Assembly of a lignin modification toolbox

Investigators
Clint Chapple, Distinguished Professor and Head, Department of Biochemistry;
Alan Friedman, Associate Professor, Department of Biological Sciences; Larisa
Avramova, Research Scientist, Bindley Bioscience Center; Shiva Hemmati and
Shinyoung Lee, Post-doctoral Fellows, Department of Biochemistry; Patrick O'Neil,
Post-doctoral Fellow, Department of Biological Sciences, Yi Li, Graduate Researcher,
Department of Biochemistry; Nick Fico, Graduate Researcher, Department of Biological
Sciences, Purdue University.

Abstract
    This project focuses on establishing an enzymatic toolbox for the production of lignin
modification molecules (LMMs). We are taking multiple approaches to the identification
and characterization of LMM-synthesizing enzymes and to date have identified three
classes of catalysts that are the focus of our work.
    In the past year we have identified and characterized a cytochrome P450-dependent
monooxygenase (P450) that is required for the synthesis of a group of novel alpha-pyrone
secondary metabolites accumulated in Arabidopsis stems. This new enzyme has recently
evolved from an enzyme in lignin biosynthesis and may provide us with another tool with
which to modify lignin biosynthesis directly and will give us greater insights into the
evolution of substrate specificity in P450s. In a parallel approach, we have generated a
large family of chimeric P450 proteins by recombining cinnamate 4-hydroxylase (C4H)
genes from two distantly related plant species. Many of these proteins retain their
enzymatic activities strongly supporting the idea that chimeric P450s can be generated
using a library of less-closely-related genes to generate novel catalysts. Following up on
the success of this approach, we are completing additional chimeric libraries between
P450s which catalyze different steps in lignin synthesis. The first of these additional
libraries, between the C4H and ferulate 5-hydroxylase (F5H) of Arabidopsis, has been
constructed and is being prepared for analysis. Other lignin-associated P450s and the
previously mentioned alpha-pyrone biosynthetic P450 will soon be employed in
comparable chimerigenesis experiments.
    Another group of proteins known as BAHD acyltransferases use Coenzyme A (CoA)
thioesters as activated donors in acyltransferase reactions that generate dimeric
phenylpropanoid conjugates. These products can be regarded as potential LMMs because
their phenolic moieties are amenable to lignin incorporation, while their internal ester or
amide bonds would introduce readily cleavable linkages into the lignin polymer. During
the past year we have successfully isolated the full length sequence of a BAHD
acyltransferase from the Asian fiber crop kenaf (*Hibiscus cannabinus*), which is known
to accumulate the candidate LMM coniferyl ferulate. Characterization of this enzyme is
currently underway.
    Finally, we have initiated a new project to identify potential LMMs through the
analysis of monolignol conjugates in natural accessions of Arabidopsis. The new
metabolites, first characterized as part of our Department of Energy-sponsored research,
are themselves potential LMMs. Using genome-wide association mapping, we expect to

be able to make rapid progress toward the identification of the enzymes and genes involved in their synthesis.

Introduction

Although the phenylpropanoid pathway produces many compounds of value to humans, a major goal of research on the pathway has been to improve our understanding of lignin biosynthesis.[5] Biofuel production has provided a new motivation for this interest because the quantity and quality of lignin in biomass crops interferes with the access of hydrolytic enzymes to the polysaccharide components of the plant cell wall, thereby inhibiting their conversion to fermentable monosaccharides. For this reason, the ability to manipulate lignification in biofuels crops would lead to substantial economic and environmental gains. The objective of this project is to learn how to fundamentally alter the chemistry of the lignin polymer in such a way that it radically alters the ease with which it can be removed from biofuel crops, thus substantially increasing yields of cell wall-derived sugars and the efficiency of their subsequent fermentation without compromising the viability of the plants themselves.

