydrolysis of coniferin to release coniferyl alcohol for lignin synthesis would require the activity of a β-glucosidase. Histochemical staining of young poplar stem x-sections incubated with a chromogenic substrate analogue of coniferin has demonstrated glucosidase activity in the differentiating xylem (Dharmawardhana, Ph.D. thesis). Furthermore, the zone of β-glucosidase activity in these stems was shown to coincide with the presence of COMT protein, and with guaiacol peroxidase activity, both activities associated with lignification. Terashima et al. have also demonstrated the poplar trees fed ^{14}C-radio-labeled coniferin are able to incorporate the ^{14}C -label into newly synthesized lignin (Terashima and Fukushima 1989). These results suggest that a β-glucosidase capable of hydrolyzing coniferin is temporally and spatially associated with lignification in the developing xylem of *Populus*.

However, preliminary attempts to identify β-glucosidase activity in protein extracts from the developing xylem of *Populus* H-11 trees were unsuccessful (data not shown; Dharmawardhana, personal communication). Therefore, I used degenerate

polymerase chain reaction (PCR) as an alternate approach to search for genes encoding a coniferin β-glucosidase in poplar. Sequence data available from the pine coniferin β-glucosidase cDNA and other cloned plant β-glucosidases allowed the identification of conserved regions among these proteins that are expected to be present within poplar β-glucosidases. After designing degenerate primers targeted to these conserved regions, PCR was used to amplify partial β-glucosidase DNA fragments from the *Populus* genome. We hoped that PCR using degenerate primers, a quick and relatively easy approach to identifying conserved genes in distant species, would enable us to identify a homologue of the pine β-glucosidase in *Populus*.

4.2 PCR Amplification of β-glucosidase Sequences from *Populus* with Degenerate Primers BGL1 and BGL2

The products generated by PCR amplification from either poplar genomic DNA or the pine coniferin β-glucosidase cDNA clone using *family 1* β-glucosidase-specific degenerate primers BGL1 and BGL2 are shown in Figure 4.1. A 780 bp band was amplified using primer BGL1, primer BGL2 and the pine coniferin β-glucosidase cDNA clone as a template. This amplification product was the size expected based on the sequence of the pine coniferin β-glucosidase clone (Dharmawardhana *et al.*, in press). PCR controls contained all of the reaction components with the following exceptions; no template, poplar genomic DNA with primer BGL1 only, and poplar genomic DNA with primer BGL2 only. These reactions did not yield any appreciable PCR products. The PCR products amplified from poplar genomic DNA by primers BGL1 and BGL2 are

indicated in Figure 4.1. The three major bands observed in this lane had approximate sizes of 1.3kb, 1.4kb and 1.7kb. Since all genomic DNA sequences that are available for plant family 1 β-glucosidases contain several introns within the region spanned by primer BGL1 and primer BGL2, these 1.3kb, 1.4kb and 1.7kb PCR products were considered reasonable candidates for partial family 1 β-glucosidase fragments.

The combined PCR fragments amplified from poplar genomic DNA with primers BGL1 and BGL2 were "shotgun cloned" into T-tailed Bluescript SK+ and recombinant plasmids were grouped into seven distinct classes (P1, P2, P3, P4, P5, P6, P7) by restriction digest analysis. One representative clone from each of the seven classes was sequenced and a BLAST-directed database search with the sequence data classified four of the clones (P1, P3, P6, and P7) as putative *family 1* β-glucosidase fragments. Clone P2 contained a nucleotide fragment with identity to plant chloroplast ribosome sequences and clones P4 and P5 contained sequences that did not show any homology to other database sequences.

Further sequencing and computer analysis indicated that clones P3 and P7 were identical in length (1406bps), shared extensive sequence homology (97%) and had nucleotide differences that occurred primarily within intronic regions or produced cryptic changes in the amino acid sequence of their derived polypeptides. These data suggested that P3 and P7 could be homologous fragments amplified from alleles of the same gene. Since they encode the same polypeptide fragments, both sequences were referred to as *POP3* for the remainder of the study. This assumption was made only to simplify subsequent amino acid sequence data analysis, recognizing the possibility that clone P3

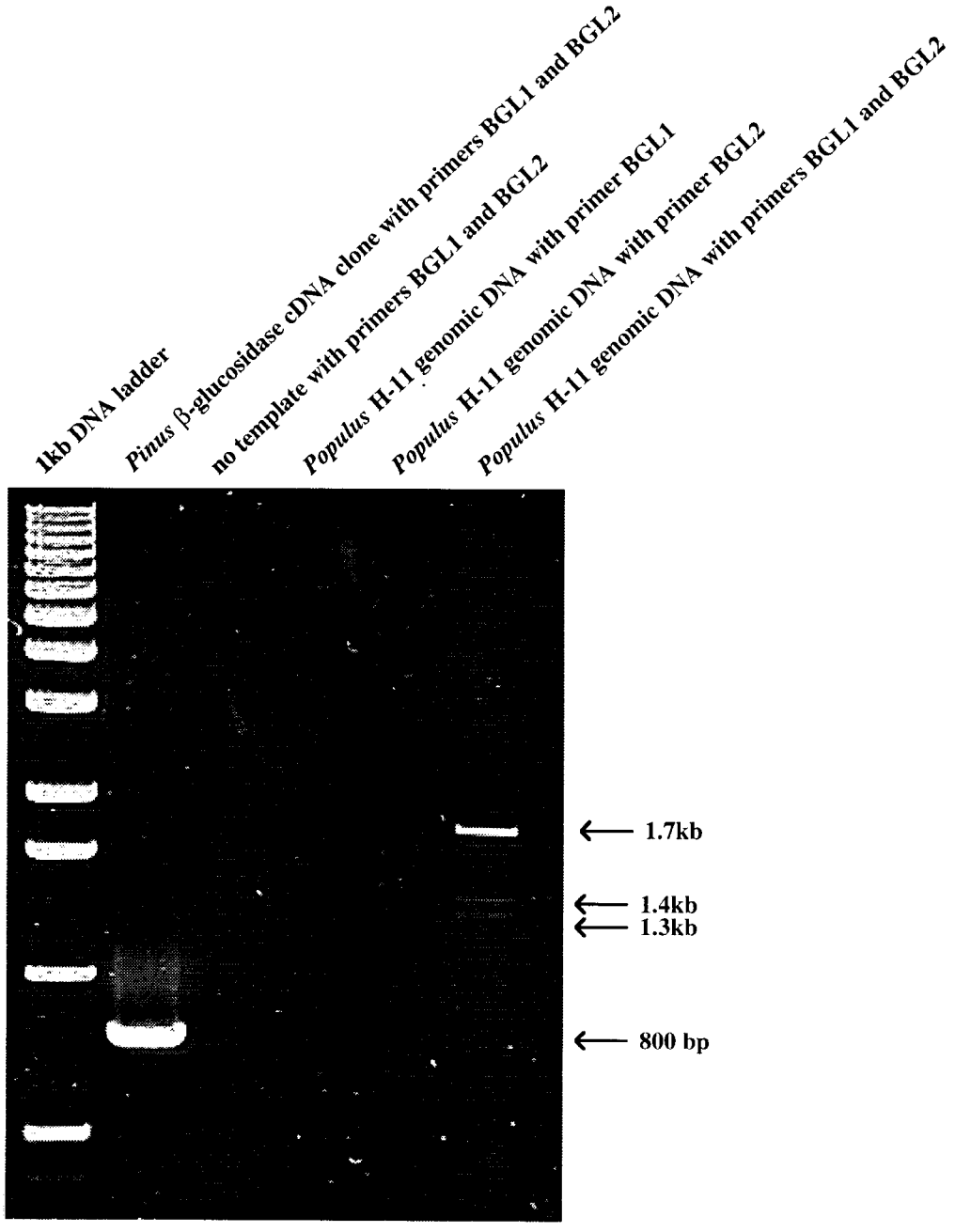


Figure 4.1 Degenerate PCR amplification. PCR amplification of *Populus* genomic DNA and *Pinus contorta* coniferin β-glucosidase cDNA clone using primers BGL1 and BGL2.

and clone P7 may actually have been amplified from separate genes that were differentially regulated or encoded catalytically distinct full-length proteins.

Clone P6, herein referred to as *POP2*, was 1320bps in length and shared 61% nucleotide sequence identity with *POP3*. Clone P1, heron referred to as *POP1*, was 1700bps in length, and shared 40% nucleotide sequence identity with *POP2* and 41% nucleotide sequence identity with *POP3* (Table 4.1). All nucleotide comparisons were made with the exonic regions only. The sequence of the *POP1* coding region was obtained during the cDNA cloning discussed in the next section, not from the *POP1* genomic clone isolated here (only the first 500bps 3' of primer BGL1 were sequenced in the genomic *POP1* clone whereas the *POP2* and *POP3* genomic clones were sequenced in their entirety). The five introns found in both *POP2* and *POP3* were located at the same positions in their respective genes, but, when analogous introns were compared, the nucleotide sequences of the introns varied greatly and some of the introns found in *POP3* were slightly larger than were their counterparts in *POP2*. The introns at the 5' end of *POP1* were found at different locations in the β-glucosidase sequence than were the introns at the 5'end of *POP2* and *POP3*. Since only the 5' end of the *POP1* genomic clone was sequenced, further information regarding the size and location of introns in this gene was not available. However, the sequence data from *POP2*, *POP3* and the 5' end of *POP1* suggests that the gene from which *POP1* was amplified may have diverged from the *POP2* and *POP3* genes (or their common ancestor) earlier than *POP2* and *POP3* genes diverged from each other.

By analyzing intron-exon borders within the genomic sequences of *POP2* and

	POP1	POP2	POP3
POP1	100%	-	-
POP2	40%	100%	-
POP3	41%	61%	100%

Table 4.1. Comparison of POP1, POP2 and POP3 nucleotide sequences over the region spanned by primers BGL1 and BGL2 (exonic regions only).

POP3 and by comparing translations of *POP1*, *POP2* and *POP3* with other plant β-glucosidase protein sequences, the partial polypeptides encoded by the PCR amplified β-glucosidase fragments were deduced. POP1, POP2 and POP3 shared extensive amino acid sequence identity with *family 1* β-glucosidases of plant, animal and bacterial origin (Table 4.2) and a multiple sequence alignment of the β-glucosidases listed in Table 4.2 illustrates the peptide regions and individual amino acid residues that are strictly conserved within this family of proteins (Figure 4.2). This alignment highlights conserved features of *family 1* β-glucosidases that are also found within the POP1, POP2 and POP3 sequences, such as, the 'NEP' and 'ENG' active site motifs, and other specific amino acid residues that are invariant among *family 1* β-glucosidase sequences. Based on the criteria of 1) sequence identity, 2) active site motif conservation and 3) other specific conserved residues, *POP1*, *POP2* and *POP3* were classified as fragments derived from three independent poplar *family 1* β-glucosidase genes.

4.3 PCR Screen of a λ-ZAP *Populus* H-11 Young Leaf Library

To address the question whether any of the enzymes encoded by the *POP1*, *POP2* or *POP3* genes were involved in the hydrolysis of monolignol glucosides during lignification, a poplar H-11 young leaf library (tissue known to undergo vascular lignification) was screened by PCR for the presence of *POP1*, *POP2* or *POP3* transcripts. Primers were designed to target a unique exonic region within each of the three β-glucosidase clones and used in combination with primer T7 (sequence found within the λ

	POP1	POP2	POP3
Pinus contorta	44%	42%	43%
Prunus avium	51%	39%	41%
Prunus serotina (1)	50%	40%	42%
Prunus serotina (2)	51%	39%	39%
Trifolium repens (1)	53%	41%	40%
Trifolium repens (2)	47%	37%	35%
Costus speciosus	51%	39%	40%
Manihot esculenta (1)	50%	33%	34%
Manihot esculenta (2)	52%	33%	34%
Manihot esculenta (3)	51%	36%	35%
Oryza sativa	44%	41%	40%
Hordeum vulgare	45%	43%	43%
Brassica napus	41%	32%	32%
Arabidopsis thaliana	39%	32%	33%
Homo sapiens	29%	27%	26%
Bacillus sp. (1)	31%	28%	28%
Bacillus sp. (2)	26%	27%	31%
Streptomyces sp.	23%	26%	28%

Table 4.2. Comparison of POP1, POP2 and POP3 predicted amino acid sequences to other family 1 β-glucosidases over the region spanned by primers BGL1 and BGL2 (amino acid identity).

Figure 4.2 Amino acid sequence alignment of *family 1* β-glucosidases in the region spanned by degenerate primers BGL1 and BGL2.

Bacillus sp. (1), Q03506; *Bacillus sp.* (2), Q08636; *Streptomyces sp.*, S45675; *Homo sapien*, P09848; *Populus* H-11 (2), POP2; *Populus* H-11 (3), POP3; *Pinus contorta*, AF072736; *Prunus avium*, U39228; *Prunus serotina* (1), U50201; *Prunus serotina* (2), U26025; *Trifolium repens*, P26205; *Populus* H-11 (1), POP1; *Manihot esculenta* (1), P2443454; *Manihot esculenta* (2), S23940; *Manihot esculenta* (3), X94986; *Costus speciosus*, D83122; *Oryza sativa*, 1143864; *Hordeum vulgare*, A57512; *Brassica napus*, U72154; *Arabidopsis thaliana*, U72155.

* - perfectly conserved amino acid, grey box indicates the location of loop C.

Active site domains are in bold.

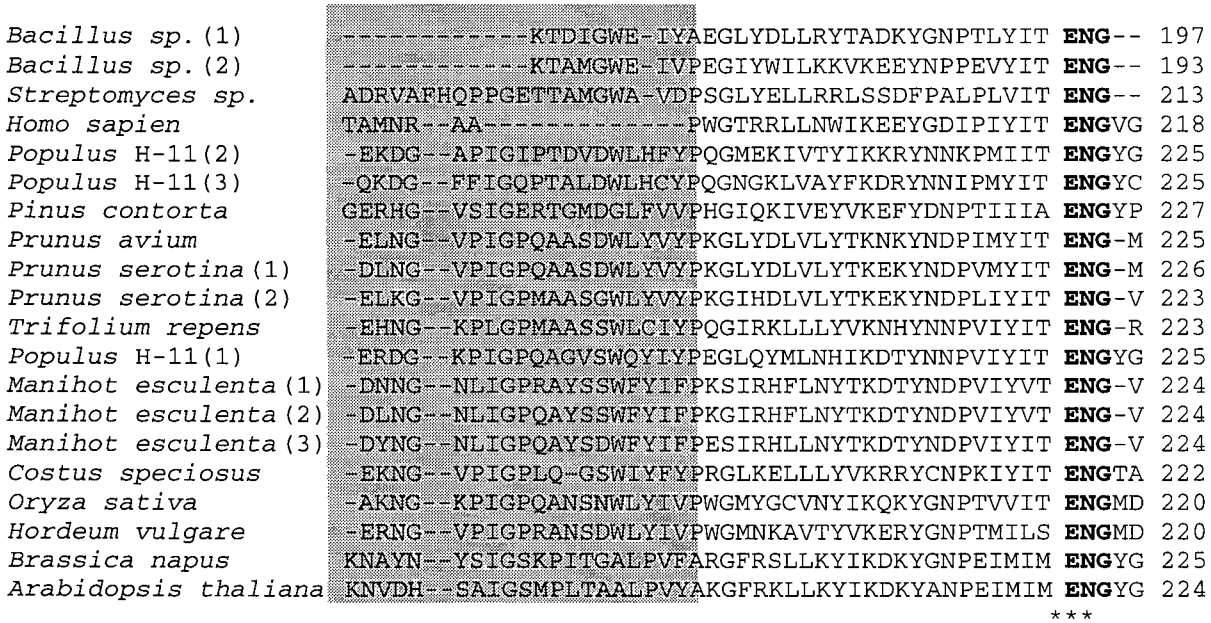
Data base accession numbers follow the species name where applicable.

