Final Technical Report
Department of Energy Biosciences Research Program

DOE Project No: DE-FG02-00ER15071
Project Title: "The control of lignin synthesis"
Principle Investigator: John E. Carlson, The Pennsylvania State University
Project Period: August 15, 2000 to August 14, 2004

Executive Summary:
In this project we tested the hypothesis that regulation of the synthesis of lignin in secondary xylem cells in conifer trees involves the transport of glucosylated lignin monomers to the wall of xylem cells, followed by de-glucosylation in the cell wall by monolignol-specific glucosidase enzymes, which activates the monomers for lignin polymerization. The information we gathered is relevant to the fundamental understanding of how trees make wood, and to the applied goal of more environmentally friendly pulp and paper production. We characterized the complete genomic structure of the Coniferin-specific Beta-glucosidase (CBG) gene family in the conifers loblolly pine (Pinus taeda) and lodgepole pine (Pinus contorta), and partial genomic sequences were obtained in several other tree species. Both pine species contain multiple CBG genes which raises the possibility of differential regulation, perhaps related to the multiple roles of lignin in development and defense. Subsequent projects will need to include detailed gene expression studies of each gene family member during tree growth and development, and testing the role of each monolignol-specific glucosidase gene in controlling lignin content.

Project Rationale:

Lignin is the component of plant cell walls that provides the strength and rigidity that is characteristic of wood and of vessels in plants that transport water from roots to leaves. Lignin also plays an important role in defense of plants against attack by pests. Lignin is a complex three-dimensional organic polymer composed of three types of subunits ("monomers"): p-coumaryl-, coniferyl- and sinapyl-alcohols. The relative proportion of the different monomers in lignin can vary greatly, which results in differences among tree species in wood quality and in the chemistries needed to produce pulp and paper. How plants regulate the biosynthesis of lignin monomers and the mechanism by which lignin monomers are transported to the cell wall for lignin polymerization are not well understood. Our pre-award results established the presence of a coniferin-specific ß-glucosidase (CBG) in lignifying (wood-forming) pine tissues, for which we obtained a full length cDNA sequence. In this project we tested the model that glucosylation of lignin monomers occurs within xylem cells prior to the transport of monolignols to the cell wall, followed by de-glucosylation in the cell wall by monolignol-specific glucosidase (MG) enzymes, activating the monomers for lignin synthesis. The information that we obtained increases our understanding of how trees make lignin and wood, and provides new opportunities to create specialty tree genotypes tailored to more environmentally friendly pulp and paper production. Our approach involved characterization of the CBG / MG gene family for subsequent detailed expression studies.

Hypothesis:

Lignin monomer glucosylation and deglucosylation regulate the supply of monomers available in the cell wall for lignin polymerization.

Project Goals:

1. Characterize the monolignol-specific glucosidase gene family in conifers.
2. Determine the subcellular localization of coniferin-specific glucosidase expression.
3. Initiate sense- and antisense- suppression experiments in transgenic spruce plants to determine the role of coniferin-specific glucosidases in lignin formation.

Results:

Characterization of the monolignol-specific glucosidase gene family in conifers:

The first phase of this project was to discover if genes for monolignol-specific glucosidase (MG) enzymes exist in other gymnosperm tree species, and if so to determine how much conservation of structure of MG genes there is among conifers and what features in the gene or in protein domains distinguished MGs from other members of the family-1 glucosidases. To accomplish this we conducted cloning and comparative sequence analysis of genomic copies of MG genes using the sequence that we had previously obtained from a 1.9 Kb full length cDNA clone for coniferin-specific ß-glucosidase (CBG) from lodgepole pine (Dharmawardhana, D.P., Ellis, B.E. and Carlson, J.E. 1999. cDNA cloning and heterologous expression of coniferin ß-glucosidase. Plant Molecular Biology. 40(2): 365 – 372).

We first conducted Southern hybridizations with cDNA probe against various genomic DNAs. The blots revealed that various conifers, including other pines, spruces, and fir species, had DNA sequences homologous to coniferin-specific ß-glucosidase (CBG). We then obtained the complete, full length genomic sequences of CBG genes from Lodgepole pine (*Pinus contorta*) and from loblolly pine (*Pinus taeda*), including the complete 5' and 3' untranslated regions of the genes, in the following manner. We first obtained fragments of the CBG gene from these species by amplification of genomic DNA using PCR primers based on our lodgepole pine cDNA sequence. The PCR primers were designed, with third base degeneracy, to amplify no more than 600 bp of cDNA sequence so that we could capture any introns present in the genomic DNA. The primers also provided overlapping ends for assembly of full length genomics sequence from the several separate PCR products (see Figure 1 for primer design). Rather than one PCR product per cDNA primer pair, we often obtained two or more fragments, each of which were cloned into a plasmid vector (Figure 2). We sequenced multiple copies (8 or more) of each amplified gene fragment to detect PCR and sequencing artifacts. We then developed consensus sequences for each gene fragment,

