Final Report on

Regulation of Guaiacyl and Syringyl Monolignol Biosynthesis

Sponsor: US Department of Energy
Grant No. DE-FG02-03ER15442

The focus of this research is to understand syringyl monolignol biosynthesis that leads to the formation of syringyl lignin, a type of lignin that can be easily removed during biomass conversion. We have achieved the three originally proposed goals for this project. (1) *SAD* and *CAD* genes (enzyme catalytic and kinetic properties) and their functional relevance to CAld5H/AldOMT pathway, (2) spatiotemporal expression patterns of *Cald5H*, *AldOMT*, *SAD* and *CAD* genes, and (3) functions of *CAld5H*, *AldOMT*, and *SAD* genes *in vivo* using transgenic aspen. Furthermore, we also found that microRNA might be involved in the upstream regulatory network of lignin biosynthesis and wood formation. The achievements are as below.

(1) Based on biochemical and molecular studies, we discovered a novel syringyl-specific alcohol dehydrogenase (SAD) involved in monolignol biosynthesis in angiosperm trees. Through CAld5H/OMT/SAD mediation, syringyl monolignol biosynthesis branches out from guaiacyl pathway at coniferaldehyde;

(2) The function of *CAld5H* gene in this syringyl monolignol biosynthesis pathway also was confirmed *in vivo* in transgenic *Populus*;

(3) The proposed major monolignol biosynthesis pathways were further supported by the involving biochemical functions of CCR based on a detailed kinetic study;

(4) Gene promoter activity analysis also supported the cell-type specific expression of *SAD* and *CAD* genes in xylem tissue, consistent with the cell-specific locations of SAD and CAD proteins and with the proposed pathways;

(5) We have developed a novel small interfering RNA (siRNA)-mediated stable gene-silencing system in transgenic plants;

(6) Using the siRNA and *P. trichocarpa* transformation/regeneration systems we are currently producing transgenic *P. trichocarpa* to investigate the interactive functions of *CAD* and *SAD* in regulating guaiacyl and syringyl lignin biosynthesis;

(7) We have cloned for the first time from a tree species, *P. trichocarpa*, small regulatory RNAs termed microRNAs (miRNAs) with likely effector roles in regulating the expression of genes involved in lignin biosynthesis and wood formation networks.

The pathway of syringyl lignin biosynthesis in angiosperm trees

We proposed a new monolignol biosynthesis model, in which the biosynthesis of syringyl monolignol is derived from guaiacyl pathway with the branch point at coniferaldehyde. This pathway is mediated by coniferaldehyde 5-hydroxylase (CAld5H), 5-hydroxyconiferaldehyde *O*-methyltransferase (AldOMT), and syringyl-specific alcohol dehydrogenase (SAD).

This hypothesis was supported by the fact that SAD in aspen had a catalytic specificity towards sinapaldehyde, and CAD was shown to be guaiacyl- or coniferaldehyde-specific. In the course of CAD cloning from xylem tissue, we identified two abundant CAD-like cDNAs that had similar but distinct sequences. Enzyme kinetics of their recombinant proteins revealed that one was specific to coniferaldehyde and the other to sinapaldehyde, supporting distinct roles for CAD and SAD in guaiacyl and syringyl monolignol biosynthesis, respectively. Furthermore, the protein immunolocalization revealed that *SAD* was preferentially expressed in fiber and ray parenchyma cells, which are enriched in syringyl lignin, while CAD protein was dominant in cells with guaiacyl lignin biosynthesis. Based on these and other lines of evidence, it was suggested that the distinct CAD and SAD functions are linked spatiotemporally to the differential biosynthesis of guaiacyl and syringyl lignins in different cell types (Li, et al. 2001). This discovery updated our knowledge of monolignol biosynthesis, and thus, opens a new avenue to investigating the regulatory network of lignin biosynthesis in association with xylem differentiation. Furthermore, the discovery of the linkage between xylem fiber cell differentiation and its associated predominant biosynthesis of syringyl monolignol also provide a basis for further investigation of the relationship between syringyl monolignol biosynthesis and xylem fiber cell function evolution in higher plants.