Our approach to the assembly of our LMM enzymatic toolkit is based upon several attributes that these enzymes have in common. First, as a result of their involvement in many secondary metabolic pathways, P450s as well BAHD and SCPL acyltransferases can be found in a wide range of plants, and exhibit broad catalytic diversity. Second, we have developed methods for the heterologous expression of these unknown proteins and robust assays with which they can be challenged with a spectrum of metabolites in a high throughput manner such that their substrate specificity, and suitability for LMM synthesis, can be determined. Finally, many proteins within each class of catalyst share enough sequence homology that we can apply our newly-developed strategies for the generation of chimeric proteins so that the proteins we identify can be further refined for the production of optimal LMMs.

When these native or engineered genes are over-expressed in biofuel crops, we expect that the phenolic nature of the LMMs will lead to their incorporation into the growing lignin polymer but will not compromise the function of this important cell wall component. Either through altered cross-linking or through the incorporation of ester or amide linkages, the LMMs will generate "weak links" within the lignin polymer, greatly enhancing the ease with which it can be deconstructed post-harvest.

Background

The lignin heteropolymer is produced via the oxidative coupling of p-coumaryl alcohol, coniferyl alcohol, and sinapyl alcohol subunits (collectively termed monolignols). The polymerization of these subunits leads to the formation of p-hydroxyphenyl (H), guaiacyl (G), and syringyl (S) lignin, respectively. H subunits are usually minor components, and the degree to which G and S units are incorporated into the polymer (commonly denoted as the S:G ratio) varies widely among species, tissue types, and even within an individual cell wall. Although H, G, and S units are widely regarded as the only monomers found in lignin, more sophisticated methods of lignin analysis applied to a broader range of plant species, mutants, and transgenic lines have revealed that other subunits, including aldehydes, side-chain-reduced monolignols, and phenylpropanoid esters and amides appear to be *bona fide* lignin components.[6-8]

This plasticity of lignin monomer composition is further demonstrated by the *fah1* mutant of Arabidopsis, the lignin of which lacks syringyl monomers [1], maize and pine mutants that deposit lignins containing 5-hydroxyguaiacyl and dihydroconiferyl alcohol monomers, respectively [9-10], engineered Arabidopsis, tobacco, and poplar lines that produce lignins derived almost solely from syringyl units [2-3], and an Arabidopsis mutant that deposits lignin containing only *p*-hydroxyphenyl units [3,11], among others. It is noteworthy that in some cases, these modified lignins are much more readily extracted than the lignins normally made by plants [12], a result that bodes well for the utility of lignin modification strategies in the improvement of biofuel crops.

Based upon our emerging understanding of the plasticity of lignin monomer composition, it seems clear that it will be possible to engineer plants to generate novel lignins that still support normal plant growth, development, and levels of biomass productivity. This might be accomplished by engineering plants to synthesize monolignols with novel substitution patterns that would lead to lignins with altered chemistries and enhanced degradability. Alternatively, there is ample precedent in the literature to suggest that plants could be engineered to incorporate dimeric phenylpropanoid ester- or amide-linked conjugates into their lignin, making it easier to cleave either chemically or enzymatically. Both of these strategies would result in a polymer that can be more effectively and efficiently removed from biomass feedstocks.

Results

This project focuses upon the assembly and use of a toolbox for lignin modification. Our highest priority objectives have been to clone and express target genes and characterize their encoded proteins with regard to substrate specificity to evaluate their potential as LMM-synthesizing enzymes. In parallel, we are using novel gene chimera construction software and a robotic gene re-assortment and reassembly method to generate synthetic P450 proteins that produce novel monolignols not found in Nature. This approach will also generate fundamental new knowledge with regard to structure-function relationships within this class of important catalysts.

BAHD acyltransferase proteins

BAHD family acyltransferases can be phylogenetically grouped into several distinct clades, one of which contains a number of enzymes which utilize hydroxycinnamic and benzoic acid derivatives as substrates. Members of this clade are particularly promising as potential LMM-synthesizing enzymes, since hydroxycinnamic and benzoic acid derivatives are frequently incorporated into the lignin polymer. We have successfully employed a degenerate PCR-based approach using conserved amino acid sequences from BAHD family members to obtain the full length clone for a novel acyltransferase-encoding gene from the Asian fiber crop kenaf (*Hibiscus cannabinus*), which is known to accumulate the candidate LMM coniferyl ferulate. Phylogenetic analysis strongly suggests that the closest relatives of the acyltransferase is involved in the synthesis of benzyl benzoate or phenethyl benzoate, both of which are strikingly similar in structure to conferyl ferulate. Further characterization of the enzyme is currently in progress.

