

Section:
Biofuels Research



U.S. DEPARTMENT OF
ENERGY

Office of Science

2012

**Genomic Science Awardee
Meeting X**

**Bethesda, Maryland
February 26-29, 2012**

[Revised: March 1, 2012]

Prepared for the
U.S. Department of Energy
Office of Science
Office of Biological and Environmental Research
Germantown, MD 20874-1290

<http://genomicscience.energy.gov>

Prepared by
Biological and Environmental Research Information System
Oak Ridge National Laboratory
Oak Ridge, TN 37830
Managed by UT-Battelle, LLC
For the U.S. Department of Energy
Under contract DE-AC05-00OR22725

Biofuels Research

Analytical Technologies, Engineering, and Production

76

Phylogenetic Distribution of Potential Cellulases in Bacteria

Renaud Berlemont* (rberlemo@uci.edu) and Adam Martiny

Earth System Science and Ecology and Evolutionary Biology, University of California, Irvine

http://allison.bio.uci.edu/DOE_Microbes.html

Project Goals: The main goal of this project is to connect diverse microbial groups with the extracellular enzyme systems that catalyze the decay of organic material. We will also determine whether different groups of microbes and their enzymes respond to environmental changes, and whether they can recover from such changes. Finally, we will develop mathematical models to predict the responses of microbial communities and their associated functions under new environmental conditions.

In most terrestrial ecosystems, the depolymerization of plant cell wall is the rate limiting step in the turnover of organic material. The composition of plant detritus is known to depend mainly on enzymes produced by microorganisms. This raises the question: which phylogenetic lineages of microorganisms can degrade plant cell wall material, including cellulose?

To address this question, we compared the distribution of Glycoside Hydrolases (GH) potentially related to the cellulose degradation among 3744 bacteria.

Some phylogenetic groups are especially rich in GHs whereas some are very poor. For example, in bacteria from the Bacteroidetes phylum ~40 GHs (from the families 1, 3, 5, 6, 8, 9, 10, 12, 16, 45, 48) are described (per genome) but in the Chlorobi phylum less than 5 of these protein-encoding genes are observed, per sequenced genome. The others bacteria that lack these GHs belong to the Aquificae, Chlamydiae, Chrysiogenetes, Cyanobacteria, Deferribacter, Elusomicrobia, Fusobacteria, Gemmatimonadetes, Synergistetes, Tenericutes and Nitrospirae phyla.

In addition, when present in genomes, not all the GH families are equally observed. Indeed, enzymes from the family 6, 8, 9, 45, and 48 are 1.5 to 3 times less abundant than enzymes from the family 1, 3 and 5. The over-representation of these enzymes is related to their numerous activities (e.g. GH1: β -glucosidase; β -galactosidase; β -mannosidase; β -glucuronidase; exo- β -1,4-glucanase; etc.) whereas the

enzymes with low abundance display only few activities (e.g. GH6: endoglucanase and cellobiohydrolase).

For each phylum, we defined an abundance profile (for GH1, 3, 5, 6, 8, 9, 10, 12, 16, 45 and 48). These profiles were compared and clustered in order to discriminate the potential cellulose degraders from the other organisms. This clustering highlights the adaptation of microorganisms to different “life-styles”. For example, autotrophic bacteria (e.g. Cyanobacteria) and/or intracellular pathogens (e.g. Tenericutes) generally lack many enzymes involved in polymers degradation, outside the cell (e.g. GH3, 5, 6, 8).

The association of ‘Carbohydrate Binding Modules’ (CBM) to catalytic domains (GH) is assumed to potentiate the catalytic activity by proximity effects. The association of GHs, from the above mentioned GH families, with CBMs from the family 2, 3 and 4 was thus investigated. Surprisingly, the CBM-GH profiles differ significantly from the profiles of their ‘free enzymes’ relatives. The association of CBM to GHs is a signature for enzymes involved in the catabolism of cellulose whereas the physiological function of ‘free’ enzymes is still pending (anabolism and/or catabolism).

