Modeling of Cellulose, Hemicellulose and Lignin-Carbohydrate Complex Formation and Regulation to Understand Plant Cell Wall Structure

Vincent Chiang[1]* (vchiang@ncsu.edu), Ron Sederoff,[1] David Muddiman,[1] Cranos Williams,[1] Fikret Isik,[1] Joel Ducoste,[1] Chris Smith,[1] John Ralph,[2] and Sunkyu Park[1]

[1]North Carolina State University, Raleigh; and [2]University of Wisconsin, Madison

Project Goals: The goal of this proposal is to use a systems approach to increase the fundamental understanding of the plant secondary cell wall. We will use multiple transgenic perturbations in *Populus trichocarpa* and measure effects on plants using advanced quantitative methods of genomics, proteomics, and structural chemistry. The combination of quantitative analysis, transgenesis, statistical inference and systems modeling provides a novel and comprehensive strategy to investigate the regulation, biosynthesis and properties of the secondary cell wall. We expect to answer two major questions: (1) To what extent can the abundance of individual transcripts and proteins of specific genes predict the quantity and composition of monomers and polymers of the cellulosic and lignin components? and (2) To what extent are there other new and necessary genes that regulate secondary cell wall properties?

This is a 3-year project and was recently funded. Here we focus our report on proposed research objectives.

SPECIFIC OBJECTIVES

1. **Transgene Perturbation:** We will generate transgenic *P. trichocarpa* with downregulated expression of 38 targeted cellulose and hemicellulose metabolic and *TF* genes. Artificial microRNA (amiRNA) and RNAi mediated gene specific knock-down experiments will be carried out in a differentiating xylem-specific manner at up to three levels: (a) the individual gene level, (b) the phylogenetic gene-pair level, and (c) the gene family level. For each transgene construct we will select transgenic lines with three distinct levels (highest suppression and two intermediates) of target gene knock-down to quantify transcriptomic, proteomic, and cell wall structural property responses to perturbations.

2. **Transcriptome Analysis:** Transgenics for each gene will be analyzed to test for specificity and interactive effects on transcripts, using new generation sequencing (Illumina GAIIX) and quantitative real-time PCR. We anticipate discovering the extent of specific and comprehensive feedback and feed-forward (loop) regulation associated with the organization and biosynthesis of cellulosic components in xylem secondary cell walls.

3. **Proteomic Analysis:** Changes in abundance of protein components will be determined by absolute protein quantification using Protein Cleavage coupled with Isotope Dilution MS (PC-IDMS)-based LC-MS/MS. Recombinant proteins will be produced from each target genes and purified for the identification of protein specific peptides (called surrogate peptides). Stable isotope labeled surrogate peptides will be synthesized and used as LC-MS/MS internal standards for absolute quantification of the target proteins in each transgenic

line. The approach allows us to quantify the response of target proteins to specific gene perturbations and to correlate changes in the proteome and transcriptome with changes in secondary cell wall properties.

4. **Quantitation of Cellulosic Characteristics:** After quantifying Klason lignin content in stem wood of each transgenic line, the remaining mass in solution will be separated by HPLC to quantify the sugar components and the polysaccharide contents. Wood cellulose crystallinity will be estimated by ^{13}C-CP/MAS solid state NMR. Composition and inter-unit linkages of LCCs will be quantified by 2D NMR.

5. **Database and Website:** A database/website will be set up at the beginning of the project. All project data will be deposited into a relational database accessible through our project website, forming the framework for statistical analysis and modeling.

6. **Statistical Analysis and Neural Networks:** Multivariate statistical (e.g. path analysis) methods will be explored to describe the degree, direction, and significance of relationships among all regulatory elements (TFs, gene transcripts and proteins) and cellulosic characteristics. Multilayer neural networks (MNNs) will be used as a quantitative approach to predict how variation in process protein concentrations influence cellulosic characteristics.

7. **Regulatory Signaling Graph**: Based on transgenic perturbations and statistical inference, modeling techniques will integrate experimental results to develop computational representations of cellulosic biosynthesis and structure. A mechanistic interaction graph will be developed to describe new regulation and to guide tests of feedback and feed-forward connections to verify functional interactions within the regulatory framework.

8. **Genome Wide Coregulated Network Analysis:** The large amount of new whole transcriptome and whole proteome information will be used for analysis of gene/protein variation to identify new coregulated associations of nontarget genes with known genes affecting secondary cell wall biosynthesis. The purpose of this analysis is to support and extend the structure of the signaling graph and the data-driven MNN models.

The outcome of this work will be a new and comprehensive description of the biosynthesis of cellulosic components of the plant secondary cell wall. This description will have predictive value and will serve to guide more efficient synthesis and degradation of the cell wall of energy crops for improved biofuels and biomass feedstocks.

This research was supported by the Office of Science (BER), U.S. Department of Energy.