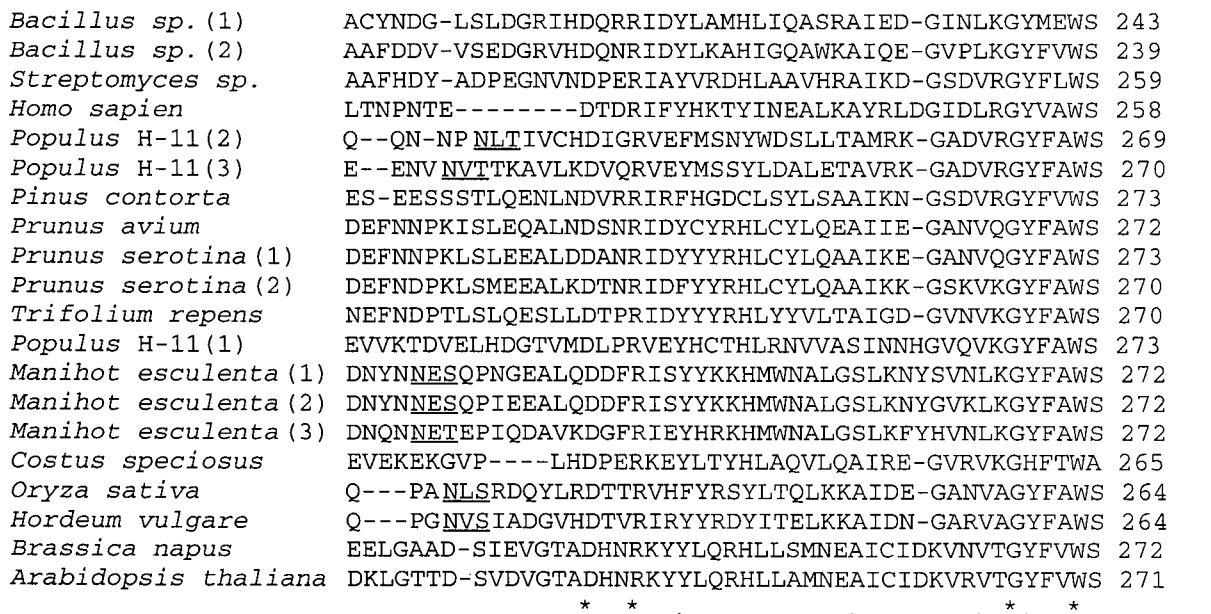
```
Bacillus sp.(1)          WITFNEPWCMAFLSNYLGVHAPGN--------------KDLQLAIDVSHH 36
Bacillus sp.(2)          WITLNEPWVVAIVGHLYGVHAPGM--------------RDIYVAFRAVHN 36
Streptomyces sp.         WTTLNEPWCSAFLGYGSGVHAPGR--------------TDPVAALRAAHH 36
Homo sapien              WMTFNEPMYLAWLGYGSGEFPPG--------------VKDPGWAPYRIAHT 37
Populus H-11(2)          WTTLNEPNIQTIKSYRSGEYPPCRCSSPFGN--CTHGD-SEKEPFIAAHN 47
Populus H-11(3)          WTTLNEPNVAAIHGYRSGIFPPSRCSRTFGY--CSSGD-SEREPFIAAHS 47
Pinus contorta           WATVNEPNLFVPLGYTVGIFPPTRCAAPHANPLCMTGNCSSAEPYLAAHH 50
Prunus avium             WTTLNEPYTISNHGYTIGIHAPGRCSSWYDPT-CLGGD-SGTEPYLVTHN 48
Prunus serotina(1)       WTTLNEPYTVSNHGYTIGIHAPGRCSCWYDPT-CLGGD-SGTEPYLVTHH 48
Prunus serotina(2)       WITLNEPYTFSSSGYAYGVHAPGRCSAWQKLN-CTGGN-SATEPYLVTHH 48
Trifolium repens         WITLNEPWGVSMNAYAYGTFAPGRCSDWLKLN-CTGGD-SGREPYLAAHY 48
Populus H-11(1)          WITLNEPQKFTGDGYDSGRFAPGRCSKWVDEKYCINGN-SSTEPYIVAHN 49
Manihot esculenta(1)     WMTFNEPSAYVGFAHDDGVFAPRRCSSWVNRQ-CLAGD-SATEPYIVAHN 48
Manihot esculenta(2)     WMTFNEPSAYVGFAHDDGVFAPGRCSSWVNRQ-CLAGD-SATEPYIVAHN 48
Manihot esculenta(3)     WMTFNEPWSLSGFAYDDGVFAPGRCSSWVNRQ-CRAGD-SATEPYIVAHH 48
Costus speciosus         WITLNEPWSLSTMGYAFGRHAPGRCSTWYG---CPAGD-SANEPYEVTHN 46
Oryza sativa             WFTFNEPRIVALLGYDQGTNPPKRCT-----KCAAGGN-SATEPYIVAHN 44
Hordeum vulgare          WFTFNEPRVVAALGYDNGFHAPGRCS-----KCPAGGD-SRTEPYIVTHN 44
Brassica napus           WITFNEPWVFSHAGYDVGKKAPGRCSKYVKEECHD--GRSGFEAYLVTHN 48
Arabidopsis thaliana     WITFNEPWVFAHAGYDLGKKAPGRCSRYVPG-CEDREGQSGKEAYLVSHN 49
                         * * ***      .   *  .*                    .    *

Bacillus sp.(1)          LLVAHGRAVTLFRE-LGI--SGEIGIAPNTSWAVPYR-RTKEDMEACLRV 82
Bacillus sp.(2)          LLRAHARAVKVFRE-TVK--DGKIGIVFNNGYFEPAS-EKEEDIRAVRFM 82
Streptomyces sp.         LNLGHGLAVQALRDRLPA--DAQCSVTLNIHHVRPLT-DSEADADAVRRI 83
Homo sapien              VIKAHARVYHTYDEKYRQEQKGVISLSLSTHWAEPKSPGVPRDVEAADRM 87
Populus H-11(2)          MILAHATAVDVYRTKYQKEQGGNIGIVLDCMWFEQIS-NSTADKLAADRA 96
Populus H-11(3)          MILSHAAAVNVYRTKYQKKQGGSIGIVMNAIWHEPIS-DSLEDKLAVERA 96
Pinus contorta           VLLAHASAVEKYREKYQKIQGGSIGLVISAPWYEPLE-NSPEERSAVDRI 99
Prunus avium             LLLAHAAAVKLYREKYQASQEGVIGITVVSHWFEPAS-ESQKDINASVRA 97
Prunus serotina(1)       LLLAHAAAVKLYREKYQASQNGVIGITIVSHWFEPAS-ESQQDKDAASRA 97
Prunus serotina(2)       QLLAHAAAVKLYKDEYQASQNGLIGITLVSPWFEPAS-EAEEDINAAFRS 97
Trifolium repens         QLLAHAAAARLYKTKYQASQNGIIGITLVSHWFEPAS-KEKADVDAAKRG 97
Populus H-11(1)          LLLSHAAAVHTYWEKYQASQNGKIGVTLNACWFEPYS-NSIEDRNAAKRS 98
Manihot esculenta(1)     LLLSHAAAVHQYRKYYQGTQKGKIGITLFTFWYEPLS-DSKVDVQAAKTA 97
Manihot esculenta(2)     LLLSHAAAVHQYRKYYQGTQKGKIGITLFTFWYEPLS-DSKVDVQAAKTA 97
Manihot esculenta(3)     LLLAHAAAVKIYRENYQETQNGKIGITLFTYWFEPLS-NSTDDMQASRTA 97
Costus speciosus         LLLAHANAVKIYRDNYKATQNGEIGITLNSLWYEPYS-KSHEDVEAATRA 95
Oryza sativa             FLLSHAAAVARYRTKYQAAQQGKVGIVLDFNWYEALS-NSTEDQAAAQRA 93
Hordeum vulgare          IILSHAAAVQRYREKYQPHQKGRIGILLDFVWYEPHS-DTDADQAAAQRA 93
Brassica napus           LLNSHAEAVEAFRQ-CEKCKGGKIGIAHSPAWFEPHDLADSQDGASIDRA 97
Arabidopsis thaliana     LLNAHAEAVEVFRQ---KVKGGKIGIAHSPAWFEPHDLKDSNDAPTVSRV 96
                            .*. .          .  ..       .       .  .
```

```
Bacillus sp.(1)          NGWSG-DWYLDPIYF-GEYPKFM---LDWYENLGYK-----PPIVDGDME 122
Bacillus sp.(2)          HQFNNYPLFLNPIYR-GDYPELV---LEFARE--YL-----PENYKDDMS 121
Streptomyces sp.         DALAN-RVFTGPMLQ-GAYPEDL---VKDTAGLTDW-----SFVRDGDLR 123
Homo sapien              LQFS-LGWFAHPIFRNGDYPDTMKWKVGNRSELQHLATSRLPSFTEEEKR 136
Populus H-11(2)          QDFF-LNWFLDPII-FGNYPAEMS----------KILGSTLPKFSSNDKE 134
Populus H-11(3)          NAFY-MNWFLDPII-LGKYPTEMR----------EILGSDLPVFSKYELE 134
Pinus contorta           LSFN-LRWFLDPIV-FGDYPQEMR----------ERLGSRLPSISSELSA 137
Prunus avium             LDFM-YGWFMDPLT-RGDYPQSMR----------SLVKERLPNFTEEQSK 135
Prunus serotina(1)       LDFM-YGWFMEPLT-RGDYPQTMR----------SIVGSRLPNFTEEQSK 135
Prunus serotina(2)       LDFI-FGWFMDPLT-NGNYPHLMR----------SIVGERLPNFTEEQSK 135
Trifolium repens         LDFM-LGWFMHPLT-KGRYPESMR----------YLVRKRLPKFSTEESK 135
Populus H-11(1)          LDFM-LGWFLNPIT-YGDYPSSMR----------ELVNDRLPTFSSLDSI 136
Manihot esculenta(1)     LDFM-FGLWMDPMT-YGRYPETMV----------DLAGDRLIGFTDEESQ 135
Manihot esculenta(2)     LDFM-FGLWMDPMT-YGRYPRTMV----------DLAGDKLIGFTDEESQ 135
Manihot esculenta(3)     LDFM-FGLWMDPIT-YGRYPRTVQ----------YLVGNRLLNFTEEVSH 135
Costus speciosus         LDFM-FGWYMDPLV-NGDYPFIMR----------ALVRDRLPFFTHAESE 133
Oryza sativa             RDFH-IGWYLDPLI-NGHYSQIMQ----------DLVKDRLPKFTPEQAR 131
Hordeum vulgare          RDFH-IGWFLDPIT-NGRYPSSML----------KIVGNRLPGFSADESR 131
Brassica napus           LDFI-LGWHLDTTM-YGDYPQIMK----------DIVGHRLPKFTEAQKA 135
Arabidopsis thaliana     LDFM-LGWHLEPTT-SGDYPQIMK----------DLLGYRLPQFTAAQKA 134
                                      .    * *.  .
```

```
Bacillus sp.(1)          LIHQPIDFIGINYYTSSMNRYNPGEAGGMLSSEAISMGAP---------- 162
Bacillus sp.(2)          EIQEKIDFVGLNYYSGHLVKFDPDAPAKV---SFVERDLP---------- 158
Streptomyces sp.         LAHQKLDFLGVNYYSPTLVSEADGSGTHNSDGHGRSAHSP-------WPG 166
Homo sapien              FIRATADVFCLNTY---YSRIVQHKTPRLNPPSYEDDQEMAEEEDPSWPS 183
Populus H-11(2)          KLKNGLDFIGINHYTSEYVQDCIFSVCEPGTGASRTEGLARR--SQ---- 178
Populus H-11(3)          RLKSGVDFIGINQYTSFYVKDCMFSTCEQGPGVSKTEGLYLR--TA---- 178
Pinus contorta           KLRGSFDYMGINHYTTLY------ATSTPPLSPDHTQYLYPD--SRVYLT 179
Prunus avium             SLIGSYDYIGVNYYSARYASAYPEDYSIPTPPSYLTDAYVNV--TT---- 179
Prunus serotina(1)       SLNGSYDYIGVNYYSARYASAYTNNYSVPTPPSYATDAYVNV--TTT--- 180
Prunus serotina(2)       LLKGSFDFIGLNYYTTRYASNAPKITSVHA--SYITDPQVN---ATA--- 177
Trifolium repens         ELTGSFDFLGLNYYSSYYAAKAPRI--PNARPAIQTDSLIN---ATF--- 177
Populus H-11(1)          NLKGSLDFVGLNYYTAYYAANANS--SSPDPRRYQTDSNCNI--TG---- 178
Manihot esculenta(1)     LLRGSYDFVGLQYYTAYYAKPNIT--VDPNFRTYKTDSGVNA--TPY--- 178
Manihot esculenta(2)     LLRGSYDFVGLQYYTAYYAEPIPP--VDPKFRRYKTDSGVNA--TPY--- 178
Manihot esculenta(3)     LLRGSYDFIGLQYYTSYYAKPNAP--YDPNHIRYLTDNRVTE--TPY--- 178
Costus speciosus         LIKGSYDFIGINYYTSNYAQHAPVTEDHTPDNSYF-DSYVNQ--SG---- 176
Oryza sativa             LVKGSADYIGINQYTASYMKG--QQLMQQTPTSYSADWQVTY--V-F---- 173
Hordeum vulgare          MVKGSIDYVGINQYTSYYMKD--PGAWNQTPVSYQDDWHVGF--V-Y---- 173
Brassica napus           KLKNSADFVGLNYYTSMFSN--HLEKPDPAKPRWMQDSLIN------WET 177
Arabidopsis thaliana     KLKDSTDFVGLNYYTSTFSN--YNEKPDPSKPSWKQDSLVS------WEP 176
                               *    .. *
```