77

Early Career Program

Engineering Robust Hosts for Microbial Biofuel Production

Mary Harrison and Mary J. Dunlop* (mjdunlop@uvm.edu)

School of Engineering, University of Vermont, Burlington

Project Goals: The goal of this project is to develop tools for improving microbial tolerance of biofuel production conditions. We aim to develop specific strategies for improving tolerance to enhance microbial synthesis of next-generation biofuels. The work is organized around three objectives: (1) Identify novel biofuel tolerance mechanisms from a targeted set of microorganisms. We are focusing on microorganisms that survive in hydrocarbon-rich environments as a source of tolerance genes, looking specifically for strategies that prevent biofuel from compromising membrane integrity and mechanisms for exporting biofuel-like compounds from the cell. (2) Engineer a synthetic feedback loop that responds to biofuel production. To optimize biofuel production yields, cells must balance several competing sources of stress. We are designing and constructing a novel feedback loop that senses biofuel production and turns on export pumps in response. (3) Finally, we are integrating multiple tolerance strategies in a biofuel production strain. In addition to

having the potential to greatly enhance biofuel yields, this work will advance understanding of how multiple stress tolerance mechanisms interact within a cell.

Microorganisms can be used to synthesize fuel from renewable materials. Microbial biofuel synthesis is a cost effective and environmentally sustainable way of producing replacements for gasoline, diesel, and jet fuel from lignocellulosic biomass. In a typical production process, biomass is deconstructed into sugars that are metabolized by a microbe engineered to convert sugar into biofuel. In this work, we focus on fuel synthesis—the final stage of biofuel production—and develop engineering tools for increasing the robustness of a biofuel production host.

A major challenge when using microorganisms to produce bulk chemicals like biofuels is that the production targets are often toxic to cells. Biofuel-like compounds are known to reduce cell viability through damage to the cell membrane and interference with essential physiological processes. Thus, cells must trade off biofuel production and survival, reducing potential yields. The majority of microbial biofuel research to-date has focused on engineering metabolic pathways for biofuel production. It is essential that we also engineer strains for biofuel tolerance. This is especially important for the large-scale production environments that will be required for generating cost effective biofuels.

Recent work on ethanol has already demonstrated that engineering strains for end-product tolerance can greatly improve overall yield. Here, we focus on engineering tolerance to advanced biofuels, taking a targeted approach toward improving biofuel tolerance and yield. Our goal is to develop a list of rational targets, rather than employing strategies like adaptation and random mutagenesis, to identify mechanisms that can be easily moved into production strains.

Recent work by the PI has indicated that microorganisms that survive in oil-rich environments are a valuable source of tolerance mechanisms [1-3]. We are testing genes from microbes that have been isolated from environments near natural oil seeps and in the vicinity of oil spills to see if they improve biofuel tolerance when expressed in *Escherichia coli*. Whole genome sequencing efforts have included several of these hydrocarbon-tolerant microbes. We are using this sequence data as a starting point for identifying tolerance genes to heterologously express in a biofuel production host.

References

1. Dunlop, M.J., et al., *Regulatory activity revealed by dynamic correlations in gene expression noise*. *Nature Genetics*, 2008. 40(12): p. 1493-1498.
2. Dunlop, M.J., et al., *Engineering microbial biofuel tolerance and export using efflux pumps*. *Mol Syst Biol*, 2011.
3. Dunlop, M.J., J.D. Keasling, and A. Mukhopadhyay, *A model for improving microbial biofuel production using a synthetic feedback loop*. *Systems and Synthetic Biology*, 2010. 4(2): p. 95-104.

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

78

Engineering Ethylmalonyl-CoA Pathway in *Methylobacterium extorquens* AM1 for Butanol Production: Identification of a Regulator Activating the Expression of a Key Gene

Bo Hu^{1*} (bohu@uw.edu), David Beck,¹ Robert Synovec,² and Mary Lidstrom^{1,3}

¹Department of Chemical Engineering, ²Department of Chemistry, and ³Department of Microbiology, University of Washington, Seattle

<https://depts.washington.edu/mlab/>

Project Goals: The main goal of this project is to characterize control points for assimilatory metabolism in a facultative methylotroph, *Methylobacterium extorquens* AM1. Methylotrophs have potential for conversion of reduced one-carbon compounds to value-added chemicals, but a more detailed understanding of assimilatory metabolism is necessary for that potential to be reached. This project will generate systems-level datasets to inform optimization approaches for a new platform for biofuels synthesis, tools for systems approaches via multi-level datasets, and starting strains for biofuels strain development using methanol + CO₂ as biofeedstocks.