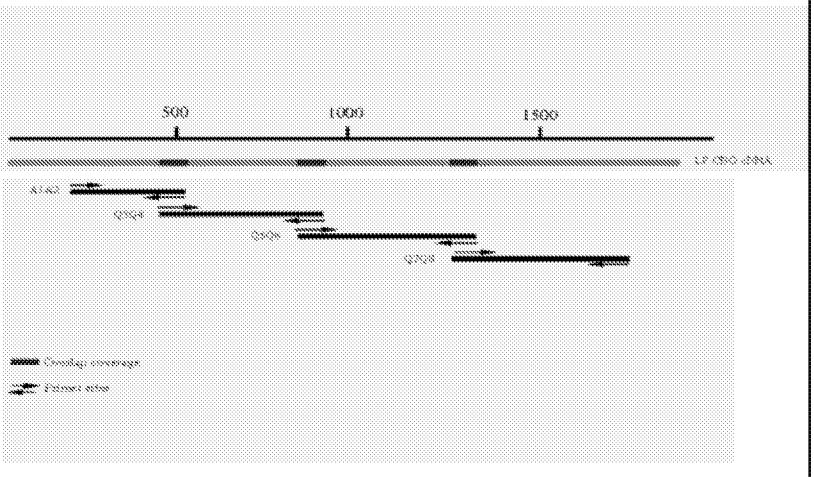

Figure 1. Design of overlapping PCR primer pairs from the fell length cDNA for genomic DNA amplification.

assembled the genes using the overlapping ends and then used the Spidey program from NCBI (http://www.ncbi.nlm.nih.gov/IEB/Research/Ostell/Spidey/index.html) to determine the intron structure of the CBG genes. The position of introns was also confirmed by BLAST alignment of the genomic sequences against the CBG cDNA sequence.

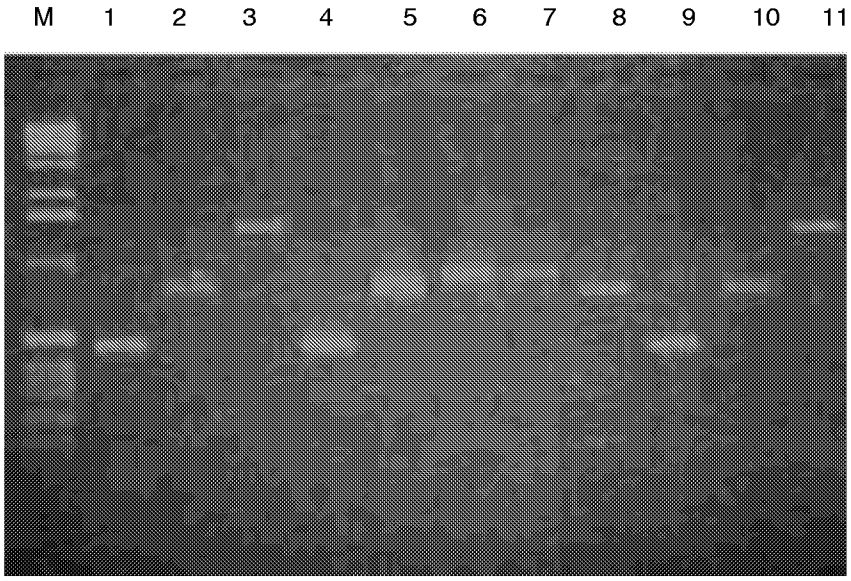

Figure 2. Agarose gel electrophoretogram of genomic DNA samples amplified with PCR primer pairs from the lodgepole pine cDNA sequence. M: Lamda/HinIII+EcoR; Lane 1-5: PCR products from lodgepole pine; Lane 6-11: PCR products from loblolly pine.