Bacillus sp. (1) ----------KTDIGWE-IYAEGLYDLLRYTADKYGNPTLYIT ENG-- 197
Bacillus sp. (2) ----------KTAMGWE-IVPEGIYWILKKVKEEYNPPEVYIT ENG-- 193
Streptomyces sp. ADRVAFHQPPGETTAMGWA-VDPSGLYELLRRLSSDFPALPLVIT ENG-- 213
Homo sapien TAMNR--AA----------PWGTRRLLNWIKEEYGDIPIYIT ENGVG 218
Populus H-11(2) -EKDG--APIGIPTDVDWLHFYPQGMEKIVTYIKKRYNNKPMIIT ENGYG 225
Populus H-11(3) -QKDG--FFIGQPTALDWLHCYPQGNGKLVAYFKDRYNNIPMYIT ENGYC 225
Pinus contorta GERHG--VSIGERTGMDGLFVVPHGIQKIVEYVKEFYDNPTIIIA ENGYP 227
Prunus avium -ELNG--VPIGPQAASDWLYVYPKGLYDLVLYTKNKYNDPIMYIT ENG-M 225
Prunus serotina (1) -DLNG--VPIGPQAASDWLYVYPKGLYDLVLYTKEKYNDPVMYIT ENG-M 226
Prunus serotina (2) -ELKG--VPIGPMAASGWLYVYPKGIHDLVLYTKEKYNDPLIYIT ENG-V 223
Trifolium repens -EHNG--KPLGPMAASSWLCIYPQGIRKLLLYVKNHYNNPVIYIT ENG-R 223
Populus H-11(1) -ERDG--KPIGPQAGVSWQYIYPEGLQYMLNHIKDTYNNPVIYIT ENGYG 225
Manihot esculenta (1) -DNNG--NLIGPRAYSSWFYIFPKSIRHFLNYTKDTYNDPVIYVT ENG-V 224
Manihot esculenta (2) -DLNG--NLIGPQAYSSWFYIFPKGIRHFLNYTKDTYNDPVIYVT ENG-V 224
Manihot esculenta (3) -DYNG--NLIGPQAYSDWFYIFPESIRHLLNYTKDTYNDPVIYIT ENG-V 224
Costus speciosus -EKNG--VPIGPLQ-GSWIYFYPRGLKELLLYVKRRYCNPKIYIT ENGTA 222
Oryza sativa -AKNG--KPIGPQANSNWLYIVPWGMYGCVNYIKQKYGNPTVVIT ENGMD 220
Hordeum vulgare -ERNG--VPIGPRANSDWLYIVPWGMNKAVTYVKERYGNPTMILS ENGMD 220
Brassica napus KNAYN--YSIGSKPITGALPVFARGFRSLLKYIKDKYGNPEIMIM ENGYG 225
Arabidopsis thaliana KNVDH--SAIGSMPLTAALPVYAKGFRKLLKYIKDKYANPEIMIM ENGYG 224
. . . . . ***


Bacillus sp. (1) ACYNDG-LSLDGRIHDQRRIDYLAMHLIQASRAIED-GINLKGYMEWS 243
Bacillus sp. (2) AAFDDV-VSEDGRVHDQNRIDYLKAHIGQAWKAIQE-GVPLKGYFVWS 239
Streptomyces sp. AAFHDY-ADPEGNVNDPERIAYVRDHLAAVHRAIKD-GSDVRGYFLWS 259
Homo sapien LTNPNTE--------DTDRIFYHKTYINEALKAYRLDGIDLRGYVAWS 258
Populus H-11(2) Q--QN-NP NLT IVCHDIGRVEFMSNYWDSLLTAMRK-GADVRGYFAWS 269
Populus H-11(3) E--ENV NVT TKAVLKDVQRVEYMSSYLDALETAVRK-GADVRGYFAWS 270
Pinus contorta ES-EESSSTLQENLNDVRRIRFHGDCLSYLSAAIKN-GSDVRGYFVWS 273
Prunus avium DEFNNPKISLEQALNDSNRIDYCYRHLCYLQEAIIE-GANVQGYFAWS 272
Prunus serotina (1) DEFNNPKLSLEEALDDANRIDYYYRHLCYLQAAIKE-GANVQGYFAWS 273
Prunus serotina (2) DEFNDPKLSMEEALKDTNRIDFYYRHLCYLQAAIKK-GSKVKGYFAWS 270
Trifolium repens NEFNDPTLSLQESLLDTPRIDYYYRHLYYVLTAIGD-GVNVKGYFAWS 270
Populus H-11(1) EVVKTDVELHDGTVMDLPRVEYHCTHLRNVVASINNHGVQVKGYFAWS 273
Manihot esculenta (1) DNYN NES QPNGEALQDDFRISYYKKHMWNALGSLKNYSVNLKGYFAWS 272
Manihot esculenta (2) DNYN NES QPIEEALQDDFRISYYKKHMWNALGSLKNYGVKLKGYFAWS 272
Manihot esculenta (3) DNQN NET EPIQDAVKDGFRIEYHRKHMWNALGSLKFYHVNLKGYFAWS 272
Costus speciosus EVEKEKGVP----LHDPERKEYLTYHLAQVLQAIRE-GVRVKGHFTWA 265
Oryza sativa Q---PA NLS RDQYLRDTTRVHFYRSYLTQLKKAIDE-GANVAGYFAWS 264
Hordeum vulgare Q---PG NVS IADGVHDTVRIRYYRDYITELKKAIDN-GARVAGYFAWS 264
Brassica napus EELGAAD-SIEVGTADHNRKYYLQRHLLSMNEAICIDKVNVTGYFVWS 272
Arabidopsis thaliana DKLGTTD-SVDVGTADHNRKYYLQRHLLAMNEAICIDKVRVTGYFVWS 271
* * . . . *. *.

-ZAP vector adjacent to the cDNA insert) to screen for the presence of these β-glucosidase cDNA's within the library. All primer combinations (POP1a+T7, POP2a+T7 and POP3a+T7) were first tested with *POP1*, *POP2* and *POP3* genomic clones do determine 1) if each primer combination amplified the desired fragment and 2) if the primers differentiated between *POP1*, *POP2* and *POP3* sequences. The only primer combination that yielded a PCR product from the young leaf library was POP1a and T7. This suggested that *POP1* may be the only β-glucosidase gene among the three identified in the genomic DNA PCR screen which was expressed in the young leaf tissue used to synthesize this library.

To further support this data, an additional PCR screen of the young leaf library was conducted using degenerate primers BGL1 and BGL2. This amplification yielded a single 850bp PCR fragment which was approximately the same size as the PCR fragment obtained from amplification of the pine β-glucosidase cDNA clone with BGL1 and BGL2 (Figure 4.3). The young leaf library PCR product obtained with primers BGL1 and BGL2 was subsequently cloned into the T-tailed Bluescript SK+ vector and bacterial colonies were screened by PCR with primers BGL1 and BGL2 for the presence of the desired insert. The PCR products from the bacterial screen were observed on an ethidium bromide stained gel and then blotted for probing with *POP1* sequence. The eighteen positive colonies obtained in this screen were all shown to contain *POP1* sequence when the blots were washed under high stringency conditions previously shown to eliminate *POP1*, *POP2* or *POP3* cross-hybridization. These results further confirmed the observation that among the β-glucosidases identified in the genomic DNA PCR screen,

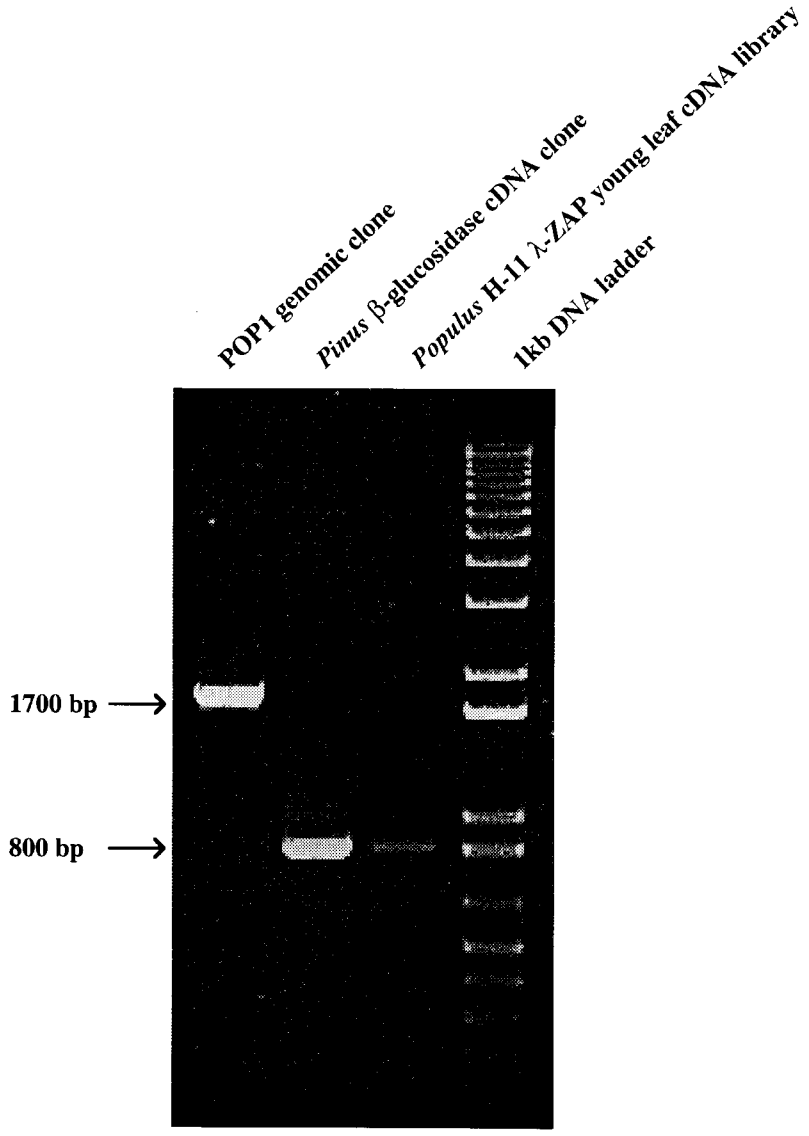


Figure 4.3 Degenerate PCR amplification from a *Populus* young leaf cDNA library using primers BGL1 and BGL2.

only *POP1* appeared to be expressed in young leaf tissue.

4.4 PCR Amplification of the 5' and 3' Ends of the POP1 cDNA

Two overlapping fragments, one containing the 5' end of the POP1 cDNA and one containing the 3' end of the POP1 cDNA, were cloned by PCR from the poplar young leaf cDNA library. An 1100bp fragment containing the 3' end of the POP1 cDNA was amplified using a POP1 exon-specific primer (POP1a) in combination with vector primer T7 (Figure 4.4) and cloned into Bluescript SK+. Ten plasmids isolated from putative recombinant colonies were screened for the presence of the 1100bp insert by restriction digestion with *Eco*RI and *Hind*III. Seven of these plasmids appeared to contain the desired insert and three of them (16P5, 16P6 and 16P8) were chosen for sequencing.

A 700bp fragment containing the 5' end of the POP1 cDNA was amplified using a second exon-specific primer (POP1b) and vector primer T3 (Figure 4.4). In addition to the predominant 700bp PCR product, several smaller PCR products were also amplified, presumably from less-than-full length POP1 cDNA clones within the library. After gel purification, the 700bp PCR fragment was cloned into the Bluescript SK+ vector and three clones that appeared to contain the desired insert when digested with BamHI and HindIII (16R2, 16R7 and 16R8) were chosen for sequencing.

The near full length sequence of the 1823bp POP1 cDNA could be deduced by integrating the 5' clone sequence and the 3' clone sequence at the 18bp overlapping region shared by these clones. Sequence information available from the partial *POP1* clone amplified previously from genomic DNA was used to confirm that the 5' and 3'

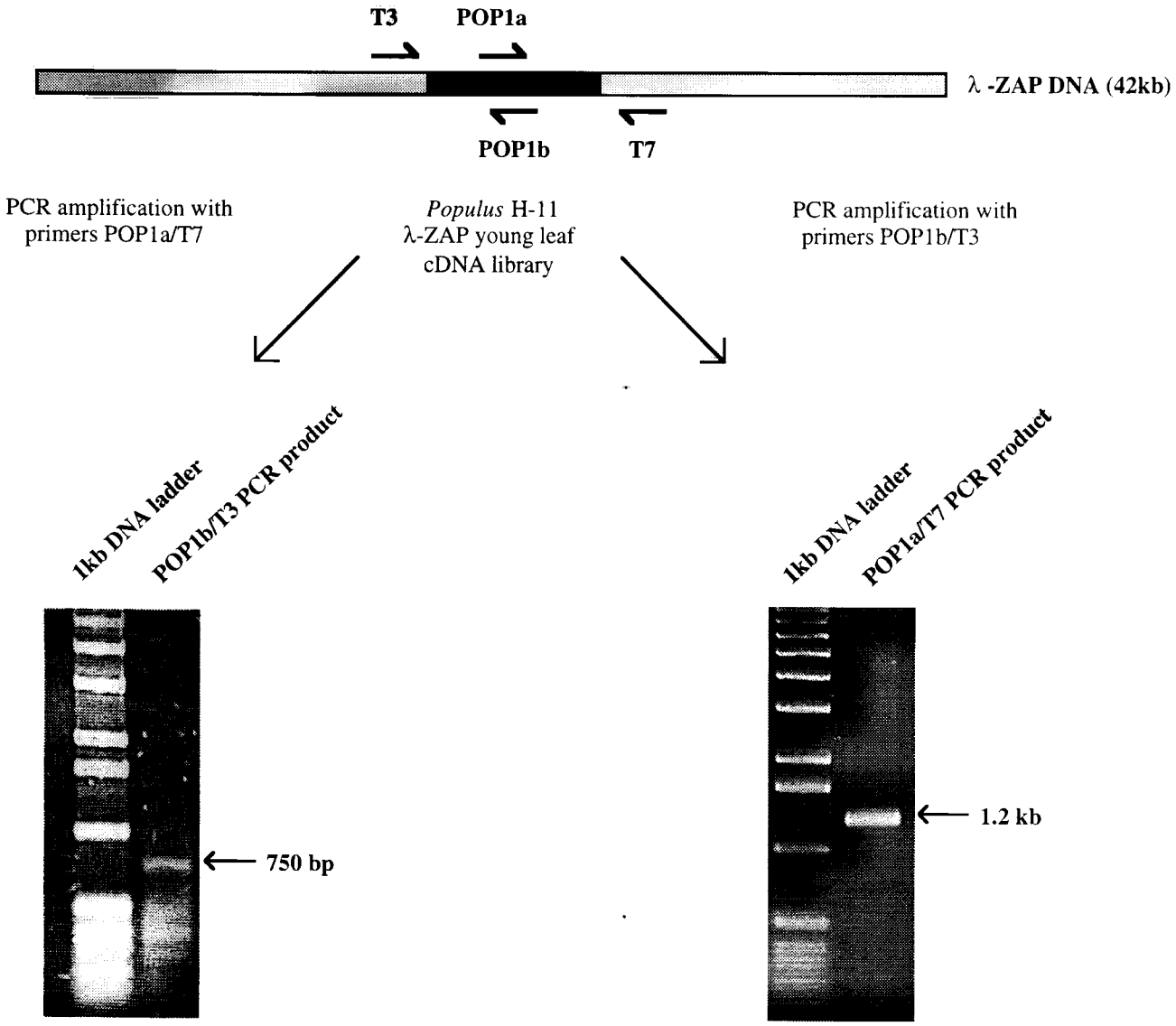


Figure 4.4 PCR amplification of the 5' and 3' ends of the POP1 cDNA from a *Populus* young leaf cDNA library.

fragments were correctly assembled. Four putative initiation ATG-codons were located within 100 nucleotides of the 5' end of the POP1 cDNA and were in frame to encode proteins characteristic of *family* 1 β-glucosidases. However, several factors suggested that the most 5' ATG would be preferentially used for initiation of protein synthesis *in vivo*; 1) the most 5' ATG codon was flanked by nucleotides that most closely matched the optimum consensus plant protein initiation sequence, 2) 90-95% of proteins are initiated from the first start codon of an mRNA sequence and 3) the predicted protein product synthesized from initiation at the first ATG codon was most characteristic of *family 1* β-glucosidases (Joshi 1987, Joshi *et al.* 1997). Taken together, this information strongly suggested that the sequence generated in this experiment contained the entire coding region for the POP1 protein.