Genetic confirmation of the proposed syringyl lignin biosynthesis pathway in aspen

As suggested by the enzyme turnover rate, the metabolic flux to syringyl lignin biosynthesis pathway might be limited by coniferaldehyde 5-hydroxylase (CAld5H), the branch point enzyme leading to syringyl monolignol biosynthesis. CAld5H has a very low enzyme turnover rate as compared with the other enzymes in the syringyl pathway. To verify these *in vitro* results, we produced transgenic aspen to demonstrate *in vivo* that CAld5H may serve as a control point for regulating syringyl lignin biosynthesis.

We transferred a sense heterologous sweetgum *CAld5H* gene into aspen under the control of a xylem-specific promoter. At the same time, we also transferred a homologous antisense *4CL* in combination with the sweetgum *CAld5H* using our developed multiple-gene transformation system. The results were published on *PNAS* (Li et al. 2003). Based on these *in vivo* results, we concluded that coniferaldehyde, the substrate of CAld5H, is the major branch point in monolignol biosynthetic pathway in angiosperm trees and that monolignol metabolic flexibility is operative downstream of this major branch point. Thus, our proposed principal monolignol biosynthetic pathway with CAld5H serving as the control point to syringyl monolignol formation was further verified by *in vivo* evidence.

Clarification of CCR catalysis, a further support of the proposed major monolignol biosynthesis fluxes

Cinnamoyl-coenzyme A reductase (CCR) has long been thought to catalyze the conversion of various cinnamoyl CoA esters to their corresponding cinnamaldehydes. This had led to the suggestion of a metabolic grid for monolignol biosynthesis. To further understand monolignol

biosynthesis in xylem development, we cloned xylem-specific and lignin-related *CCR* and *CCoAOMT* cDNAs from aspen developing xylem. The recombinant CCR protein was produced through an *E. coli* expression system and purified to electrophoretic homogeneity. CCR biochemical properties were characterized through direct structural corroboration and quantitation of the reaction products using a liquid chromatography-mass spectrometry system. The enzyme kinetic profiles demonstrated that CCR is highly specific for catalyzing the reduction of feruoyl-CoA as measured by Km and Vmax/Km values, and that when various cinnamoyl CoA esters were used as a pool of mixed substrates, feruoyl-CoA became a strong competitive inhibitor for CCR reactions with other CoA esters. Moreover, when caffeoyl-CoA *O*-methyltransferase (CCoAOMT) was coupled with CCR in a reaction with caffeoyl-CoA ester as the substrate, caffeoyl-CoA was converted into coniferaldehyde but not caffealdehyde. These results suggested that the CCoAOMT/CCR-mediated conversion of caffeoyl-CoA ester to coniferaldehyde via feruoyl-CoA constituted a main metabolic flux toward monolignol biosynthesis in lignifying xylem, providing further evidence for the proposed major fluxes for monolignol biosynthesis.

The promoter activities of CAD and SAD in xylem

Differential production of CAD and SAD proteins were found in different types of xylem cells (Li, et al. 2001). To elucidate the mechanisms underlying this differential expression, we cloned aspen CAD and SAD promoter sequences. Transcription start site (TSS) analysis identified a 2.0 kb *PtCAD* and a 2.1 kb *PtSAD* promoter fragments, respectively. Sequence analysis showed that these two promoters exhibited little homology, although the expression of their cDNAs is predominant in xylem tissue as previously demonstrated. To assess the promoter function, *PtCAD* and *PtSAD* gene promoter fragments were serially deleted and each deletion was constructed into a binary *GUS* report-gene vector for transforming tobacco. Transgenic tobacco plants were analyzed in several repeats for each promoter-GUS deletion construct. The analysis of these transgenic tobacco plants for promoter activities based on GUS staining provided the following results. (1) Both *PtCAD* and *PtSAD* promoters were specifically active in tobacco xylem tissue, consistent with PtCAD and PtSAD protein expression in aspen; (2) The GUS activity in *PtCAD* promoter plants was particularly intensified in cells adjacent to vessel cells, while the GUS activity appeared mainly in xylem ray cells in *PtSAD* promoter plants. The preferential distribution of the GUS activity in various cell types of the xylem tissue suggests the presence of distinct tissue-specific regulatory elements in these two promoters. Although it is difficult to precisely identify the GUS activity to a particular cell type due to the diffusing nature of the GUS staining, the observed differential distribution of GUS signal between these two promoters suggested that the expression of PtCAD and PtSAD is associated with cell type differentiation in xylem tissue. (3) The shortest deletion of *PtCAD* promoter, which contained only 0.4 kb sequence upstream of TSS, was still able to confer the promoter activity in the transgenic tobacco as did the 2.0 kb promoter. This suggests that most of the effective tissue specific *cis*-elements in *PtCAD* promoter may be located in a short sequence stretch close to the TSS. In contrast, deletion of *PtSAD* promoter sequence from 2.1 to 0.9 kb completely abolished the promoter activity, suggesting that the *cis*-elements of *PtSAD* gene promoter may be located 1.0 kb upstream of TSS. These promoter activity results further imply that specific elements in these two promoters may be involved in regulatory networks responsible for tissue specificity and thus the developmental control of guaiacyl and syringyl lignin biosynthesis in trees. It would be interesting to identify the *cis*-elements in these promoter sequences and their binding factor genes to understand the effectors for regulating the expression of these guaiacyl- and syringyl-specific monolignol genes involved in lignin biosynthetic networks.