Our results for the structure of the *Pinus contorta* and *Pinus taeda* genes, indicate that at least two versions of CBG genes are present in conifers, which we refer to as the cbgA and cbgB genes (Figure 3). There appears to be one copy of each gene in Lodgepole pine and two copies of each gene in loblolly pine. The structure of the CBG genes is rather complex, with multiple introns in both 5' and 3' portions of the genes. The position and length of the introns and exons distinguishes A from B type genes. The homology of the pine cDNA sequence was very high (>98%) to the exons in the genomic DNA clones from both pine species. All the exons from the six genes are very similar in sequence (the average identity is 98.7%), while the introns varied in size and number. *Pinus taeda* and *P. contorta* appear to both have two basic types of CBG genes – cbgA and cbgB. The cbgA genes have three more introns in the C-terminal end of the gene than do the cbgB genes. *Pinus taeda* has two copies of the cbgA genes (PtcbgA1 and PtcbgA2) and two of the cbgB genes (PtcbgB1 and PtcbgB2), however, while *P. contorta* carries only one copy of each (PccbgA and PccbgB). All copies of the cbgA genes (PccbgA, PtcbgA1 and PtcbgA2) share the same number introns at the same positions, while all versions of the cbgB genes (PccbgB, PtcbgB1 and PtcbgB2) have the same intron/exon organization. The two copies of cbgA and cbgB genes in *Pinus taeda* differ in size, however, due to a larger eighth intron in the PtcbgA2 and PtcbgB2 genes versus the PtcbgA1 and PtcbgB1 genes.

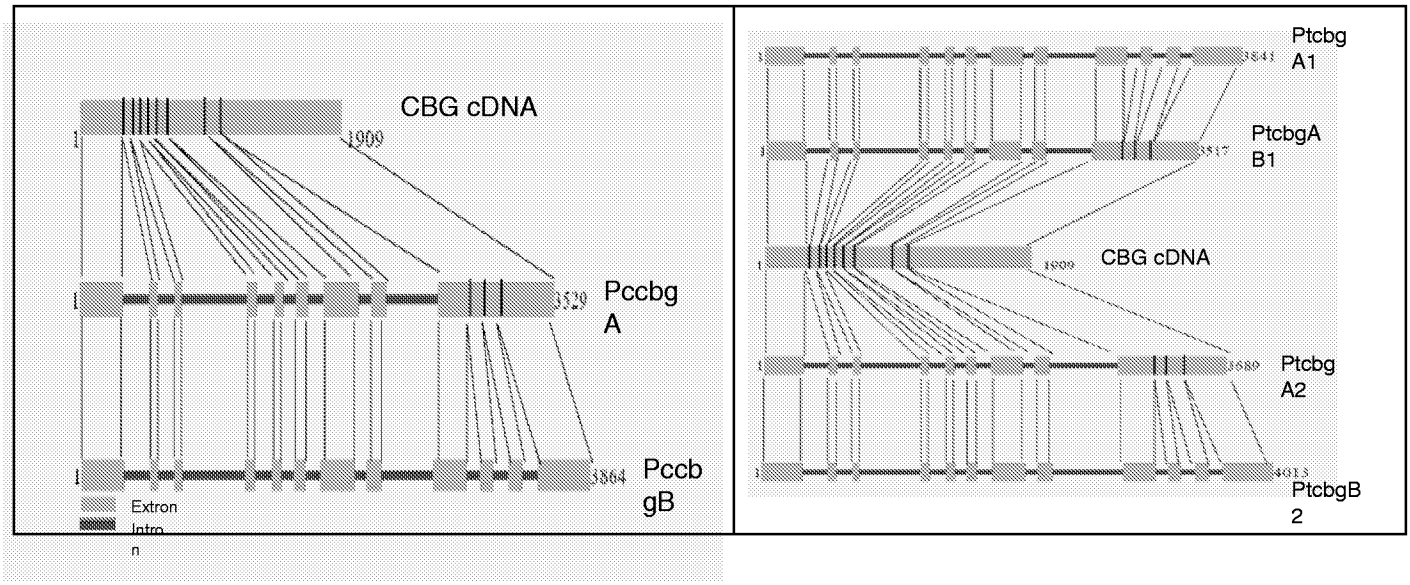

Figure 3. The structure of the pine CBG gene family. *Pinus contorta* (left) contains one copy each of two CBG genes (cbgA and cbgB) while *Pinus taeda* (right) contains two copies of each. Mauve bars, exons; Blue lines, introns; Gene lengths in base pairs.

We also obtained and assembled the partial genomic sequences from interior spruce (*Picea glauca x engelmannii*), Sitka spruce (*Picea sitchensis*), Douglas-fir (*Pseudotsuga menziesii*), jack pine and white pine. Although our sequences for the CBG genes in interior spruce, Sitka spruce, Douglas-fir, jack pine and white pine are incomplete, the CBG genes in these species also appear to be present in a two-gene family structure similar to *Pinus contorta* and *Pinus taeda*. Completion of the CBG genes for these species will require genome walking to fill gaps where the PCR primers from the lodgepole pine cDNA do not amplify.