Our initial work focused on potential regulatory genes for the ethylmalonyl-CoA (EMC) pathway, the part of assimilatory metabolism that involves crotonyl-CoA as an intermediate. We have identified CcrR, a TetR-type activator, which has been shown to regulate the expression of crotonyl-CoA reductase/carboxylase, a key enzyme of EMC pathway. The *ccrR* mutant is impaired in its ability to grow on C1 and C2 compounds, correlating with the reduced activity of crotonyl-CoA reductase/carboxylase. The *ccr* gene was found to be cotranscribed with an upstream gene (*katA*) and the promoter strength was found to reduce as much as 50% in the absence of CcrR compared to that in wild type *M. extorquens* AM1. Gel retardation assays with purified His-tagged CcrR showed that Ccr binds to the promoter-regulatory region of *ccr*. A palindromic sequence upstream of *katA* at position -334 to -321 with respect to the translational start site was found to be the binding site and mutations in this region eliminated gel retardation with CcrR. These results show that CcrR stimulates expression of the *katA-ccr* promoter on the order of two-fold but is not required for this expression. The identification of a specific activator protein regulating expression of one of the genes of the EMC pathway is a first step in understanding how the pathway as a whole is regulated, and this work also generates information for manipulating flux to end products that use crotonyl-CoA as precursor, such as butanol.

As a proof of principle example for manipulation of carbon flux through crotonyl-CoA, we have initiated studies to introduce a two-step pathway into *M. extorquens* AM1 to convert crotonyl-CoA into butanol. The initial step would be reduction of crotonyl-CoA to butyryl-CoA by

a crotonyl-CoA specific trans-enoyl-CoA reductase from *Treponema denticola* (Ter), coupled with another reduction step catalyzed by alcohol dehydrogenase from *Clostridium acetobutylicum* ATCC8244 (AdhE2) to produce butanol. Genes encoding Ter and AdhE2 were synthesized with codon usage optimized using a codon usage table derived from highly expressed genes of *M. extorquens* AM1 grown on methanol. Each gene was subcloned into *M. extorquens* AM1-adaptable expression vectors for expression. Enzyme assays showed that both Ter and AdhE2 activities were detected in recombinant *M. extorquens* AM1 strains containing plasmids with *ter* or *adhE2*. The immediate future plan is to integrate both genes into a single operon with various expression capabilities to develop a strain with high butanol productivity.

This work is supported by Office of Biological and Environmental Research in the DOE Office of Science.

79

Integration of Carbon, Nitrogen, and Oxygen Metabolism in *Escherichia coli*

Yi-Fan Xu^{1,2*} (yifanxu@Princeton.EDU), Christopher D. Doucette,^{1,3} Junyoung O. Park,^{1,4} Nathaniel Roquet,¹ Daniel Amador-Noguez,¹ Marshall L. Reaves,^{1,3} Xiaojiang Feng,² Herschel Rabitz,² David J. Schwab,^{1,3} Ned S. Wingreen,^{1,3} and **Joshua D. Rabinowitz**^{1,2}

¹Lewis-Sigler Institute for Integrative Genomics, ²Department of Chemistry, ³Department of Molecular Biology, and ⁴Department of Chemical and Biological Engineering, Princeton University, Princeton, N.J.

Project Goals: A key challenge for living systems is balancing utilization of different elemental nutrients, such as carbon, nitrogen, and oxygen. Although the regulation of specific nutrient assimilation systems has been extensively studied, how these systems coordinate with each other remains poorly understood. Here we aim to obtain a quantitative understanding of such coordination in *Escherichia coli* with a focus on fast-acting regulatory mechanisms. To this end we are (1) quantifying metabolic responses to nutrient perturbations using metabolomics, (2) building differential equation models that bridge multiple nutrient systems, and (3) discovering regulatory principles from this unified combination of metabolomics and modeling.