Cytochrome P450-dependent monooxygenases

We have developed high-throughput *in vivo* assays for the P450 enzymes that are involved in phenylpropanoid metabolism derived from *Arabidopsis thaliana* including C4H, F5H, and *p*-coumaroyl shikimate 3' hydroxylase (C3'H) and C4H and F5H from *Selaginella moellendorffii*. Each of these assays are conducted in heterologous WAT11 yeast strains that also express the corresponding P450 reductase from Arabidopsis.

We are using a high-throughput version of a previously-established HPLC-based assay that has been used extensively in our lab. A seven minute LC-MS separation has been developed, suitable for the identification of reaction products from chimeras with known or novel activities. In the past year we have validated this method for C4H activity to ensure its reproducibility, robustness, and sensitivity in the large-scale assay of activity for more than a thousand chimeric protein variants that we expect to generate. Its reproducibility and robustness has been confirmed, yielding similar data over many months of assays. After combining a number of small modifications, the sensitivity of this assay has been increased more than 10-fold, yielding an assay that has a dynamic range of more than 100.

All 64 hybrids from the first set of site-directed chimeras between the C4H enzymes from *Arabidopsis thaliana* and *Selaginella moellendorffii* have been characterized for activity in vivo. Many of them retain significant activity, and these are not only hybrids that are nearly exclusively derived from one parent. As noted above, these results strongly support the idea that chimeric P450s can be generated from diverse genes to generate novel catalysts. We have also taken the in vivo activity data and used it to generate a model of the origins of C4H activity. We find that, using only five parameters, the species identity of four of the fragments and a score derived from our graphical models of amino acid interactions, we can develop a linear model that explains a highly significant fraction of the total variation in activity among the chimeras. The ability to conduct such analyses should aid future engineering of novel P450's.

Our initial experiments with the C4H chimeras revealed that the expression levels in our initially-proposed yeast system (which employed a Gateway-modified version of pYeDP60) yielded levels of in vivo activity which were less than desirable (significantly less than in the original pYeDP60 without Gateway modification) and amounts of protein that made analysis of the quantity and stability of the P450 enzymes by CO-difference spectroscopy problematic. Switching to a previously-described variant of pYeDP60 with the USER cloning system [13], has reversed the reduction of activity seen in the Gateway modifications, obtaining several-fold higher expression than the original pYeDP60, while retaining some of the ease-of-cloning benefits. In some cases, we have further enhanced the expression by the use of synthetic genes that contained optimized codons for yeast expression.

The chimeric C4H enzymes described above are currently being recloned into the USER system to enable a confirmatory and more detailed analysis of activity in vivo and a determination of expression and stability by CO-difference spectroscopy of microsomal extracts. Microsomal extracts from the most interesting enzymes will also be analyzed for their kinetic parameters in vitro. We anticipate being able to model the stability and kinetic parameters of these enzymes, much as we've done for in vivo activity.

In further work, and as a control and reference for the proposed site-directed recombination methods, we have conducted initial trials implementing a random chimera generating approach (gene shuffling) for recombining the same two divergent

C4H genes from *Arabidopsis thaliana* and *Selaginella moellendorffii*. Our initial experiments have produced approximately 30 confirmed shuffled genes. We are currently recloning these genes into the pYeDP60-User expression system to evaluate their stability and catalytic capacity. We are also currently developing assays of P450 activity employing alternative, cinnamic acid-related substrates. All the C4H chimeras from both our site-directed and random libaries will be analyzed for activities against these substrates. We have also completed the first library between two different P450 enzymes, the C4H and F5H of *Arabidopsis* and are preparing to transfer these chimeras into yeast and assay their activity as both C4H and F5H enzymes and their ability to hydroxylate novel substrates.