Stable siRNA gene silencing system in plants

One common mechanism of gene expression silencing in plants and animals involves the generation of siRNAs of ~21-23 nucleotides (nt) that can recognize specific target mRNA sequences and then lead to their degradation. This mechanism has applied to the transient knockout of gene expression in mammalian cells using synthetic siRNA molecules that can be injected into cells. In addition, the *in vivo* expression of siRNA in cell through transformation has been investigated in animal systems, however, this technology has not been reported in plants. To understand functions of specific genes in plants, especially those involved in longer-term developmental processes, a heritable and specific siRNA silencing technology is needed. We designed two siRNA expression vectors under the control of a human and a plant Pol III RNA gene promoter, respectively. The vectors are amenable to *Agrobacterium*-mediated plant transformation for a selective, stable gene silencing. Using tobacco system, we tested the effectiveness and heritability of these systems in silencing a target gene encoding β–glucuronidase (GUS) (Lu, *et al*. 2004).

Interactive roles of CAD and SAD in regulating guaiacyl and syringyl lignin biosynthesis

We used reverse genetic approaches to further understand the interactive roles of CAD and SAD *in vivo*. We would like to silence the expression of CAD and SAD genes individually and in combination. However, due to the high sequence similarity between *CAD* and *SAD* genes, the conventional gene silencing technologies, such as the antisense cDNA approach, may be difficult to distinctly knockout one without affecting the other. This was actually one of the main reasons that prompted us to seek other alternatives to this non-specificity issues associated with most of the gene knockout technologies, and developed the siRNA technique described above. Because genome sequence of *Populus trichocarpa* is now available, we decided to use the identical *P. trichocarpa* clone (Nisqually-1) used for genome sequencing as the model to investigate the *in vivo* interactive roles of CAD and SAD. For this purpose, we also have developed *Agrobacterium*-mediated transformation/regeneration systems for this particular *P. trichocarpa* clone.

After poplar genome sequence data from JGI and NCBI were partially assembled into contig sequences, we analyzed all sequences similar to those of *CAD* and *SAD* genes in the genome and identified *PtCAD* and *PtSAD* homologs in *P. trichocarpa*. According to the sequence information, we selected two 19-nt fragments having perfect sequence complementarity with two distinct sites in the *CAD* and the *SAD* mRNAs, designated CAD1, CAD2, SAD1 and SAD2, for preparing the siRNA templates. These 19-nt sequences do not share any significant sequence homology with any *P. trichocarpa* genes other than their targeted fragments. The siRNA templates were individually cloned into the expression module pGPSL-HPT to give the corresponding binary vectors, pBICAD1, pBICAD2, pBISAD1, and pBISAD2. A control binary vector pBCK was constructed from pBI121 by deleting its *GUS* expression cassette. These binary vectors were individually mobilized into *A. tumefaciens* C58 for transforming *P. trichocarpa*. Three transformation events were conducted. Event A, pBICAD1-harboring *Agrobacterium* strain was mixed with an equal volume of the same density of the pBISAD1-containing *Agrobacterium* and used to co-cultivate with *P. trichocarpa*. In this way, in one transformation event, transgenic trees with downregulated (1) *CAD*, (2) *SAD* and (3) *CAD* and *SAD*, respectively, can be generated simultaneously for characterizing the interactive roles of CAD and SAD *in vivo*. To ensure the success of specific gene silencing, another set of siRNA sequences were used. Thus, in Event B, pBICAD2 and