Results:

Biomass production requires the integration of multiple external nutrients, whose availability is subject to environmental fluctuations. As growth can be limited by the scarcity of any one nutrient, the rate at which each nutrient is assimilated must be sensitive not only to its own availability, but also to that of other nutrients. Remarkably, across diverse nutrient conditions, *E. coli* grow nearly optimally: they balance effectively the conversion of carbon into energy versus biomass, and excrete only small amounts of waste.

What type of regulatory architecture might (i) enable homeostasis of intracellular metabolite concentrations and (ii) produce metabolic fluxes that nearly optimize growth? The most basic homeostatic regulatory mechanism is feedback inhibition. Through a series of reductionist differential equation models, with fluxes expressed in Michaelis-Menten form, we found that *feedback inhibition alone is sufficient to meet both of the above goals*. While such a result was expected for a linear biosynthetic pathway, it was striking that feedback inhibition can in theory also effectively regulate metabolic cycles, the portioning of carbon between energy production and biomass, and the integration of multiple nutrients (e.g., nitrogen and carbon).

We then considered how actual feedback regulatory architectures compare with those predicted to be effective based on our conceptual analysis. *E. coli*'s ammonia assimilation cycle, involving the enzymes glutamine synthetase (GS) and glutamate synthase (GOGAT), is among the best-studied metabolic cycles from the perspective of regulation. Despite this, existing literature focuses on a single feedback (from glutamine on GS), whereas our analysis indicated the need also for a feedback from glutamate on GOGAT. Through detailed metabolomic analysis and modeling of the cycle, including analysis of various metabolic and regulatory mutants, we showed that a feedback on GOGAT is indeed required, and that surprisingly it is provided by aspartate, whose concentration mirrors that of glutamate.

To investigate the link between carbon and nitrogen metabolism, we measured changes in the glucose uptake rate in response to changing nitrogen availability. We found that glucose uptake in *E. coli* is subject to rapid regulation by nitrogen. Remarkably, nitrogen up-shift doubles glycolytic flux in 2 min without substantial changes in the concentration of any glycolytic intermediate. What type of feedback might enable such a response? The most elegant possibility involved the carbonaceous substrate of nitrogen assimilation α -ketoglutarate directly inhibiting the phosphotransferase system, which simultaneously uptakes glucose and drains the bottom of the glycolysis pathway. Such a feedback loop, never previously proposed, was verified biochemically and shown genetically to be both *necessary and sufficient for nitrogen-based regulation of sugar uptake*.

Much of central carbon metabolic flux is altered by changing oxygen levels. Nevertheless, there are parallels to the case of nitrogen metabolism: α -ketoglutarate, the product of carbon metabolism that directly feeds nitrogen metabolism, controls carbon metabolic response to nitrogen availability; similarly, we believe that NADH, the product of carbon metabolism that immediately feeds into oxygen metabolism, controls carbon metabolic response to oxygen availability. NADH levels impact carbon metabolism in multiple ways, including shutting off the right side of the tricarboxylic acid (TCA) cycle, reversing the left side of the TCA cycle, and inhibiting glycolysis. This is basically feedback inhibition writ large.

Finally, we considered when feedback inhibition might be insufficient as a metabolic regulatory strategy. Our analyses suggested its sufficiency for all steady-state conditions.

Therefore, we considered the possible importance of other regulatory motifs in oscillating conditions. A proposed feed-forward circuit involves activation of phosphoenolpyruvate carboxylase (the anapleurotic enzyme in *E. coli* that provides new 4 carbon units into the TCA cycle) by fructose-1,6-bisphosphate. We verified genetically the existence of this regulation, found that the regulation is ultrasensitive, and showed that ultrasensitive activation of this enzyme by fructose-1,6-bisphosphate markedly enhances *E. coli* growth specifically during oscillating glucose conditions.