Further characterization of the predicted POP1 protein product is based on the assumption that protein synthesis began at the first ATG codon located 25 bps from the 5' end of the POP1 cDNA sequence. Since most mRNAs contain 5' untranslated sequences ranging from 50-200 nucleotides in length, it is considered likely that the POP1 sequence presented here does not represent a completely full-length clone (Joshi *et al.* 1997). However, for the purposes of this study additional 5'-sequence information was unnecessary and therefore not pursued. The 3'-untranslated region of *POP1* is 175bp in length and contains a conserved polyadenylation signal.

The POP1 cDNA had an open reading frame spanning from nucleotide 26 to nucleotide 1649 and was predicted to encode for a 541 amino acid protein with a molecular mass of 61373 Da and a calculated isoelectric point of 5.26. The predicted

protein product contained the conserved 'NEP' and 'ENG' motifs that are believed to be essential for the catalytic activity of *family 1* β-glucosidases. Five additional residues that are related to catalytic activity and are strictly conserved among family 1 β-glucosidases are located at Gln65, His169, Tyr359, Trp481 and Gln488 in the POP1 protein (Figure 4.5). Six putative N-asparagine glycosylation sites matching the consensus glycosylation motif (Asn-X-(Ser-Thr)), where X can be any amino acid except proline, are found in POP1 (Figure 4.5). Similar to most characterized plant *family 1* β-glucosidases, the putative POP1 protein appears to contain a twenty-two amino acid N-terminal signal peptide (Figure 4.5).

Within the large *family 1* group, the POP1 protein sequence showed the highest similarity to plant sequences, with amino acid identities ranging from 43% to 52%. The β-glucosidases most similar to *POP1* have been cloned from *Manihot esculenta* (cassava), *Prunus serotina* (black cherry) and *Costus speciosus*. The POP1 protein and the pine coniferin β-glucosidase were 46% identical, a level of identity that is essentially the same as that seen between different angiosperm sequences. A BLAST search with *POP1* in the recently created *Populus* developing xylem EST database produced a hit against an expressed sequence (clones A001P37U and A001P60) that did not, however, correspond to any of the putative β-glucosidases identified in this study (*POP1*, *POP2* or *POP3*). An alignment of the deduced polypeptide encoded by the EST database sequence against several plant β-glucosidases as well as *POP1*, *POP2* and *POP3* strongly suggested that the sequence identified in the EST database encoded a fragment of a previously unidentified *family 1* β-glucosidase that was expressed in poplar developing

```
   1 ----------------------------------------------------------CATATACACATTACCAACTAGCTTG

   1  M  A  I  P  T  G  L  L  L  L  G  M  L  F  L  M  F  S  F  A  S  A  Q  I  E  M  P  R  G  I
  26 ATGGCTATTCCTACTGGCTTGCTTTTATTAGGAATGCTCTTCCTAATGTTCTCGTTTGCTTCTGCTCAGATAGAGATGCCAAGAGGTATC

  31  T  E  D  F  D  T  T  C  I  T  D  L  K  F  G  L  S  R  K  S  F  P  G  D  F  I  F  G  A  A
 116 ACTGAGGACTTTGATACGACTTGTATCACTGATCTTAAATTCGGGTTATCTCGTAAATCCTTTCCGGGGGATTTCATATTCGGGGCAGCG

  61  A  S  A  Y  Q  T  E  G  H  A  N  K  S  C  R  G  P  S  I  W  D  T  F  T  Q  D  F  P  E  R
 206 GCATCCGCTTACCAGACCGAGGGTCATGCAAACAAAAGTTGCAGAGGTCCGAGTATATGGGACACTTTTACTCAAGATTTCCCAGAAAGG

  91  I  A  D  G  C  N  G  D  L  G  I  D  L  Y  N  R  Y  E  S  D  L  E  E  M  K  D  M  N  M  D
 296 ATAGCTGATGGCTGCAACGGAGACTTGGGAATTGATTTGTACAATCGCTACGAGAGTGATCTTGAAGAAATGAAGGATATGAATATGGAT

 121  A  F  R  F  S  I  S  W  S  R  V  I  P  S  G  K  I  R  A  G  V  N  K  D  G  I  E  F  Y  N
 386 GCTTTCAGATTCTCGATCTCCTGGTCCAGAGTAATTCCCAGTGGGAAAATAAGGGCAGGAGTGAACAAAGATGGAATTGAGTTTTACAAC

 151  K  L  I  D  A  T  I  A  K  G  L  Q  P  Y  A  T  L  F  H  W  D  V  P  Q  A  L  E  D  K  Y
 476 AAGCTAATCGATGCAACCATAGCTAAAGGTTTACAGCCCTATGCAACTCTCTTTCATTGGGATGTTCCTCAAGCACTTGAAGACAAGTAC

 181  G  G  F  L  S  D  N  I  V  S  D  F  R  D  F  A  E  L  C  F  K  E  F  G  D  R  V  K  Y  W
 566 GGTGGCTTTCTCAGTGATAATATTGTGAGCGACTTCCGAGATTTTGCTGAGCTTTGCTTCAAGGAATTTGGTGACCGAGTGAAGTACTGG

 211  I  T  L  N  E  P  Q  K  F  T  G  D  G  Y  D  S  G  R  F  A  P  G  R  C  S  K  W  V  D  E
 656 ATTACTTTGAATGAGCCACAAAAGTTCACCGGTGATGGCTACGATTCAGGCCGCTTTGCACCAGGCCGATGTTCCAAGTGGGTGGATGAA

 241  K  Y  C  I  N  G  N  S  S  T  E  P  Y  I  V  A  H  N  L  L  L  S  H  A  A  A  V  H  T  Y
 746 AAGTACTGCATAAATGGGAACTCTTCCACCGAGCCTTACATAGTTGCCCATAATCTCCTCCTTTCCCATGCAGCAGCAGTACATACATAT

 271  W  E  K  Y  Q  A  S  Q  N  G  K  I  G  V  T  L  N  A  R  W  F  E  P  Y  S  N  S  I  E  D
 836 TGGGAAAAGTATCAGGCATCTCAAAATGGGAAGATTGGGGTAACACTCAATGCCCGCTGGTTTGAACCGTACTCTAACAGTATAGAAGAT

 301  R  N  A  A  K  R  S  L  D  F  M  L  G  W  F  L  N  P  I  T  Y  G  D  Y  P  S  S  M  R  E
 926 CGTAATGCTGCCAAAAGATCTCTTGATTTCATGCTTGGTTGGTTCCTGAACCCTATAACATATGGTGACTATCCGAGCAGCATGCGGGAA

 331  L  V  N  D  R  L  P  T  F  S  S  L  D  S  I  N  L  K  G  S  L  D  F  V  G  L  N  Y  Y  T
1016 TTAGTTAATGATAGGCTACCAACATTCTCTTCACTGGATTCTATAAATCTCAAAGGATCGTTGGACTTTGTTGGATTGAATTACTATACT

 361  A  Y  Y  A  A  N  A  N  S  S  S  P  D  P  R  R  Y  Q  T  D  S  N  C  N  I  T  G  E  R  D
1106 GCATATTATGCAGCAAATGCGAATTCTTCTAGTCCAGACCCTCGTAGATACCAAACAGATTCTAATTGCAATATTACAGGAGAACGAGAT

 391  G  K  P  I  G  P  Q  A  G  V  S  W  Q  Y  I  Y  P  E  G  L  Q  Y  M  L  N  H  I  K  D  T
1196 GGCAAACCCATCGGTCCACAGGCTGGGGTATCATGGCAGTATATCTATCCAGAGGGGTTGCAGTATATGTTGAATCACATCAAAGATACA

 421  Y  N  N  P  V  I  Y  I  T  E  N  G  Y  G  E  V  V  K  T  D  V  E  L  H  D  G  T  V  M  D
1286 TACAATAATCCAGTAATTTACATCACTGAAAATGGCTACGGTGAAGTCGTTAAGACTGATGTAGAACTACATGATGGGACCGTGATGGAT

 451  L  P  R  V  E  Y  H  C  T  H  L  R  N  V  V  A  S  I  N  N  H  G  V  Q  V  K  G  Y  F  V
1376 CTTCCAAGGGTAGAATATCATTGTACTCATCTTAGGAATGTTGTAGCATCTATCAACAATCATGGAGTTCAAGTCAAAGGCTATTTTGTA

 481  W  S  F  A  D  N  F  E  F  T  D  G  Y  T  I  G  F  G  L  L  Y  V  N  R  T  S  N  F  T  R
1466 TGGTCCTTTGCCGACAATTTCGAATTTACAGATGGATATACCATAGGATTCGGTTTGTTGTACGTAAACCGTACGTCTAATTTCACAAGA

 511  I  K  K  L  S  S  H  W  F  T  E  F  L  G  D  Q  P  A  N  P  V  P  L  Y  F  K  R  L  N  I
1556 ATAAAGAAGCTCTCATCACATTGGTTTACAGAATTCCTTGGGGACCAACCAGCTAATCCAGTACCATTATATTTCAAGCGATTGAATATA

 541  A
1646 GCTTGATGGAAGAGAAATTAAGTATTTGGCCGCCAATATAAATAAATAAATAATCACTAGATAACTTTTATCCAGTGAAGTCAATTGTAT

1736 AATTTTTCCTAATTAAGTGTGTGATGTATTATACTGTATTGTTTCAATAAAAACTTGGTTATAAAAAAAAAAAAAAAAAAAAAAAAAA
```

Figure 4.5 Nucleotide and predicted amino acid sequence of the POP1 cDNA. A proposed signal peptide is boxed in grey. Active site residues are highlighted. Potential N-linked glycosylation sites are underlined. Amino acid residues that are perfectly conserved among family 1 β-glucosidases are in bold.

xylem tissue (Figure 4.6).

4.5 Discussion

Glucosyl hydrolases form an extremely large and diverse group of enzymes with representatives found in all kingdoms. These enzymes are thought to act by a general acid catalytic mechanism in which two amino acid residues participate in hydrolysis, with either retention or conversion of configuration at the anomeric carbon atom of the hydrolyzed glucose (Sinnott 1990). A classification system based on amino acid similarity has been used to organize these enzymes into at least 57 different families (Henrissat 1991, Henrissat and Bairoch 1993, 1996). Grouping these families in this fashion, rather than based on substrate specificity, has made it possible to detect conserved structural features, mechanistic similarities and evolutionary relationships.

Based on this classification scheme, one major class of glucosyl hydrolases has been labeled as *family 1* β-glucosidases (Henrissat 1991, Henrissat and Bairoch 1993, 1996). *Family 1* β-glucosidases share the same basic $(\alpha/\beta)_8$ barrel tertiary conformation, they contain several perfectly conserved amino acid residues, and they employ the same mechanistic strategy of hydrolyzing β-glucosidic linkages with overall retention of configuration at the anomeric carbon of the hydrolyzed glucose (Sanz-Aparicio 1998, Day and Withers 1986). A striking characteristic of this family of enzymes is their collective ability to accommodate a wide range of substrates despite the significant sequence conservation between family members found in organisms as diverse as bacteria, mammals and plants.

```
POP1          HAAAVHTYWE KYQASQNGKI GVTLNACWFE PYSNSIEDRN AAKRSLDFML
POP2          HATAVDVYRT KYQKEQGGNI GIVLDCMWFE QISNSTADKL AADRAQDFFL
POP3          HAAAVNVYRT KYQKKQGGSI GIVMNAIWHE PISDSLEDKL AVERANAFYM
xylem EST     NAAAVQRYRE KYQEKQKGRI GILLDFVYYE PLTRSKADNL AAQRARDFHI
               * **  *   ***  * * * *        *     *  *   *

POP1          GWFLNPITYG DYPSSMRELV NDRLPTFSSL DSINLKGSLD FVGLNYYTAY
POP2          NWFLDPIIFG NYPAEMSKIL GSTLPKFSSN DKEKLKNGLD FIGINHYTSE
POP3          NWFLDPIILG KYPTEMREIL GSDLPVFSKY ELERLKSGVD FIGINQYTSF
xylem EST     GWFIHPIVYG EYPXTMQNIV GNRLPQFTKE EVKMVKGSMD FVGINQYTTY
               **  **  *  **  *       ** *         *   * * * * **

POP1          YAANA
POP2          YVQDC
POP3          YVKDC
xylem EST     YLYDP
               *
```

Figure 4.6 Amino acid alignment of POP1, POP2 and POP3 with the *Populus* EST clone identified in developing xylem. *- indicates perfectly conserved residues.

A hallmark for identification of *family 1* β-glucosidases is the presence of signature 'NEP' and 'ENG' motifs at conserved regions in their amino acid sequence (Figure 4.2). It was determined by the analysis of mutant enzymes and by covalent inhibitor trapping studies that, during the first step of hydrolysis, the glutamate residue in the ENG motif is responsible for the initial nucleophilic attack on the substrate (Withers *et al.* 1990, Trimber *et al.* 1992). This attack results in covalent attachment of the glucose moiety to the glutamate residue while the aglycone is released as the leaving group. In the second step of hydrolysis, the glucose is released from the enzyme by acid-base catalysis (Withers and Aebersold 1995), catalyzed by the glutamate of the 'NEP' motif (Wang *et al.* 1995). All *family 1* β-glucosidases contain both signature active site sequences, with the exception of the Brassicaceae myrosinases which contain conserved 'NQL' motifs instead of 'NEP' motifs. This difference may reflect the role of myrosinase in hydrolyzing S-linked rather than O-linked glucosides. The conserved 'NEP' and 'ENG' motifs are present in the amino acid sequence of all three poplar family 1 β-glucosidases identified in this study.

Additional information about active site residues contributing to the catalytic activity of *family 1* β-glucosidases has been obtained from the co-crystallization of a *Bacillus* β-glucosidase covalently bound to its transition state inhibitor (Sanz-Aparicio 1998). The predicted catalytic glutamate residues encircle the anomeric sugar C1 atom of the bound substrate while it sits within the active site pocket formed at the bottom of the ($\alpha/\beta)_8$ barrel. In addition, several residues contribute through hydrogen bonding to substrate binding in the active site and to stabilization of the catalytic glutamates. These

residues are strictly conserved among *family 1* β-glucosidases and correspond to Trp398, His121, Tyr296, Gln20, and Glu405 of the *Bacillus* enzyme. The corresponding conserved residues (Trp481, His169, Tyr359, Gln65and Glu488) are also present in the POP1 β-glucosidase identified in this study (Figure 4.5).