The confirmation and extension of the CBG gene family to the six additional species provides support to our hypothesis that CBG has a conserved role in monolignol synthesis and transport among conifers. However, using the heterologous PCR primer approach to clone the genomic copies of the CBG gene was difficult and extremely time consuming even among these closely-related Pinaceae tree species. The heterologous PCR primer approach has not worked as well for us for some of the more distantly-related tree species that we tried, most likely because of DNA sequence variation at the PCR primer sites.

Determine the subcellular localization of coniferin-specific glucosidase expression.

The second goal for the project, after characterizing the monolignol-specific glucosidase gene family, was to determine the subcellular localization of coniferin-specific glucosidase expression followed by studies on the regulation of expression of the genes to better understand the role of monolignol-glucosidase activity in lignin formation.

Early in the period of this project, however, the subcellular localization of coniferin-specific glucosidase expression was reported by colleagues at the University of British Columbia. They used immunolocalization (Samuels et al, 2002, Planta 216:72-82) to demonstrate that the CBG protein is found only in the secondary cell walls of the cambium and secondary xylem in

Lodgepole pine. Their cell wall localization results supports our original hypothesis that coniferin-β-glucosidase provides a mechanism to de-glucosylate monolignols *in muro*, i.e. after the transport of the monolignols to the cell wall space.

Initiate sense and anti- sense suppression experiments.

To test the role of monolignol-glucosidases in lignin formation, transgenic plants will be created to modify the expression of the CBG genes. The gene family data provide a guide for preparing optimal anti-sense constructs and for the production of gene-specific PCR primers for quantification of the expression of each member of the CBG gene family by RT-PCR. During the first year of this project, the PI's Ph.D. student Madoka Gray-Mitsumune, completed the first gene regulation experiments with CBG in which she transformed white spruce (*Picea glauca)* using anti-sense CBG cDNA sequence and biolistic transformation of somatic embryos. Several transgenic lines were produced and the regenerated plants were assessed for level of endogenous CBG mRNA (northern blots), amount of CBG protein (western blots), CBG enzyme activity (with appropriate cinnamyl alcohol substrates), total lignin content (thioglycolate analysis; acetyl bromide analysis) and lignin composition (GC analysis of nitrobenzene oxidation products). She did not detect any CBG transcripts early in development in the transformed plants, but did observe transcripts and enzyme activity in later stages of transgenic seedling development when lignification would be more likely to occur. Constitutive expression of the anti-sense CBG did not depress activity of native CBG transcription nor did it alter enzyme activity enough to affect lignin content. This may have been due to weak expression of the CBG gene from our anti-sense vector, which used the angiosperm 35S promoter. This was the first known attempt to use the anti-sense technology in a gymnosperm species at the time. Improvements in the technology and in our vector constructs, using information from our gene family study, should improve the results that we obtain from future experiments. We will continue to explore opportunities for cooperating with forest biotechnology industry partners on the testing CBG anti-sense constructs in spruce and pine using their efficient transformation and regeneration systems.

Personnel trained on the project:

Year one: Cyrus Abdmishani, visiting scientist from Tehran University, worked on the project for 8 months but did not make much progress. He did learn valuable research skills in the process, though.

Year two: Song Liu began masters thesis studies with the PI, including course work. His assistantship and tuition were supported by the project.

Years three and four: The project was continued by Song Liu, with assistance from a second graduate student, Yun Wu. Song Liu and Yun Wu generated all of the new genomic NA sequences for conifer CBG gene family members.

Project Outputs

Liu, S. and Carlson, J.E. "Identification And Comparison Of Coniferin B-Glucosidase Genomic Sequences In Lodgepole Pine And Loblolly Pine," Poster at Plant and Animal Genome Conference XII, San Diego, CA, Jan. 10-14, 2004.

Liu, S. 2004. Variation in The Amount of Chinese vs. American Genetic Material Remaining in Resistant Backcross Progeny and Identification and Comparison of Coniferin B-Glucosidase Genomic Sequences in Lodgepole Pine and Loblolly Pine. Master of Science Thesis in Genetics, submitted August 2004, Pennsylvania State University, 68pp.

Liu, S. and Carlson, J.E. Identification and Comparison Of Coniferin B-Glucosidase Genes In Lodgepole Pine and Loblolly Pine. Manuscript in review.

Gray-Mitsumune, M. 2000. Towards Genetic Modification Of the Lignin Biosynthetic Pathway in Interior Spruce (*Picea glauca* x *engelmanni* complex). Ph.D. in Plant Biotechnology, University of British Columbia, 191 pp.