Taken together, these findings support a general mechanism for nutrient integration: *Limitation for a nutrient other than carbon leads to build-up of the most closely related product of carbon metabolism, which in turn feedback inhibits further carbon metabolism.* Our efforts to translate this principle into a predictive, quantitative model of central carbon, nitrogen, and oxygen metabolism are ongoing. Eventual success will require intensive re-examination of the regulation of many enzymes, with the ultimate payoff being enhanced understanding and rational engineering of bacterial metabolism.

80

Metabolomics of Clostridial Biofuel Production

Daniel Amador-Noguez,¹ Ludmilla Aristilde,¹ Caleb Bastian,² Herschel Rabitz,² Leonid Kruglyak,¹ and **Joshua D. Rabinowitz**,^{1,2*} (josh@genomics.princeton.edu)

¹Lewis-Sigler Institute for Integrative Genomics and
²Department of Chemistry, Princeton University, Princeton, N.J.

Project Goals: The production of biofuels from cellulosic biomass holds promise as a source of renewable clean energy. Members of the genus *Clostridium* collectively have the ideal set of the metabolic capabilities for biofuel production from cellulosic biomass: *C. acetobutylicum* rapidly ferments glucose to biofuels (butanol, hydrogen) and *C. cellulolyticum* effectively degrades cellulose. Here we aim to integrate metabolomics, genomics and genetic engineering to dramatically advance understanding of metabolism in *C. acetobutylicum* and *C. cellulolyticum*. In so doing, we will lay basic science groundwork for engineering of an organism that cost-effectively converts cellulose into solvents and/or hydrogen gas.

Results and Plans:

The fermentation carried out by *C. acetobutylicum* is characterized by production of acids during exponential growth ("acidogenesis"), followed by production of solvents as growth slows down ("solventogenesis"). We previously used metabolomics and quantitative flux modeling to map the metabolic changes associated with this acidogenic-solventogenic transition. This effort revealed a marked reorganization of central metabolism that involved both large changes in intracellular metabolite concentrations and metabolic

fluxes. Solventogenesis was characterized by a dramatic down regulation of fluxes through pyruvate carboxylase, the reductive TCA cycle, and amino acid biosynthesis. These flux changes favor build-up of pyruvate and reducing power, the substrates of solventogenesis. Recently, we have knocked out pyruvate carboxylase, and find that this facilitates transition into solventogenesis, improves acid re-assimilation, and slightly increases final solvent yields.

Another key aspect in biofuel production from cellulose is hemicellulosic sugar catabolism. We previously mapped the pathways of glucose metabolism in *C. acetobutylicum* by following the dynamics of ¹³C-labeled glucose assimilation. Here we are conducting similar isotope tracer experiments to explore the simultaneous utilization of glucose and hemicellulosic sugars (xylose, arabinose, mannose, and galactose). Our initial experiments show that while galactose is not assimilated in the presence of glucose, the organism simultaneously and non-discriminantly catabolizes both glucose and mannose. Similarly, both xylose and arabinose are assimilated in the presence of glucose not only into the pentose phosphate pathway, but also lower glycolysis. There is no assimilation into upper glycolysis, however. This suggests that, in the presence of pentose sugars, there is minimal activity of the transaldolase reaction, which yields fructose-6-phosphate. Follow-up experiments with different types of ¹³C-labeled glucose and xylose are underway to confirm this hypothesis. In addition, we are working to model quantitatively the associated fluxes.

Finally we aim to understand *C. cellulolyticum*'s catabolic limitations. While an impressive cellulose degrader, *C. cellulolyticum* is a sluggish fermenter (20-fold slower than *C. acetobutylicum*). We will apply metabolomics and genomics to investigate the reasons for its slow sugar catabolism.

Funding: DOE DE-SC0006839.