Identification of a novel phenylpropanoid P450 in Arabidopsis

F5H (CYP84A1) catalyzes the 5-hydroxylation of coniferaldehyde and coniferyl alcohol to yield 5-hydroxy-coniferaldehyde and 5-hydroxy-conferyl alcohol, respectively. [2,5] In Arabidopsis, there is a paralog of CYP84A1 that has been designated CYP84A4. When we heterologously expressed CYP84A4 we found that it did not show detectable activity towards either coniferyl alcohol or coniferyl aldehyde, both of which are substrates of CYP84A1. Based upon these data, we concluded that the catalytic activity of CYP84A4 is distinct from that of CYP84A1 and that CYP84A4 has lost its ancestral activity as an F5H. CYP84A4 appears to be recently evolved in Arabidopsis because few angiosperms have CYP84 family members other than those that encode F5H, and specifically, a CYP84A1 paralog like CYP84A4 is not present in members of the closely
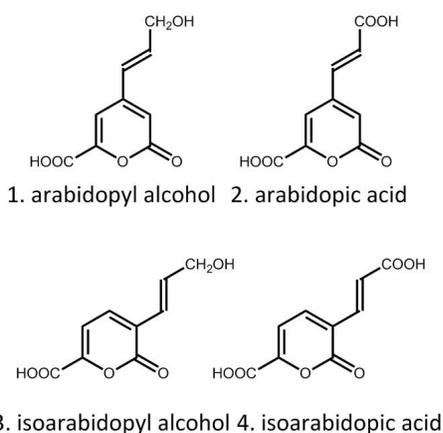

1. arabidopyl alcohol 2. arabidopic acid

3. isoarabidopyl alcohol 4. isoarabidopic acid

Figure 1. APs identified in Arabidopsis.

related Brassica genus. To elucidate the function of CYP84A4, we first identified a homozygous CYP84A4 T-DNA insertional mutant (Salk_064406), which we later named *apd1-1* for in *arabidopyrone deficient1*. Plants homozygous for *apd1* accumulate normal levels of sinapate esters which, together with the strong phenotype of the CYP84A1-deficient *fah1* mutant, indicate that CYP84A1 and CYP84A4 are not redundant genes. Reverse genetics and metabolic profiling have revealed that CYP84A4 is required for the synthesis of four previously unknown metabolites that we have named arabidopyrones (APs). To determine the identity of the unknown compounds we purified them by HPLC, analyzed them by LC-MS, and elucidated their structures by NMR. To our surprise, the metabolites had not been previously described in Arabidopsis or any other plants to date.

The four metabolites are typified by a 6-carboxy-2-pyrone core moiety, decorated with either a 3-hydroxyprop-1-en-1-yl or a 2-carboxyvinyl side chain at the 3 or 4 position of the pyrone ring, reminiscent of the side chain of the general phenylpropanoid alcohols or acids (Fig. 4). We have named these newly discovered pyrone compounds arabidopyl alcohol, iso-arabidopyl alcohol, arabidopic acid and iso-arabidopic acid. The structure of arabidopyl alcohol was confirmed by synthesis.

When we assayed heterologously-expressed CYP84A4 against a range of possible phenylpropanoid pathway intermediates, only *p*-coumaraldehyde was found to be a substrate for CYP84A4, yielding caffealdehyde as a product. In view of these results it is interesting to note that CYP84A1, which is presumably the ancestral protein from which CYP84A4 is derived, exhibits slight activity as a *p*-coumaraldehyde and *p*-coumaryl alcohol 3-hydroxylase with Km values several hundred fold higher than those for its optimal substrates, coniferaldehyde and coniferyl alcohol. This novel catalyst provides an important additional input for our chimeragenesis work, and may shed light on how substrate specificities can evolve among the P450 superfamily.