Structural superposition of x-ray crystal structures for *family 1* β-glucosidases from *Bacillus*, *Lactococcus*, *Trifolium repens* and *Sinapis alba* has provided information about the putative substrate recognition region in these enzymes (Sanz-Aparicio 1998). When comparing these four enzymes which utilize dramatically different natural substrates a striking conservation of overall topography was observed except in four variable loop regions that contain structural rearrangements derived from insertions and deletions. These loops appeared to be important in establishing the size and shape of the substrate binding pocket. Loop C, which spans amino acids 297 to 328 in the *Bacillus* enzyme, was predicted by its proximity to the catalytic glutamates to be the most important region in determining substrate specificity (Figure 4.2). Further analysis of these predicted substrate recognition regions should prove invaluable in determining how *family 1* β-glucosidases recognize their preferred substrate.

A common characteristic of plant *family 1* β-glucosidases is the presence of an N-terminal signal sequence that is cleaved from the mature protein. Full length cDNA sequences are available for nine plant *family 1* β-glucosidases for which the amino terminus sequence of the corresponding mature proteins is also known (Dharmawardhana *et al.* 1995, Brzobohaty *et al.* 1993, Hughes *et al.* 1992, Falk and Rask 1995, Gus-Mayer *et al.* 1994, Zheng and Poulton 1992, Wiersma 1996). All nine of these cDNAs encode

enzymes containing N-terminal signal peptides that are not present in their mature peptides. Furthermore, a preliminary computer analysis of the 50+ plant *family 1* β-glucosidase sequences available in the GeneBank database suggests that most of these enzymes contain N-terminal signal peptides.

Signal sequences in plants serve to direct proteins to many different locations in the cell, including the cell wall, the vacuole, the plasma membrane, the chloroplast, the mitochondria and the tonoplast (Nakamura and Matsuoka 1993, Rusch and Kendall 1995, Chrispeels and Raikhel 1992). Initially, a signal sequence directs newly synthesized proteins into the ER where the signal peptide is removed and the nascent protein is modified (glycosylation, protein folding, disulfide bond formation, and oligomerization). Proteins that are destined for secretion, which is the bulk flow pathway in plant cells as in all eukaryotic cells, then pass through the ER, the Golgi apparatus, the *trans*-Golgi network, and transport vesicles before being secreted from the cell. Proteins destined for compartments within the cell contain additional targeting information but follow the same bulk flow pathway until they are sorted from secretory proteins, likely in the trans-Golgi network, to specific cellular locations such as the vacuole or chloroplast.

Signal peptides vary greatly in both their primary amino acid sequence and their overall length (Martoglio and Dobberstein 1998). However, most cleaved N-terminal signal sequences have a characteristic hydrophobic core (h-) region comprised of six to fifteen amino acid residues. The h-region is flanked on it's C-terminal side by a polar (c-) region which often contains proline and glycine residues as well as small uncharged residues in positions -3 and -1 (alanine, valine, glycine, serine, threonine or cysteine).

These apparently determine the site of signal peptide cleavage. The h-region is flanked on its N-terminal side by a polar n-region, which usually has a net positive charge.

The most common type of β-glucosidase signal peptide (found in POP1) is 20-25 amino acids long and appears to serve as an ER retention signal targeting these enzymes to the bulk flow secretion pathway. A comparison within this class of signal peptides reveals a striking retention of hydrophobicity and charge properties (characteristics outlined in the previous paragraph) but little conservation of primary amino acid sequence. A second group of β-glucosidases contain N-terminal signal peptides that are generally 50-60 amino acids long. In addition to functioning as ER retention signals, these longer sequences have been shown to target these enzymes to the plastid.

What information does the presence of a signal peptide in almost all plant *family 1* β-glucosidases reveal about the function of these proteins? It appears from several examples that these enzymes are synthesized and compartmentalized at the cellular or sub-cellular level to isolate the β-glucosidase enzymes from their substrates. In oat (*Avena sativa*), the As-P60 β-glucosidase has been shown by electron microscopy and immunochemistry to be located predominantly in the plastid (Nisius and Ruppel 1987). As-P60 and its native substrate, avenacoside B, only come together after tissue damage. This mixing of enzyme and substrate releases the anti-fungal saponin, 26-desgluco-avenacoside B. Similarly, in cassava and white clover it has been shown by vacuum infiltration and immuno-cytofluorescence that the β-glucosidase responsible for hydrolysis of the cyanogenic glucosides, lotaustralin and linamarin, is located in the cell wall and apoplast (Mkpong OE *et al.* 1990, Kakes P 1985). Mechanical leaf damage

results in the mixing of enzyme and substrate, leading ultimately to the release of cyanide, a proposed defense chemical in these species. These data suggest that the signal peptide in the POP1 protein may also act to compartmentalize this β-glucosidase in a separate location from its natural substrate. However, the subcellular location of POP1 in poplar leaves is currently unknown.

An additional characteristic of *family 1* β-glucosidases is the presence of several N-linked glycosylation sites. These sites are identified by their NX(S/T) motif, where X can be any residue except proline (Bause 1983). Within the ER, high-mannose oligosaccharides are attached to the asparagine (N) side chain of secreted proteins by the oligosaccharyltransferase complex. It is believed that glycosylation plays a role in protein folding, resistance to proteolytic cleavage and secretory pathway targeting. Mature β-glucosidases from several plant species contain N-linked oligosacarides. The putative POP1 protein contains six potential N-linked glycosylation sites but whether any of these are utilized *in planta* is unknown.

The experiments conducted in this portion of the study identified the partial genomic DNA sequence of three unique family 1 β-glucosidases in H-11 poplar. Although they are different from each other, a detailed analysis of the *POP1*, *POP2* and *POP3* sequences indicates that the genes from which they are derived most likely encode *family 1* β-glucosidases. Furthermore, a near full-length sequence for the POP1 cDNA was identified in a poplar young leaf library. As the tissue from which the library was generated is known to undergo lignification, the *POP1* gene was considered to be a candidate for a poplar β-glucosidase that might be involved in the hydrolysis of

monolignol glucosides during lignin synthesis.

Chapter 5
Expression Studies and the Recombinant POP1 Protein

5.1 Introduction

To assess the potential role of POP1 as a lignification specific β-glucosidase it was necessary to establish its catalytic function. The accumulation of *POP1* transcripts was investigated in poplar developing xylem harvested during May of 1995 and 1996. In the spring, poplar undergoes an intensive growth phase that involves the lignification of newly developing xylem tissue. Since many lignification-specific enzymes are known to be up-regulated in xylem tissue during this growth period, it was predicted that POP1 could also be up-regulated if this enzyme were a component of the lignification pathway in poplar.

Many enzymes are also extremely specific in the substrates on which they can act. Although cloning a homologous sequence for one of these enzymes does not guarantee its identity, a fairly confident prediction of the substrate preference can be made. On the other hand, plant *family 1* β-glucosidases appear to be capable of hydrolyzing an extremely wide range of glycosylated molecules. Predicted natural substrates for cloned plant *family 1* β-glucosidases include saponins, cyanogenic glucosides, conjugated

growth hormones, glycosylated lignin precursors, glucosinolates and starch metabolites (Inoue *et al.* 1996, Gus-Mayer *et al.* 1994, Hughes *et al.* 1992, Zheng and Poulton 1995, Falk and Rask 1995, Dharmawardhana *et al.* 1995, Thangstad *et al.* 1993, Leah *et al.* 1995). This variability in hydrolysis capabilities has made it difficult to predict *family 1* β-glucosidase substrate preferences. The enzyme activity of recombinant POP1 (expressed in baculovirus) was therefore investigated directly in an attempt to determine this enzyme's natural substrates.

5.2 Expression Analysis of POP1

Expression of the *POP1* gene in poplar H-11 species was examined in young leaves and in the developing xylem of trees harvested in May of 1995 and May of 1996. Northern blot analysis showed that *POP1* mRNA accumulated to low levels in young leaf tissue but was not present at detectable levels in developing xylem tissue in either sampling year (Figure 5.1). Furthermore, attempts to amplify an internal region of *POP1* from total poplar developing xylem RNA by RT-PCR were unsuccessful (data not shown). In this experiment, a poly $T_{(18)}$ primer was used for the reverse transcription step and primers BGL1 and BGL2 were used for the PCR amplification step. From the data generated in these experiments it appears that *POP1* mRNA is not present in developing xylem tissue sampled from *Populus* in the spring.

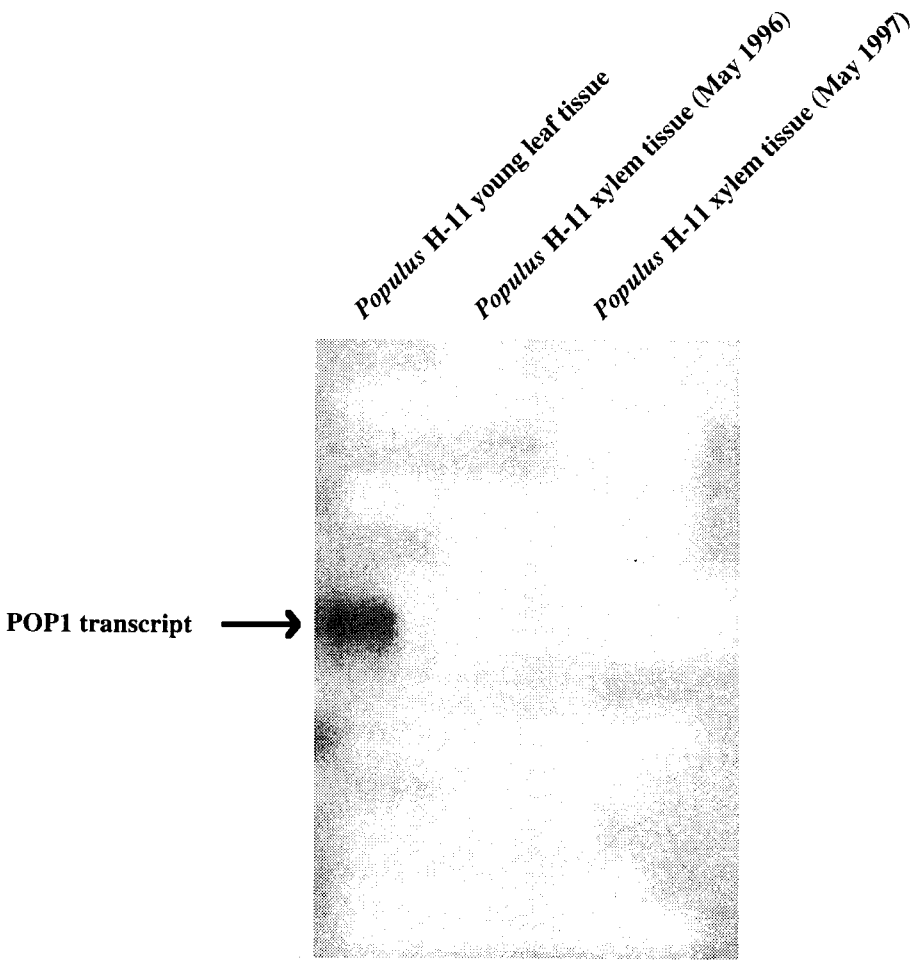


Figure 5.1 Northern blot analysis of POP1 transcript accumulation in *Populus* lignifying tissue.

5.3 Recombinant POP1

5.3.1 Cloning of the POP1 Coding Region into the pVL1392 Baculovirus Expression Vector

The putative coding region of the *POP1* gene was amplified by PCR from a poplar H-11 young leaf library using primers CODf and CODr. The 1559bp PCR product was cloned into the *Pst*I/*Bam*HI restriction sites in the polycloning site of the baculovirus pVL1392 expression vector. The cloning of *POP1* into pVL1392 was confirmed by restriction digest analysis with PstI/BamHI and clone POP1-4 was chosen for use in the baculovirus expression system.

5.3.2 SDS Page

Sf-9 (*Spodoptera frugiperda*) insect cells were infected with either wild type AcMNPV baculovirus or the pVL1392:POP1 expression construct, and harvested 72 h post-infection. A 72 h collection time point was chosen based on a preliminary time-course protein expression experiment (12 h sampling intervals) in which 72 h was shown to be the time point at which the POP1 protein was maximally expressed (data not shown). Cells infected with the pVL1392:POP1 expression construct were shown to overexpress an approximately 60 kDa protein (estimated by migration relative to protein standards) that was not present in Sf-9 cells infected with wild type baculovirus (Figure 5.2). The recombinant POP1 protein appeared to be about the same molecular weight as was predicted for the derived polypeptide (61.4 kDa), suggesting that post translational

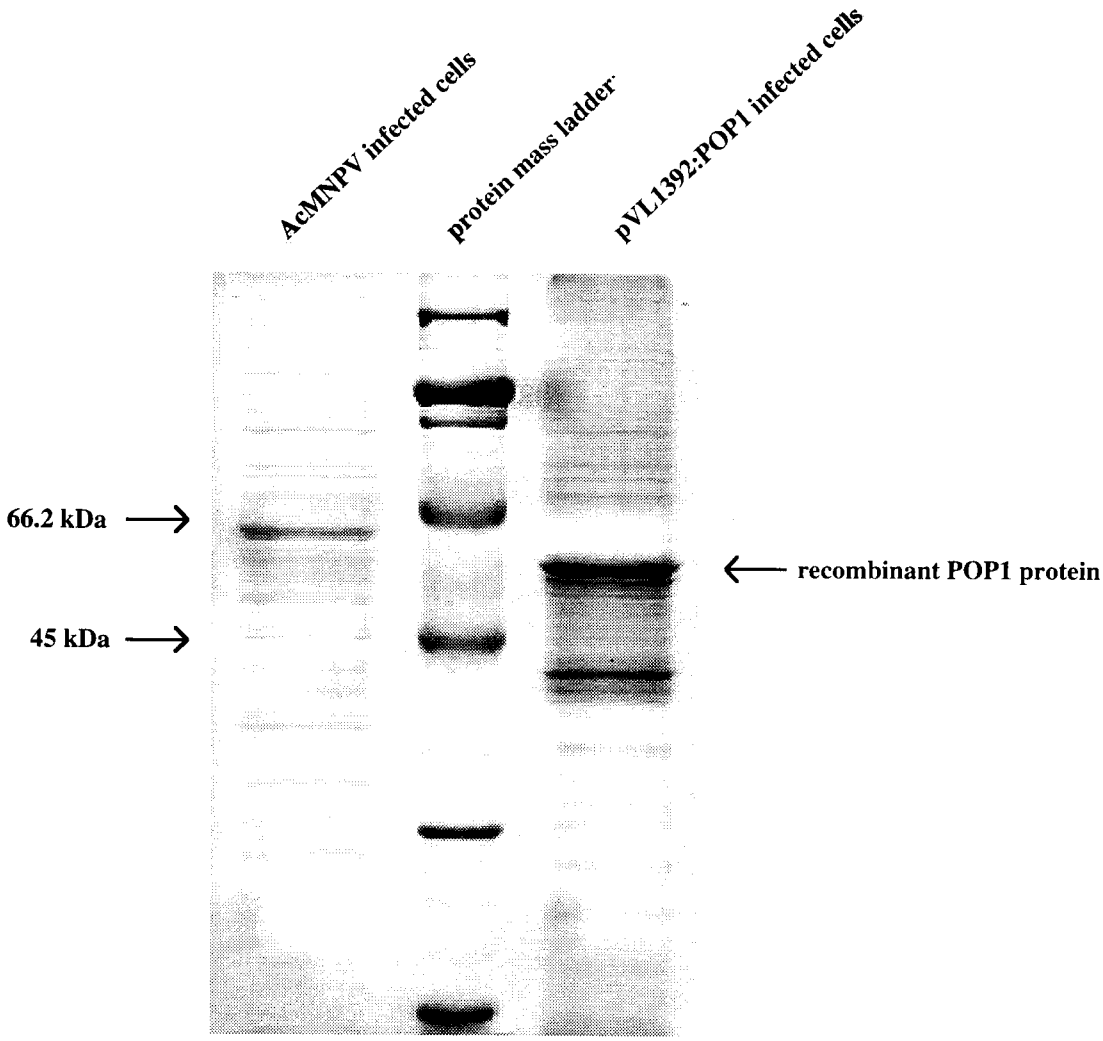


Figure 5.2 SDS-Page analysis of recombinant POP1 expression.

modifications effecting the migration rate of the POP1 protein did not occur during expression of this protein in insect cells.