81

Experimental Systems Biology Approaches for Clostridia-Based Bioenergy Production: The Metabolite Stress-Response System in Solventogenic Clostridia

E. Terry Papoutsakis^{1,2*} (epaps@udel.edu), Blake C. Meyers,^{2,3} Kelvin H. Lee,^{1,2} Maciek R. Antoniewicz,¹ Costas D. Maranas,⁴ H. Huang,² and **Cathy H. Wu**^{2,5}

¹Department of Chemical Engineering, University of Delaware, Newark; ²Delaware Biotechnology Institute, Newark; ³Department of Plant and Soil Sciences, University of Delaware, Newark; ⁴Department of Chemical Engineering, Pennsylvania State University, University Park; and ⁵Department of Biochemistry and Molecular and Cellular Biology, Georgetown University Medical Center, Washington, D.C.

<http://research.bioinformatics.udel.edu/clostridia-stress-response>

Project Goals: The objectives of this project are to engage enabling experimental systems-biology approaches to support the development of integrated, predictive models of the metabolic and regulatory networks underlying the metabolite stress response in solventogenic clostridia. Clostridia are Gram⁺, obligate anaerobic, endospore-forming bacteria of major importance to fermentative biofuel production. Here, we focus on understanding and modeling the stress-response of *Clostridium acetobutylicum* to two important toxic metabolites: butanol and butyrate. Systems-level understanding is expected to lead to better strategies for industrial-strain development, as well as bioprocessing strategies taking advantage of the stress response to achieve superior bioprocessing outcomes.

Solventogenic and other clostridia are of major importance for developing technologies for biofuel production (3). A major and unique advantage is their ability to utilize a large variety of substrates (hexoses, pentoses, oligosaccharides, xylan, and starches). Among the two sequenced solventogenic clostridia, *C. acetobutylicum* is the only one that contains a full cellulosome (2) and may thus directly utilize cellulosic material for production of fuels and chemicals.

The toxic-metabolite stress response is a problem of major and general importance not only in clostridial biotechnologies but in all microbial systems of interest to bioenergy production (1). Achieving the goals of this project will solve a long-standing problem of both fundamental and practical importance. The significant broader impact is that the approaches and tools developed here can be applied to other cellular systems, aiming to achieve the ultimate goal of comprehensive multidimensional understanding of cells via Genome Scale Model (GSM) multidimensional platforms and models. The outcomes of this project aim to become an enabling paradigm for modeling complex programs of organisms and biological systems of importance to DOE's mission on energy and the environment.

To achieve these goals, we apply genomic tools for collecting extensive transcriptomic (based on both deep sequencing and microarray analyses) and targeted fluxomic and proteomic data, and couple omics data integration with building stress models and modeling platforms that can be linked, as an added modeling dimension, to a 2nd generation GSM of this organism. Successful completion of this project aims to the following deliverables: (i) experimental omics data of the metabolite stress response in *C. acetobutylicum* for deposition to public repositories and that can be mined for systems-level predictive modelling; (ii) an in-depth systems-level molecular understanding at multiple genome-scale levels of the metabolite stress response; (iii) two models and modeling frameworks for the metabolite stress responses linking them to the constructed 2nd generation GSM; and, (iv) computational and bioinformatic tools and frameworks that can be applied to modeling other cellular programs.

References

1. Nicolaou, S. A., S. M. Gaida, and E. T. Papoutsakis. 2010. A comparative view of metabolite and substrate stress and tolerance in microbial bioprocessing: From biofuels and chemicals, to biocatalysis and bioremediation. *Metab Eng* 12:307-331.
2. Nölling, J., G. Breton, M. V. Omelchenko, K. S. Makarova, Q. Zeng, R. Gibson, H. M. Lee, J. Dubois, D. Qiu, J. Hitti, Y. I. Wolf, R. L. Tatusov, F. Sabathe, L. Doucette-Stamm, P. Soucaille, M. J. Daly, G. N. Bennett, E. V. Koonin, and D. R. Smith. 2001. Genome sequence and comparative analysis of the solvent-producing bacterium *Clostridium acetobutylicum*. *J Bacteriol* 183:4823-4838.
3. Tracy, B. P., S. W. Jones, A. G. Fast, D. C. Indurthi, and E. T. Papoutsakis. 2011. Clostridia: the importance of their exceptional substrate and metabolite diversity for biofuel and biorefinery applications. *Curr Opin Biotechnol* doi: 10.1016/j.copbio.2011.10.008.

