

## ABSTRACT

Author: Jaini, Rohit. PhD

Institution: Purdue University

Degree Received: December 2017

Title: A Computational and Experimental Investigation of Lignin Metabolism in *Arabidopsis thaliana*.

Major Professor: John A. Morgan

Predominantly localized in plant secondary cell walls, lignin is a highly cross-linked, aromatic polymer that imparts structural support to plant vasculature, and renders biomass recalcitrant to pretreatment techniques impeding the economical production of biofuels. Lignin is synthesized via the phenylpropanoid pathway where the primary precursor phenylalanine (Phe) undergoes a series of functional modifications catalyzed by 11 enzyme families to produce *p*-coumaryl, coniferyl, and sinapyl alcohol, which undergo random polymerization into lignin. Several metabolic engineering efforts have aimed to alter lignin content and composition in order to make biofuel feedstock more amenable to pretreatment techniques. Despite significant advances, several questions pertaining to carbon flux distribution and control in the phenylpropanoid network remain unanswered. Furthermore, complexity of the metabolic pathway and a lack of sensitive analytical tools add to the challenges of mechanistically understanding lignin synthesis.

In this work, I describe improvements in analytical techniques used to characterize phenylpropanoid metabolism that have been applied to obtain a comprehensive quantitative mass balance of the phenylpropanoid pathway. Finally, machine learning and artificial intelligence were utilized to make predictions about optimal lignin amount and composition for improving saccharification. In summary, the overarching goal of this thesis was to further the understanding

of lignin metabolism in the model system, *Arabidopsis thaliana*, employing a combination of experimental and computational strategies.

First, we developed comprehensive and sensitive analytical methods based on liquid chromatography coupled with tandem mass spectrometry (LC-MS/MS) to quantify intermediates of the phenylpropanoid pathway. Compared to existing targeted profiling techniques, the methods were capable of quantifying a wider range of phenylpropanoid intermediates, at lower concentrations, with minimal sample preparation. The technique was used to calculate flux maps for wild type and mutant *Arabidopsis* stems that were supplied  $^{13}\text{C}_6$ -Phe exogenously. Flux maps computed in this work; (i) suggest the presence of a hitherto uncharacterized alternative route to caffeic acid and lignin synthesis, (ii) shed light on flux splits at key branch points of the network, and (iii) indicate presence of inactive pools for a number of metabolites.

Finally, we present a machine learning based model that captures the non-linear relationship between lignin content and composition, and saccharification efficiency. A support vector machine (SVM) based regression technique was developed to predict saccharification efficiency and biomass yields as a function of lignin content, and composition of monomers that make up lignin, namely *p*-coumaryl (H), coniferyl (G), and sinapyl (S) sub-units. The model was trained on data obtained from the literature and validated on *Arabidopsis* mutants that were excluded from the training data. Functional forms obtained from SVM regression were further optimized using genetic algorithms (GA) to maximize total sugar yields. Our efforts resulted in two optimal solutions with lower lignin content and interestingly varying H:G:S composition that were conducive to high saccharide extractability.