Gaining access to novel phenylpropanoid metabolism in Arabidopsis

In the context of our Department of Energy-funded research, we have developed methods for the LC-MS analysis of a group of novel phenylpropanoids in Arabidopsis which hyper-accumulate in a mutant we had previously characterized in our laboratory. A striking commonality observed in these structures is that they include a hydroxycinnamyl alcohol moiety that appears to have been involved in a radical-coupling reaction reminiscent of those involved in lignin polymerization. We are now using our LC-MS
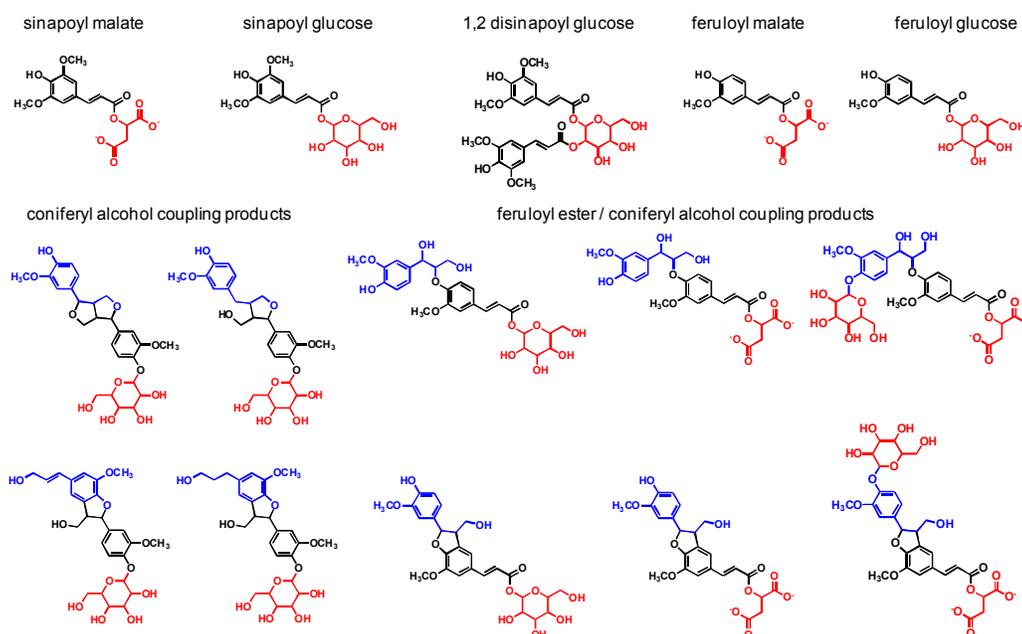

Figure 2. Novel monolignol coupling products identified in Arabidopsis. Known metabolites are shown in the top row, with hydroxycinnamic acids shown in black and alcohol acceptors (malate or glucose) shown in red. In the bottom two rows, novel cross-linking products are illustrated with their monolignols moieties shown in blue.

based profiling strategy to identify Arabidopsis natural accessions that are altered in their accumulation of lignans and monolignol conjugates. Using genome-wide association mapping, we will be able to identify genes required for their synthesis in only 1-2 months after the profiling is complete. The enzymes encoded by these genes will then be characterized in detail, and hold distinct promise as catalysts to generate useful LMMs.

Progress

In this progress report, we have outlined our initial efforts towards isolating genes of potential value in the production of LMMs, lignin building blocks which we anticipate will improve the degradability of lignin while leaving intact its critical function in the cell wall. Specifically, we have taken major steps towards the development of an *in vitro* mutagenesis approach for the generation of novel P450s to modify hydroxylation patterns in lignin, and have made significant progress in elucidating the function of novel P450s in Arabidopsis metabolism.

These preliminary results represent essential first steps toward the identification of candidate genes that produce LMMs. It is our hope that when these native or engineered genes are over-expressed in biofuel crops, the phenolic nature of the LMMs produced will lead to their incorporation into the growing lignin polymer. If such experiments come to fruition, we envision a future in which LMM-containing crops could vastly increase the efficiency of cellulosic biofuel production and dramatically curb our excessive emission of greenhouse gasses.