5.3.3 Western Blot

Total protein extracts from Sf-9 cells expressing the POP1 protein (harvested 72 h post infection) and from Sf-9 cells infected with wild type AcMNPV baculovirus (harvested 72 h post-infection) were examined by western blot analysis (Figure 5.3). Antibodies raised against the purified, denatured pine coniferin β-glucosidase strongly cross-reacted with several proteins present in these extracts. In the lane containing proteins extracted from Sf-9 cells infected with the pVL1392:POP1 expression vector an immuno-reactive band was observed at about 61 kDa, the predicted size of the POP1 protein. These results indicate that the pine β-glucosidase antibody is capable of recognizing epitopes present in the POP1 protein. However, it should be noted that the strength of the signal observed in lane 1 is likely due to the abundance of POP1 protein present in this extract and is not due to the high specificity of the pine β-glucosidase antibody for the POP1 protein. A similar band was not observed in protein extracts from Sf-9 cells infected with wild type AcMNPV baculovirus.

Additional bands that cross reacted with the pine β-glucosidase antibody were observed in protein extracts from Sf-9 cells that had been infected with either the wild type AcMNPV baculovirus or the pVL1392:POP1 expression construct (in particular, two proteins, at about 20 kDa and 6 kDa, had strong cross-reactions with the pine antibody). These bands were believed to arise from the cross-reaction of antibodies

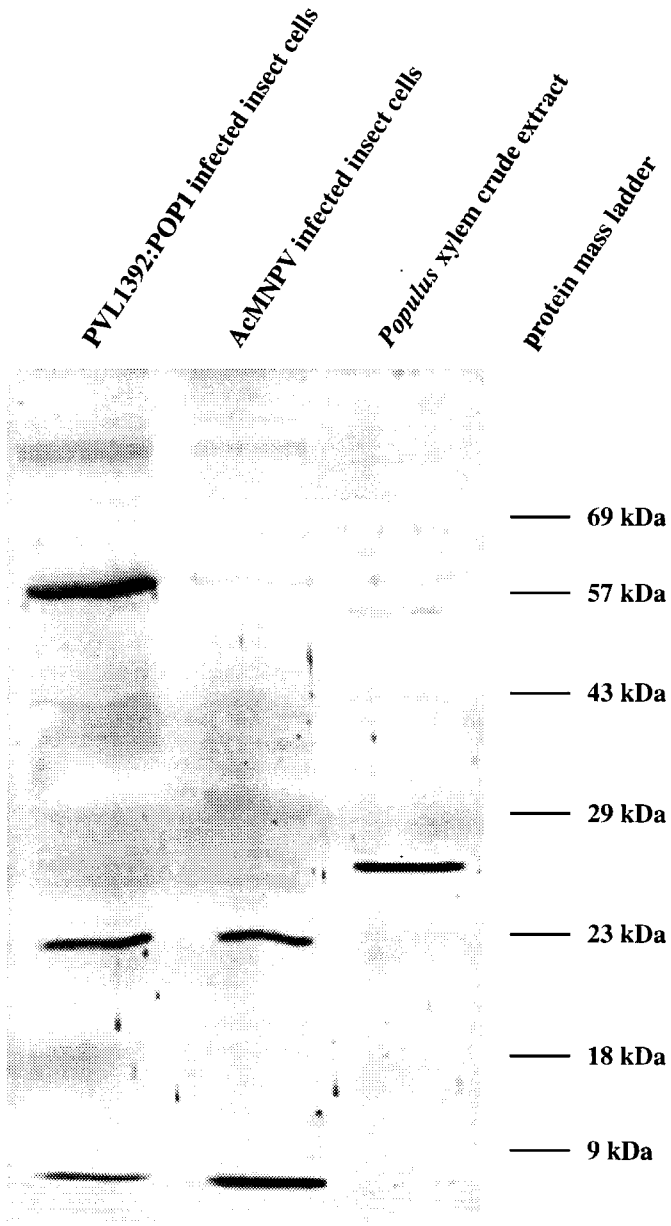


Figure 5.3 Western blot analysis of recombinant POP1 and *Populus* developing xylem. Crude protein extracts were detected with a coniferin β-glucosidase primary antibody.

present in the rabbit serum (other than those specific to the pine β-glucosidase) to insect cell proteins.

Protein extracts from poplar developing xylem (harvested May 1996) were also examined by western blot analysis. The pine β-glucosidase antibody strongly cross-reacted with a 27 kDa protein present in the protein extract from poplar xylem. A fainter immuno-reactive band with a molecular weight of about 55 kDa was also observed. It is not known whether the bands observed in these extracts correspond to β-glucosidases or to other unrelated proteins present in the extract.

5.3.4 Enzymatic Characterization of the POP1 Recombinant Protein

Assays carried out with the recombinant POP1 protein and varying concentrations of test substrates (coniferin and 4-nitrophenyl β-glucoside) did not provide any insights as to the enzymatic activity of POP1. Crude protein extracts from insect cells expressing the POP1 recombinant protein did not show activity against either substrate over a range of concentrations (1-4mM) when compared to crude protein extracts from cells infected with wild type baculovirus. Either the recombinant POP1 does not have activity against these substrates, or the recombinant protein is being expressed in an inactive form.

5.3.5 Sequencing of the POP1:pVL1392 Expression Construct

The *POP1* fragment used for expression was sequenced in its entirety to assess the fidelity of the coding region PCR amplification. DNA sequencing determined that the *POP1* sequence used for expression contained four nucleotide changes when compared to the originally determined POP1 sequence. These changes resulted in four

amino acid changes in the derived polypeptide. The changes were as follows: cDNA nt. #449, A→T resulting in an amino acid change from Asn→Thr; cDNA nt. #918, T→C resulting in an amino acid change from Ile→Thr; cDNA nt. #1046, T→C resulting in an amino acid change from Ser→Pro and cDNA nt. #1562, A→G resulting in an amino acid change from Lys→Ala. It is not known whether the observed changes are due to PCR amplification errors or to allelic variation between *POP1* transcripts.

5.4 Discussion

Northern blot and RT-PCR studies suggest that *POP1* is expressed in the young leaves of poplar but not in the spring developing xylem. During May and June, poplar undergoes a dramatic increase in biomass that includes shoot elongation, new leaf formation, and an increase in tree girth (i.e. branches and the main trunk of the tree). Newly formed tissue becomes heavily lignified during this period and an up-regulation of phenylpropanoid enzymes involved in lignification (e.g. PAL, C4H and 4CL) occurs. Based on these observations, it seemed that this would be the most likely time to detect the transcription of a β-glucosidase involved in lignification. Furthermore, the gene encoding the pine coniferin β-glucosidase, which is believed to be involved in the lignification of *Pinus contorta*, is known to be actively transcribed during this growth period (data not shown). However, the results from expression experiments conducted in this study do not indicate that *POP1* transcripts are present in developing xylem, suggesting that POP1 may not be involved in the lignification pathway of poplar.

Several *family 1* plant β-glucosidases found primarily in the leaves are implicated

in the mobilization of preformed defense responses. *Trifolium repens* (white clover), *Manihot esculenta* (Cassava), *Linum usitatissimum* (Flax), *Hevea braziliensis* (Rubber tree), *Phaseolus lunatus* (Lima bean) and *Lotus* species contain the cyanogenic glucosides lotaustralin and linamarin (Hughes *et al.* 1992). The leaves of these species also contain a β-glucosidase that is responsible for the hydrolysis of lotaustralin and linamarin. Because β-glucosidases and substrates are found in distinct cellular or subcellular locations in the cell, it is only upon tissue damage that hydrolysis occurs. Hydrolysis, in combination with the activity of a nitralase, results in the formation of HCN that is believed to function as a feeding deterrent for small herbivores. Similarly, in the Brassicaceae family, leaf tissue contains spatially separated non-toxic glucosinilates and the β-glucosidase (myrosinase) necessary for their breakdown to toxic defense molecules (Thangstad *et al.* 1993). In *Avena sativa*, an analogous defense system exists in which saponins stored in the leaves are hydrolyzed by a specific β-glucosidase only upon tissue damage (Gus-Mayer *et al.* 1994). However, in this case, the toxic hydrolysis products appear to be antifungal compounds rather than herbivore deterrents.

The results from this study do not provide any clarification of the role that POP1 plays in poplar physiology, but they do raise questions about it's potential involvement in lignification. If POP1 is a leaf-specific enzyme, it's true role may be in a preformed defense response as is observed with several other plant *family 1* β-glucosidases. Since poplar leaves contain a large variety of phenolic glucosides of unknown function, any one (or more) of which could serve as a potential substrate for POP1, a detailed analysis of POP1 enzyme activity against potential *in vivo* substrates may help to determine the

function of POP1 in poplar leaves.

SDS-Page and Western blot data further supported the hypothesis that the POP1 cDNA cloned from poplar young leaves encodes a *family 1* β-glucosidase. In particular, the cross-reaction observed between the heterologously expressed POP1 protein and the pine coniferin β-glucosidase antibody demonstrates that the pine and poplar β-glucosidases share common epitopes despite there being considerable amino acid sequence divergence between these enzymes. However, the pine coniferin glucosidase antibody cross-reacted relatively weakly to the POP1 protein suggesting that although these protein share similar epitopes, there are likely significant differences in their overall tertiary conformation.

The recognition of a particular plant *family 1* β-glucosidase by antibodies raised to a different plant *family 1* β-glucosidase is not unprecedented in the literature. In black cherry (*Prunus serotina*), antisera obtained by immunizing rabbits with deglycosylated amygdalin hydrolase (*family 1* β-glucosidase) not only recognized amygdalin hydrolase isozymes, but also reacted strongly towards prunasin hydrolase, a different *family 1* β-glucosidase present in the same species (Ping Li *et al.* 1992). In addition, antisera against a linamarase (*family 1* β-glucosidase) in white clover (*Trifolium repens*) was observed to cross-react with other unidentified β-glucosidases present in the same species (Kakes 1985). A cross-reaction between maize-anti-β-glucosidase sera and dhurrinases from *Sorghum bicolor* L. Moench has also been reported (the amino acid sequence identity between the maize β-glucosidase to which the sera was raised and the dhurrinases is approximately 70%) (Cicek and Esen 1998). These results are consistent with the

concept that β-glucosidases within a species, and among species, share conserved structural features. Furthermore, the data presented in this study demonstrate that structural conservation is found between *family 1* β-glucosidases in species as distantly separated as poplar and pine.

Additional cross-reactivity of the coniferin β-glucosidase antibody to poplar xylem proteins was observed. In particular, a 27kDa band in the *Populus* xylem extract strongly cross-reacted with the pine antibody. This band is smaller than excpected for a *family 1* β-glucosidase but it is interesting to note that the pine β-glucosidase, when analysed by SDS-PAGE/western blot, appears as two smaller bands (29 kDa and 26kDa) and two larger bands (52kDa and 41kDa) (Dharmawardhana *et al.* 1999). The smaller bands are presumed to arise from protease cleavage. Perhaps a similar event occurred with the poplar xylem extract analysed here, suggesting that investigation into the identity of the 27kDa band may be valuable.

The recombinant POP1 enzyme did not show activity *in vitro* against either 4-nitrophenyl β-glucoside (4-NPG) or coniferin. 4-NPG, a synthetic, chromogenic glucoside, is frequently used to detect β-glucosidase activity since it is hydrolyzed by most β-glucosidases. Furthermore, all characterized plant *family 1* β-glucosidases are capable of hydrolyzing 4-NPG, although with varying efficiencies. Coniferin was also chosen as a test substrate to assess the possible role of POP1 in lignification. The inability of recombinant POP1 to hydrolyze these test substrates suggests that either POP1 is being expressed in an inactive form or POP1 is incapable of hydrolyzing either of these compounds.

The *POP1* coding region was sequenced in its entirety to determine if PCR amplification errors were responsible for recombinant POP1's apparent lack of enzyme activity. Upon sequencing, four nucleotide changes were observed between the originally amplified *POP1* fragments, and the coding region of *POP1* amplified for protein expression. These differences result in four amino acid substitutions. *Taq* DNA polymerase does not possess proofreading capability and has been reported to incorporate an incorrect nucleotide in approximately every $1x10^5$ bases it amplifies (Cline *et al* 1996). It is possible that the four nucleotide substitutions are a result of PCR error.

However, allelic variation among *POP1* transcripts from which the young leaf library was created could also be responsible for the observed nucleotide differences. Poplar H-11 trees are derived from a cross between the species *Populus deltoides* and *Populus trichocarpa*. Genetic variation between the *POP1* genes in the parent species would result in allelic variation in their H-11 offspring and may account for the nucleotide differences observed between the originally amplified *POP1* sequence and the *POP1* coding region used for heterologous expression. Alternatively, gene duplication may have given rise to a second, highly homologous copy of *POP1* in the *Populus* genome. The nucleotide variation observed in the *POP1* coding sequence, whether derived from PCR errors, allelic variation or gene duplication, may provide an explanation as to why the recombinant POP1 protein does not show enzyme activity against 4-NPG or coniferin.

Another possible explanation as to why the recombinant POP1 protein appears to be inactive is that it was expressed as a cytoplasmic protein in insect cells (no signal peptide) whereas it appears to be expressed as a secreted protein *in planta* (contains a

signal peptide). The baculovirus expression vector used in this study (pVL1392) does not have the option of fusing a leader sequence to the expressed protein for targeting to the secretion pathway. By synthesizing POP1 as a cytoplasmic protein, none of the modifications associated with the secretory pathway, perhaps necessary to synthesize a functional POP1 protein, would have occurred (e.g. glycosylation, disulphide bond formation, myristoylation). However, plant *family 1* β-glucosidases from *Pinus contorta*, *Zea mays* and *Costus speciosus* have been successfully expressed in bacteria (Inoue *et al.* 1996, Dharmawardhana *et al.* 1995, Cicek and Esen 1999). These recombinant enzymes showed no differences in substrate preference or enzyme activity when compared to their purified counterparts, despite the absence of eukaryotic post-translational modifications. Because of these precedents, it was assumed that modifications that normally occur to secreted β-glucosidase proteins (usually glycosylation) were not a requirement for *in vitro* enzyme activity. Perhaps, however, the POP1 enzyme, unlike other plant *family 1* β -glucosidases, does require specific post-translational modifications in order to be expressed as an active enzyme.