pBISAD2 *Agrobacterium* strains were mixed for co-cultivation. Event C, the control pBC *Agrobacterium* was used. The production of these transgenics is in progress.

MiRNA functions in wood formation

Lignin biosynthesis and wood formation are developmental events associated with cell differentiation in plants. Although the knowledge of how monolignols are synthesized in wood has been greatly advanced in recent years, little is known about the regulatory effectors/factors that trigger monolignol biosynthesis for lignin deposition. We also do not know the regulatory mechanisms of how newly divided cambial elements differentiating into specialized cells recruit a concomitant biosynthesis of preferred lignin types to form the xylem or wood. While our research on biochemical aspects of monolignol biosynthesis continues, we began to engage in research at the gene regulation levels, using *P. trichocarpa* as a model system, trying to understand the regulatory networks controlling the developmentally coordinated lignin biosynthesis and cell function specialization during wood formation. This research had strong knowledge and technology supports as a result of recent advances in related research: (1) *P. trichocarpa* genome sequence, the only one for a tree species, is now available; (2) efficient genetic transformation and regeneration systems are established for *P. trichocarpa*; (3) an siRNA-mediated gene-specific silencing technique has been developed for plants; (4) tremendous progress has been made in understanding the roles of small, noncoding RNAs, or miRNAs, in regulating gene expression in eukaryotes; and (5) a variety of miRNAs have been isolated from lignin/wood forming tissues of *P. trichocarpa* (Lu et al. 2005). MiRNAs have been demonstrated to play critical roles in regulatory networks of a variety of processes associated with development, such as developmental transitions and patterning, cell differentiation, and maintaining fitness to various environmental conditions. We believe that miRNAs might also coordinate lignin biosynthesis regulatory network in wood development, as supported by our preliminary results that miRNAs are associated with lignin biosynthesis. Thus, we are in a unique position to extend our current biochemical investigations on monolignol biosynthesis to cover research on regulatory mechanisms that underlie lignin biosynthesis as the first step towards achieving our long-term research goal of understanding wood formation.

Major publications supported by this grant:

Lu, S., Sun, Y., Clark, C., Li, L., and Chiang, VL. 2005, Novel and mechanical stress responsive microRNAs in *Populus trichocarpa* that are absent from Arabidopsis. The Plant Cell, 17(8): 2186-203.

Li, L., Cheng, X., Lu, S., Umezawa, T., and Chiang, V. L. (2005), Clarification of Cinnamoyl Co-enzyme A Reductase Catalysis in Monolignol biosynthesis. Plant & Cell Physiology, 46(7): 1073-82.

Lu, SF., Shi, R., Tsao, CC., Yi, X., Li, L., and Chiang, VL., (2004). Small interfering RNA (siRNA)-mediated heritable gene silencing in plants. Nucleic Acid Research, 32: e171.

Nakatsubo, T., Li, L., Chiang, V.L., Hattori, T., Shimada, M., Umezawa, T. (2003). Basic studies towards elucidation of heartwood formation mechanisms. Wood Research. *90*, 5-6.

Li, L., Zhou, Y., Cheng, X. F., Sun, J., Marita, J., Ralph, J., and Chiang, V.L. (2003). Combinatorial genetic reduction of lignin quantity and augmentation of lignin syringyl constituents in trees through multigene co-transformation. PNAS, 100: 4939-4944.

Li, L.,Cheng, X.F., Leshkevich, J., Umezawa, T., Harding, S.A., and Chiang, V.L. (2001). The last step of syringyl monolignol biosynthesis in angiosperm dicots is regulated by a novel gene encoding sinapyl alcohol dehydrogenase. The Plant Cell, 13(7)1567-1585.