Future Plans

In the remaining term of the project we will continue to generate and characterize P450 chimeras. As described in our last progress report, using other funding we have recently identified P450s from the lycophyte *Selaginella moellendorrffii* and have prioritized them for chimerigenesis along with the novel Arabidopsis P450 CYP84A4 described in this report. We will use our LC-MS based profiling strategy to identify Arabidopsis natural accessions that are altered in their accumulation of lignans and monolignol conjugates. Finally, we have determined that another potential LMM of great interest, p-coumarylferulate, is accumulated in celery. Rather than employing our previous degenerate PCR strategy to isolate the BAHD acyltransferase involved, we will deep sequence the celery transcriptome to identify candidate genes, express them using the high throughput assay described in our last progress report, and characterize the enzyme(s) of interest in detail.

Publications
1. Bioenergy Systems Research Institute, University of Georgia, Athens GA. April 2011. Using Arabidopsis to learn about lignin.
2. Department of Biology, Virginia Polytechnic Institute and State University, Blacksburg VA, October 2010. A tale of three P450s.
3. ASPB 2010, Montreal PQ, July 2010. Lignin: an extracellular polymer with unique biochemical plasticity.

References
1. Chapple, C.C., Vogt, T., Ellis, B.E. and Somerville, C.R. An Arabidopsis mutant defective in the general phenylpropanoid pathway. Plant Cell, 4, 1413-1424, 1992.

2. Meyer, K., Shirley, A.M., Cusumano, J.C., Bell-Lelong, D.A. and Chapple, C. Lignin monomer composition is determined by the expression of a cytochrome P450-dependent monooxygenase in Arabidopsis. Proc. Natl. Acad Sci. USA, 95, 6619-6623, 1998.
3. Franke, R., McMichael, C.M., Meyer, K., Shirley, A.M., Cusumano, J.C. and Chapple, C. Modified lignin in tobacco and poplar plants over-expressing the Arabidopsis gene encoding ferulate 5-hydroxylase. Plant J., 22, 223-234, 2000.
4. Franke, R., Hemm, M.R., Denault, J.W., Ruegger, M.O., Humphreys, J.M. and Chapple, C. Changes in secondary metabolism and deposition of an unusual lignin in the *ref8* mutant of Arabidopsis. Plant J., 30, 47-59, 2002.
5. Humphreys, J.M. and Chapple, C. Rewriting the lignin roadmap. Curr. Opin. Plant Biol., 5, 224-229, 2002.
6. Lu, F.C., and Ralph, J. Detection and determination of *p*-coumaroylated units in lignins. J Agr. Food Chem. 47, 1988-1992, 1999.
7. Lu, F.C., and Ralph, J. Preliminary evidence for sinapyl acetate as a lignin monomer in kenaf. Chem. Commun. 1, 90-91, 2002.
8. Boerjan, W., Ralph, J. and Baucher, M. Lignin Biosynthesis. Annu. Rev. Plant Physiol. Plant Mol. Biol., 54, 519-546, 2003.
9. Vignols, F., Rigau, J., Torres, M.A., Capellades, M. and Puigdomenech, P. The *brown midrib3* (*bm3*) mutation in maize occurs in the gene encoding caffeic acid O-methyltransferase. Plant Cell, 7, 407-416, 1995.
10. Ralph, J., MacKay, J.J., Hatfield, R.D., O'Malley, D.M., Whetten, R.W., and Sederoff, R.R. Abnormal lignin in a loblolly pine mutant. Science, 277, 235-239, 1997.
11. Franke, R., Humphreys, J.M., Hemm, M.R., Denault, J.W., Ruegger, M.O., Cusumano, J.C., and Chapple, C. The Arabidopsis *REF8* gene encodes the 3-hydroxylase of phenylpropanoid metabolism. Plant J., 30, 33-45, 2002.
12. Huntley, S.K., Ellis, D., Gilbert, M., Chapple, C. and Mansfield, S.D. Significant increases in pulping efficiency in C4H-F5H-transformed poplars: Improved chemical savings and reduced environmental toxins. J. Agric. Food Chem., 51, 6178-6183, 2003.
13. Hamann T, Møller BL. Improved cloning and expression of cytochrome P450s and cytochrome P450 reductase in yeast. Protein Expr. Purif. 2007 56:121-127.

Contacts

Clint Chapple: chapple@purdue.edu
Alan Friedman: afried@exchange.purdue.edu