Several other general problems with assay conditions or protein preparations could be envisioned to effect POP1 activity. For instance, insect cells may contain molecules that inhibit POP1 activity. The chosen assay conditions (e.g. pH, temperature or lack of necessary co-factors) may also have affected POP1 activity. However, the standard assay conditions chosen in this study have been used to detect β-glucosidase activity in enzymes ranging in source from bacteria to plants. These assay conditions include: a pH of 5.5 (characterized *family 1* β-glucosidases are generally most active

somewhere between pH 4.5 and 6.5), an assay temperature of 37° (characterized *family 1* β-glucosidases are active at temperatures ranging from 25°-50°) and no cofactor additions (characterized *family 1* β-glucosidases do not appear to require co-factors). Although the POP1 enzyme could require unique reaction conditions outside these standard requirements, it seemed unlikely that this could account for the total absence of POP1 activity.

It is also possible that the POP1 protein was expressed in an active form but was not being assayed with a substrate which it is capable of hydrolyzing (i.e. it's natural substrate). Poplar species are known to contain an extraordinary range of phenolic glycosides. In particular, the ethyl acetate extract of the hot-water-soluble portion of an 80% ethanol extractive of leaves have been shown to contain at least eight different phenolic glucosides (salicin, salireposide, trichocarpin, populin, salicortin, tremuloidin, tremulacin and rutin) (Pearl and Darling 1970, Dommisse *et al.* 1985). The function of most of these molecules is unknown, although salicin and salireposide exhibit antiviral properties against polio and Semliki forest virus (Van Hoof *et al.* 1989). Any of these glycosides could be the natural substrate for POP1. An interesting experiment that is being pursued outside the scope of this thesis is the investigation of POP1 activity against a crude preparation of glycosides isolated from poplar leaf tissue. This experiment may show that the recombinant POP1 protein has been expressed as an active enzyme and may give an idea of its natural substrate(s).

Chapter 6
Conclusions and Future Directions

Two interesting characteristics of *family 1* β-glucosidases deserve further comment within this thesis. First, each individual β-glucosidase enzyme can hydrolyze a wide range of glucosides and, second, plant β-glucosidases are almost all targeted to the secretory system. Based on this information, an interesting hypothesis about the mode of action for β-glucosidases can be proposed.

Most research into plant *family 1* β-glucosidases has been motivated by interest in a particular glucoside that is abundant within a plant species and is of interest to humans. For instance, the roots of cassava (*Manihot esculenta*) become cyanogenic when tissue damage results in the breakdown of the glucosides lotaustralin and linamarin (Hughes *et al.* 1992). However, despite the health hazard associated with this food source, it is a staple carbohydrate for many communities in the tropics. Interest in the biochemistry behind the hydrolysis of lotaustralin and linamarin led to the isolation of a *family 1* β-glucosidase (linamarase) that was capable of hydrolyzing these cyanogenic glucosides.

Purified linamarase can hydrolyze a wide range of glucosides, however, including linamarin, 4-nitrophenyl β-glucoside, 2-nitrophenyl β-galactoside, prunasin and cellobiose. Some of these substrates are synthetic and do not represent natural 'cassava' glucosides, but they do highlight the aglycone variability that is tolerated by this enzyme. Furthermore, this substrate 'wobble' is seen with all plant *family 1* β-glucosidases that have been characterized in detail.

However, despite the observed 'wobble' with synthetic substrates, there are no

reports in the literature examining the potential for substrate 'wobble' among natural glucosides relevant to the species from which the β-glucosidase in question was identified. This experiment could reveal that within one species, plant *family 1* β-glucosidases can hydrolyze several glucosides. If this is the case, it may be that substrate usage in the cell is determined by the co-localization of β-glucosidase and glucoside, not by the enzyme's active site substrate binding constraints (i.e. spatial regulation instead of substrate specificity regulation). This model would predict that linamarase is capable of hydrolyzing several cassava glucosides in addition to linamarin.

If this model were correct, β-glucosidases would present an interesting system to examine how enzymes are sequestered from, but co-localized with, their substrates. It is known that almost all plant *family 1* β-glucosidases contain signal peptides that target the newly synthesized proteins to the secretory system. This suggests that β-glucosidases can be directed to discrete locations in the cell, but where and when does the enzyme come in contact with its glucoside substrates? In cassava, β-glucosidases and their substrates (lotaustralin and linamarin) are separated within the cell and mechanical damage is thought to be responsible for their co-localization. However, as more and more data becomes available on the complexity of intracellular trafficking it is tempting to speculate that in other situations a more elegant process, involving membrane bound vesicles and protein scaffolding, could be involved in coordinating enzyme-substrate co-localization.

The research presented in this thesis does not allow for analysis of these proposed ideas but if enzyme activity can be achieved with recombinant POP1, poplar would make

an excellent species to examine these possibilities. The leaves of poplar contain in excess of eight phenolic glucosides. It would be interesting to determine whether POP1 is specific for only one of these glucosides, or if, like other plant *family 1* β-glucosidases acting on synthetic substrates, POP1 can hydrolyze several of the naturally occurring glucosides.

This study has identified the first three partial sequences from *family 1* β-glucosidase genes of poplar (*POP1, POP2* and *POP3*). Additionally, the cDNA of one of these genes (*POP1*) was cloned, sequenced and shown to be expressed in young leaf tissue. An inability to detect recombinant POP1 activity against coniferin, in combination with the absence of *POP1* transcript in developing xylem tissue, has brought into question the involvement of POP1 in lignification. Since *in vivo* natural substrate(s) of POP1 are unknown the role of POP1 in poplar physiology remains speculative.

However, *family 1* β-glucosidase research has begun to highlight the size and complexity of this ancestral enzyme family that contains representatives from all known kingdoms. Furthermore, plant species contain multiple family members (at least three in *Populus*, but likely more if the identification of eighteen unique β-glucosidase genes in the arabidopsis genome is representative of other species). Further study of this enzyme family will likely touch diverse areas of metabolism and provide exciting information about glucoside usage in plant physiology.

The data presented in this thesis does not contribute strong evidence for the existence of a monolignol glucoside transport system in poplar. However, this work has only begun to characterize the *family 1* β-glucosidases of poplar and in no way excludes

the exsistence of a coniferin specific β-glucosidase in this species. Of particular intrest is the recent identification of a *family 1* β-glucosidase in poplar developing xylem by EST sequencing. If further work is pursued on monolignol transport this gene would be an excellent candidate for investigation. Furthermore, a *family 1* β-glucosidase has been identified by a differential screen of xylem transcripts from stressed wood (compared to no stressed wood) (A. Plovanich-Jones, personal communication). It is not yet known whether the gene that is upregulated in stressed wood is the same as the EST clone, *POP1*, *POP2* or *POP3*, or if it is novel. These results highlight that several avenues remain to be explored before the hypothesis of monolignol transport via monolignol glucosides can be rejected.

References

Altschul SF, Gish W, Miller W, Myers EW and Lipman DJ (1990) Basic local alignment search tool. J. Mol. Biol. 215:403-410

Atanassova R, Favet N, Martz F, Chabbert B, Tollier MT, Monties B, Fritig B and Legrand M (1995) Altered lignin composition in transgenic tobacco expressing *O*-methyltransferase sequences in sense and antisense orientation. Plant J. 8:465-477

Barcelo RA, Morales M and Pedreno MA (1998) Specific compartmentalization of peroxidase enzymes in relation to lignin biosynthesis in the plant cell. In Lignin and Lignan Biosynthesis, ACS symposium Series 3697, N.G. Lewis and S. Sarkanen. eds. (Washington, DC: American Chemical Society): pp.84-95

Barnes RF, Muller LD, Bauman LF and Colenbrander VF (1971) *In vitro* dry matter disappearance of brown midrib mutants of maize (*Zea mays* L.). J. Anim. Sci. 33:881-884

Bate NJ, Orr J, Ni W, Meromi A, Nadler-Hassar T, Doerner PW, Dixon RA, Lamb CJ and Elkind Y (1994) Quantitative relationship between phenylalanine ammonia-lyase levels and phenylpropanoid accumulation in transgenic tobacco identifies a rate-determining step in natural product synthesis. Proc. Natl. Acad. Sci. 91:7608-7612

Baucher M, Chabbert B, Pilate G, Van Doorsselaere J, Tollier MT, Petit-Conil M, Cornu D, Monties B, Van Montagu M, Inze D, Jouanin L and Boerjan W (1996) Red xylem and higher extractability by down-regulating a cinnamyl alcohol dehydrogenase in poplar. Plant Physiol. 112:1479-1490

Bause E (1983) Structural requirements of N-glycosylation of proteins. Studies with proline peptides as conformational probes. Biochem. J. 209:331-316

Boudet AM, Lapierre C and Grima-Pettenati J (1995) Biochemistry and molecular biology of lignification. New Phytol. 129:203-236

Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principles of protein dye binding. Anal. Biochem. 72:248-254

Bradshaw HD, Villar M, Watson BD, Otto KG, Stewart S and Stettler RF (1994) Molecular genetics of growth and development in *Populus.* III. A genetic linkage map of a hybrid poplar composed of RFLP, STS and RAPD markers. Theor. Appl. Genet. 89:167-178

Bradshaw HD and Stettler RF (1995) Molecular genetics of growth and development in *Populus*. IV. Mapping QTLs with large effects on growth, form and phenology traits in a forest tree. Genetics 139:963-973

Brzobohaty B, Moore I, Kristoffersen P, Bako L, Campos N, Schell J and Palme K (1993) Release of active cytokinin by a β-glucosidase localized to the maize root meristem. Science 262:1051-1054

Buxton DR and Redfearn DD (1997) Plant limitations to fiber digestion and utilization. J. Nutr. 127:814S-818S

Chapple CCS, Vogt T, Ellis BE, Sommerville CR (1992) An Arabidopsis mutant defective in the general phenylpropanoid pathway. Plant Cell 4:1413-1424

Chrispeels MJ and Raikhel NV (1992) Short peptide domains target proteins to plant vacuoles. Cell 68:613-616

Cicek M and Esen A (1998) Structure and expression of a dhurrinase (beta-glucosidase) from sorghum. Plant Physiol. 116:1469-1478

Cicek M and Esen A (1999) Expression of soluble and catalytically active plant (monocot) beta-glucosidases in *E. Coli*. Biotechnol Bioeng. 63:392-400

Cline J, Braman JC and Hogrefe HH (1996) PCR fidelity of pfu DNA polymerase and other thermostable DNA polymerases. Nuc. Acid Res. 24:3546-3551

Davin LB, et al (1997). Stereoselective bimolecular phenoxy radical coupling by an auxiliary (dirigent) protein without an active center. Science. 275:362-6.

Day AG and Withers SG (1986) The purification and characterization of a beta-glucosidase from *Alcaligenes faecalis*. Biochem. Cell Biol. 64:914-922

Dean JFD, LaFayette PR, Rugh C, Tristram AH, Hoopes JT, Eriksson KL and Merkle SA (1998) Laccases associated with lignifying vascular tissue. In Lignin and Lignan Biosynthesis, ACS symposium Series 3697, N.G. Lewis and S. Sarkanen. eds. (Washington, DC: American Chemical Society): pp.96-108

Dharmawardhana DP, Ellis BE, Carlson JE (1992) Characterization of vascular lignification in *Arabidopsis thaliana*. Can J Bot 70:2238:2244

Dharmawardhana DP, Ellis BE and Carlson JE (1995) A β-glucosidase from lodgepole pine xylem specific for the lignin precursor coniferin. Plant Physiol. 107:331-339

Dharmawardhana DP, Ellis BE and Carlson JE (1999) cDNA cloning and heterologous expression of coniferin beta-glucosidase. Plant Mol. Biol. 40:365-72

Dommisse RA, Van Hoof L and Vlietinck AJ (1985) Structural analysis of phenolic glucosides from Salicaceae by NMR spectroscopy. Phytochemistry 25:1201-1204

Downes et al. (1991) Lignin distribution across tracheid cell walls of poorly lignified wood from deformed copper deficient *Pinus radiata* (D. Don). Wood Sci. Technol. 25:7-14 (1991)

Doyle JJ and Doyle JL (1990) Isolation of plant DNA from fresh tissue. Focus 12:13-15

Fahey GC (1997) Symposium introduction. J. Nutr. 127:809S

Falk A and Rask L (1995) Expression of a zeatin-*O*-glucoside-degrading β-glucosidase in *Brassica napus*. Plant Physiol. 108:1369

Foxon GA (1998) Brown-midrib maize (bm1) - a mutation affecting the cinnamyl alcohol dehydrogenase gene. Plant J. 14:545-553

Freudenberg K and Harkin JM (1963) The glucosides of cambial sap of spruce. Phytochemistry 2:189-193

Goffner D, Campbell MM, Campargue C, Clastre M, Borderies G, Boudet A and Boudet AM (1994) Purification and characterization of cinnamoyl-CoA:NADP oxidoreductase in Eucalyptus gunnii. Plnt Physiol. 106:625-632

Goffner D, Van Doorsselaere J, Yahiaoui N, Samaj J, Grima- Pettenati J and Boudet AM (1998) A novel aromatic alcohol dehydrogenase in higher plants-molecular cloning and expression. Plant Mol. Biol. 36: 755-765

Gus-Mayer S, Brunner H, Schneider-Poetsch HAW, Lottspeich F, Eckerskorn C, Grimm R and Rudiger W (1994) The amino acid sequence previously attributed to a protein kinase or a TCP-1 related molecular chaperone and co-purified with phytochrome is a β-glucosidase. FEBS Lett. 347:51-54

Gus-Mayer S, Brunner H, Schneider-Poetsch HAW and Rudiger W (1994) Avenacosidase from oat: purification, sequence analysis and biochemical characterization of a new member of the BGA family of β-glucosidases. Plant Mol. Biol. 26:909-921

Halpin C, Knight ME, Foxon GA, Campbell MM, Boudet Am *et al.* (1994) Manipulation of lignin quality by down-regulation of cinnamyl alcohol dehydrogenase. Plant J. 6:339-350

Halpin C, Holt K, Chojecki J, Oliver D, Chabbert B, Monties B, Edwards K, Barakate A and Henrick N, Engelbrecht J, Brunak S and Heijne G (1997) Identification of

prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein engineering 10:1-6

Henrissat B (1991) A classification of glucosyl hydrolases based on amino acid sequence similarities. Biochem J. 280:309-316.

Henrissat B and Bairoch A (1993) New families in the classification of glycosyl hydrolases based on amino acid sequence similarity. Biochem J. 293:781-788

Henrissat B and Bairoch A (1996) Updating the sequence-based classification of glycosyl hydrolases. Biochem. J. 316:695-696

Higuchi T and Brown SA (1963) Studies of lignin biosynthesis using isotopic carbon. Can. J. Biochem. Phsiol. 41:621-628

Holton TA and Graham (1991) A simple and efficient method for direct cloning of PCR products using ddT-tailed vectors. Nuc. Acids Res. 191:1156-1158

Hosel W, Surholt E, Borgmann E (1978) Characterization of β-glucosidase isoenzymes possibly involved in lignification from chick pea (*Cicer arietinum* L.) cell suspension cultures. Eur. J. Biochem. 84:487-492

Hosel W and Todenhagen R (1980) Characterization of a β-glucosidase from *Glycine max* which hydrolyses coniferin and syringin. Phytochemistry 19:1349-1353

Hrazdina G and Wagner GJ (1985) Metabolic pathways as enzyme complexes: Evidence for the synthesis of phenylpropanoids and flavanoids on membrane associated enzyme complexes. Arch. Biochem. Biophys. 237:88-100

Hughes MA, Brown K, Pancoro A, Burray BS, Oxtoby E and Hughes J (1992) A molecular and biochemical analysis of the structure of the cyanogenic β-glucosidase (Linamarase) from cassava (*Manihot esculenta* Cranz). Arch. Biochem. Biophys. 295:273-279

Ibrahim RK and Grisebach H (1976) Purification and properties of UDP-glucose:coniferyl alcohol glucosyltransferase from suspension cultures of Paul's Scarlet Rose. Arch. Biochem. Biophys. 176:700-708

Ibrahim RK (1977) Glucosylation of lignin precursors by uridine diphosphate glucose:coniferyl alcohol glucosyltransferase in higher plants. Z.Pflanzenphysiol. 85:253-262

Inoue K, Shibuya M, Yamamoto K and Ebizuka Y (1996) Molecular cloning and bacterial expression of a cDNA encoding furostanol glycoside 26-*O*-β-glucosidase of *Costus speciosus*. FEBS Lett. 389:273-277

Inoue K, Sewalt VJH, Ballance GM, Ni W, Sturder C and Dixon RA (1998) Developmental expression and substrate specificities of Alfalfa caffeic acid 3-*O*-methyltransferase and caffeoyl coenzyme A 3-*O*-methyltransferase in relation to lignification. Plant Physiol. 117:761-770

Joshi CP (1987) An inspection of the domain between putative TATA box and translation start site in 79 plant genes. Nucl. Acid Res. 15:6643-6653

Joshi CP, Zhou H, Huang X and Chiang VL (1997) Context sequences of translation initiation codon in plants. Plant Mol. Biol. 35:993-1001

Jung HG and Allen MS (1995) Characteristics of plant cell walls affecting intake and digestibility of forages by ruminants. J. Anim. Sci. 73:2774-2790

Jung HG (1997) Analysis of forage fiber and cell walls in ruminant nutrition. J. Nutr. 127:810S-813S

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

Kakes P (1985) Linamarase and other β-glucosidases are present in the cell walls of *Trifolium repens* L. leaves. Planta 166:156-160

Kneusel RE, Mattern U and Nicolay K (1989) Formation of trans-caffeoyl CoA from trans-4-coumaryl CoA by Zn^{++} dependent enzymes in cultured plant cells and its activation by an elicitor induced pH shift. Arch. Biochem. Biophys. 269:455-462

Leah R, Kigel J, Svendsen I and Mundy J (1995) Biochemical and molecular characterization of a barley seed β-glucosidase. J. Biol. Chem. 270:15789-15797

Lechtenberg VL, Muller LD, Bauman LF, Rhykerd CL and Barnes RF (1972) Laboratory and in vitro evaluation of inbred and F_2 populations of brown midrib mutants of *Zea mays* L. Agronomy J. 64:657-660

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

Leinhos V and Savidge RA (1993) Isolation of protoplasts from developing xylem of *Pinus banksiana* and *Pinus strobus*. Can. J. For. Res. 23:343-348

Leinhos V, Preethi V, Randeniya U and Savidge RA (1994) Purification of an acidic coniferin-hydrolyzing β-glucosidase from developing xylem of *Pinus banksiana*. Phytochemistry 37: 311-315

Lewis NG and Davin LB (1998) The biochemical control of monolignol coupling and structure during lignan and lignin biosynthesis. In Lignin and Lignan Biosynthesis, ACS symposium Series 3697, N.G. Lewis and S. Sarkanen. eds. (Washington, DC: American Chemical Society): pp.334-361

Li L, Popko JL, Zhang XH, Osakabe K, Tsai CJ, Joshi CP and Chiag VL (1997) A novel multifunctional O-methyltransferase implicated in a dual methylation pathway associated with lignin biosynthesis in loblolly pine. Proc. Natl. Acad. Sci. 94:5461-5466

Liang X, Dron M, Cramer CL, Dixon RA and Lamb CJ (1989) Differential regulation of Phenylalanine ammonia-lyase genes during plant development and by environmental cues. J. Biol. Chem. 264:14486-14492

MacKay JJ, O'Malley DM, Presnell T, Booker FL, Campbell MM, Whetton RW and Sederoff RR (1997) Inheritance, gene expression, and lignin characterization in a mutant pine deficient in cinnamyl alcohol dehydrogenase. Proc. Natl. Acad. Sci. 94:8255-8260

Maher EA, Bate NJ, Ni W, Elkind Y, Dixon RA and Lamb CJ (1994) Increased susceptibility of transgenic tobacco plants with suppressed levels of preformed phenylpropanoid products. Proc. Natl. Acad. Sci. 91:7802-7806

Marcinowski S and Grisebach H (1978) Enzymology of lignification. Eur. J. Biochem. 87:37-44

Martoglio B and Dobberstein B (1998) Signal sequences: more than just greasy peptides. Trends Cell Biol. 8:410-415

Martz F, Maury S, Pincon G and Legrand M (1998) cDNA cloning, substrate specificity and expression study of tobacco caffeoyl-CoA 3-*O*-methyltransferase, a lignin biosynthetic enzyme. Plant Mol. Biol. 36:427-437

Meng H and Campbell WH (1998) Substrate profiles and expression of caffeoyl coenzyme A and caffeic acid *O*-methyltransferases in secondary xylem of aspen during seasonal development. Plant Mol. Biol. 38:513-520

Meyer K, Shirley AM, Cusumano JC, Bell-Lelong DA and Chapple C (1998) Lignin monomer composition is determined by the expression of a cytochrome P450-dependent monooxygenase in Arabidopsis. Proc. Natl. Acad. Sci. 95:6619-6623

Mkpong OE, Yan H, Chism G and Sayre RT (1990) Purification, charaterization and localization of linamarase in cassava. Plant Physiol. 93:176-181

Nakamura K and Matsuoka K (1993) Protein targeting to the vacuole in plant cells. Plant Physiol. 101:1-5

Neish AC (1961) Formation of *m*- and *p*-coumaric acids by enzymatic desamination of the corresponding isomers of tyrosine. Phytochemistry 1:1-24

Nielsen H, Engelbrecht J, Brunak S and Von Heijne G (1997) Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Engineering 10:1-6

Nisius A and Ruppel HG (1987) Immunocytological and chemical studies on the stromacentre-forming protein from *Avena* plastids. Planta 171:443-452

Pallas JA, Paiva NL, Lamb CJ and Dixon RA (1996) Tobacco plants epigenetically suppressed in phenylalanine ammonia-lyase expression do not develop systemic acquired resistance in response to infection by tobacco mosaic virus. Plant J. 10:281-293

Pearl IA and Darling SF (1970) Hot water phenolic extractives of the bark and leaves of diploid *Populus tremuloides*. Phytochemistry 10:483-484

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

Ralph J, MacKay JJ, Hatfield RD, O'Malley DM, Whetton RW and Sederoff RR (1997) Abnormal lignin in a loblolly pine mutant. Science 277:235-239

Rasmussen S. and Dixon RA (1999) Transgene-mediated and elicitor-induced perturbation of metabolic channeling at the entry point into the phenylpropanoid pathway. Plant Cell. 11(8):1537-1552.

Rosler J, Krekel F, Amrhein N, Schmid J (1997) Maize phenylalanine ammonia-lyase has tyrosine ammonia-lyase activity. Plant Physiol. 113:175-179

Rusch SL and Kendall DA (1995) Protein transport via amino-terminal targeting sequences: common themes in diverse systems (Review). Mol. Mem.Biol. 12:295-307

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

Sanz-Aparicio J, Hermoso JA, Martinez-Ripoll M, Lequerica JL and Polaina J (1998) Crystal structure of a beta-glucosidase A from *Bacillus polymyxa*: insights into the catalytic activity in family 1 glycosyl hydrolases. J. Mol. Biol. 275:491-502

Savidge RA, Randeniya PV, Xu Y, Leinhos V and Forster H (1998) Coniferyl alcohol oxidase: a new enzyme spatio-temporally associated with lignifying tissue. In Lignin and Lignan Biosynthesis, ACS symposium Series 3697, N.G. Lewis and S. Sarkanen. eds. (Washington, DC: American Chemical Society): pp.109-130

Sewalt VJH, Ni W, Blount JW, Jung HG, Masoud SA, Howles PA, Lamb CJ and Dixon RA (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. 115:41-50

Schmid G and Grisebach H (1982) Enzymatic synthesis of lignin precursors. Eur. J. Biochem. 123:363-370

Sinnott M (1990) Catalytic mechanisms of glycosyl transfer. Chem. Rev. 90:1171-1202

Solovyev VV, Salamov AA and Lawrence CB (1994) Predicting internal exons by oligonucleotide composition and discriminant analysis of spliceable open reading frames. Nucl. Acid Res. 22:5156-5163

Somssich IE, Wernert P, Kierdrowski S and Hahlbrock K (1996) *Arabidopsis thaliana* defense-related protein ELI3 is an aromatic alcohol:$NADP^+$ oxidoreductase. Proc. Natl. Acad. Sci. 93:14199-14203

Srivastava DK and Bernhard SA (1986) Metabolite transfer via enzyme-enzyme complexes. Science 234:1081-1086

Sterky F et al. (1998) Gene discovery in the wood forming tissues of poplar: Analysis of 5692 expressed sequence tags. Proc. Natl. Acad. Sci. 95:13330-13335

Surholt E and Hosel W (1981) Identification of coniferin-specific-glucosidases *in Cicer arietinum* L. seedlings. Z. Pflanzenphysiol. 101:207-214

Terashima N (1989) An improved radiotracer method for studying formation and structure of lignin. In Plant Cell Wall Polymers: Biogenesis and Biodegradation, ACS symposium Series 399, N.G. Lewis and M.G. Paice. eds. (Washington, DC: American Chemical Society): pp.148-159

Terashima N and Fukushima K (1989) Biogenesis and structure of macromolecular lignin in the cell wall of tree xylem as studied by microautoradiography. In Plant Cell Wall Polymers: Biogenesis and Biodegradation, ACS Symposium Series 399, N.G. Lewis and M.G. Paice. Eds. (Washington, DC: American Chemical Society): pp.160-168

Terazawa M, Okuyama H and Miyake M (1984) Phenolic compounds in living tissue of woods I. Mokuzai Gakkaishi 30:322-328

Teutsch HG, Hasenfratz MP, Lesot A, Stoltz C, Garnier JM, Jeltsch JM, Durst F and Werck-Reichhart D (1993) Isolation and sequence of a cDNA encoding the Jerusalem artichoke cinnamate 4-hydroxylase, a major plant cytochrome P450 involved in the general phenylpropanoid pathway. Proc. Natl. Acad. Sci. 90:4102-4106

Thangstad OP, Winge P, Husebye H and Bones A (1993) The myrosinase (thioglucoside glucohydrolase) gene family in Brassicaceae. Plant Mol. Biol. 23:511-524

Thompson JD, Higgins DG and Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignments through sequence weighting, position-specific gap penalties and weight matrix choice. Nucl. Acid Res. 22:4673-4680

Trimber DE, Warren RAJ and Withers SG (1992) Region-directed mutagenesis of residues surrounding the active site nucleophile in β-glucosidase from *Agrobacterium faecalis*. J. Biol. Chem. 267: 10248-10251

Uden W, Pras N, Batterman S, Visser JF and Malingre TM (1990) The accumulation and isolation of coniferin from a high-producing cell suspension of *Linum flavum* L. Planta 183:25-30

Van Doorsselaere J, Baucher M, Chognot E, Chabbert B, Tollier M-T, *et al.* (1995) A novel lignin in poplar trees with reduced caffeic acid/ 5-hydroxyferulic acid *O*-methyltransferase activity. Plant J. 8:855-864

Van Hoof L, Totte J, Corthout J, Pieters LA, Mertens F, Vanden Berghe DA and Vlietinck AJ (1989) Plant antiviral agents, VI. Isolation of antiviral phenolic glucosides from *Populus* cultivar beaupre by droplet counter-current chromatography. J. Nat. Prod. 52:875-878

Vignols F, Rigau J, Torres MA, Capellades M and Puigdomenech P (1995) The *brown mirib*3 (bm3) mutation in maize occurs in the gene encoding caffeic acid *O*-methyltransferase. Plant Cell 7:407-416

Wagner GJ and Hrazdina G (1984) Endoplasmic reticulum as a site of phenylpropanoid and flavanoid metabolism in *Hippeastrum*. Plant Physiol. 74:901-906

Wallace G and Fry SC (1994) Phenolic components of the plant cell wall. Int. Rev. Cytol. 151:229-267

Wang Q, Trimbur D, Graham R, Warren RAJ and Withers SG (1995) Identification of the acid/base catalyst in Agrobacterium faecalis β-glucosidase by kinetic analysis of mutants. Biochemistry 34:14554-14562

Weirsma PA (1996) Molecular cloning and nucleotide sequence of a β-glucosidase cDNA from ripening cherry fruit. Plant Physiol. 110:337

Withers SG, Warren RAJ, Street IP, Rupitz K, Kempton JB and Aebersold R (1990) Unequivocal demonstration of the involvement of a glutamate residue as a nucleophile in the mechanism of a "retaining" glycosidase. J. Am. Chem Soc. 112:5887-5889

Withers SG and Aebersold R (1995) Approaches to labeling and identification of active site residues in glycosidases. Protein Science 4:361-372

Whetton R and Sederoff R (1995) Lignin Biosynthesis. Plant Cell 7:1001-1013

Whetten RW, MacKay JJ and Sederoff RR (1998) Recent advances in understanding lignin biosynthesis. Annu. Rev. Plant Physiol. Plant Mol. Biol. 49:585-609

Ye ZH, Kneusel RE, Matern U and Varner JE (1994) An alternative methylation pathway in lignin biosynthesis in *Zinnia*. Plant Cell 6:1427-1439

Ye ZH and Varner JE (1995) Differential expression of two *O*-methyltransferases in lignin biosynthesis in *Zinnia elegans*. Plant Physiol. 108: 459-467

Ye ZH (1997) Association of caffeoyl coenzyme A 3-*O*-methyltransferase expression with lignifying tissues in several dicot plants. Plant Physiol. 115:1341-1350

Zheng L and Poulton JE (1992) Temporal and spatial expression of amygdalin hydrolase and (R)-(+) -mandelonitrile lyase in black cherry seeds. Plant Physiol. 100:282-290

Zhong R, Morrison, WH, Negrel J and Ye ZH (1998) Dual methylation pathways in lignin biosynthesis. Plant Cell 10:2033-2045

Zhou C, Yang Y and Jong AY (1990) Mini-prep in ten minutes. Biotechniques 8:172-173