82

Population Level Analysis of Mutations Underlying Improvements in Biofuel Production by *Clostridium phytofermentans*

Supratim Mukherjee, Stephen Godin, and **Jeffrey L. Blanchard*** (blanchard@microbio.umass.edu)

Department of Microbiology, University of Massachusetts, Amherst

<http://www.bio.umass.edu/micro/blanchard/>

Project Goals: Although cellulosic plant biomass is plentiful and cheap, the cost of degrading the cellulosic and hemicellulosic components of the plant cell wall currently limits the use of plant biomass as a competitive renewable alternative to gasoline and is a key challenge in developing a U.S. biomass-based industry for manufacturing biofuels from agricultural and forestry wastes. The goal of this project is to determine the mechanistic basis of improvements in *C. phytofermentans* strains evolved to produce biofuels more rapidly. Understanding the mechanistic details of individual mutations will be of particular interest to the larger scientific community for improving rates of biomass degradation and biofuel production.

Clostridium phytofermentans is a genetically tractable, anaerobic bacterium isolated from forest soil near the Quabbin Reservoir and the Harvard Forest Long Term Ecological Research site in Massachusetts, U.S.A. *C. phytofermentans* (1) can saccharify all major carbohydrate components of plant biomass; (2) has acquired many biofuel-related enzymes through horizontal gene transfer; and (3) produces ethanol as the primary product of plant carbohydrate fermentation. The combination of its broad nutritional versatility coupled with its high levels of ethanol production distinguishes *C. phytofermentans* from all other cultured microbes characterized to date. Using long-term experimental evolution we have selected for strains of *C. phytofermentans* that grow and produce ethanol faster on plant carbohydrates. Full genome re-sequencing through the DOE Community Sequencing Program has allowed us to identify mutations present at varying allele frequencies in cultures with increased rates of biomass degradation. The results reveal

new strategies for evolving and engineering microorganisms for the production of biofuels from plant feedstocks.

This research was funded by an Isenberg Scholar Award to S. Mukherjee, a Howard Hughes Medical Institute Fellowship to S. Godin and a DOE Community Sequencing Program award to J. Blanchard.

83

Understanding Fundamental Aspects of Butanol Production by *Clostridium beijerinckii*

Caroline B. Milne,^{1,2} Yi Wang,¹ Beth Papanek,¹ Yong-Su Jin,¹ Nathan D. Price,^{1,2} and Hans P. Blaschek^{1*}
(blaschek@illinois.edu)

¹University of Illinois, Urbana and ²Institute for Systems Biology, Seattle, Wash.

<http://bioenergy.illinois.edu>

Project Goals: The solventogenic clostridia offer a sustainable approach to petroleum-based production of n-butanol, an important chemical feedstock and potential fuel. With the availability of the genome sequence for *Clostridium beijerinckii* 8052, we can now employ the tools of systems biology in order to gain increased insight into the metabolic and regulatory networks relevant to solvent production. Project goals include examination of: 1) the mutations underlying the *C. beijerinckii* BA101 butanol-overproduction phenotype, 2) the molecular basis for the global shift from acidogenesis to solventogenesis, 3) the genetic basis of butanol tolerance in *C. beijerinckii* and 4) RNA-seq technology for single-nucleotide resolution analysis of the transcriptome of this microorganism.

We have constructed the first genome-scale metabolic model (*i*CM925) for *C. beijerinckii*, containing 925 genes, 938 reactions, and 881 metabolites. To build the model we employed a semi-automated procedure that integrated genome annotation information from KEGG, BioCyc, and The SEED, and utilized computational algorithms with manual curation to improve model completeness. To validate *i*CM925, we conducted fermentation experiments using the NCIMB 8052 strain, and evaluated the ability of the model to simulate measured substrate uptake and product production rates. Experimentally observed fermentation profiles were found to lie within the solution space of the model; however, under an optimal growth objective, additional constraints were needed to reproduce the observed profiles—suggesting the existence of selective pressures other than optimal growth. Notably, a significantly enriched fraction of actively utilized reactions in simulations—constrained to reflect experimental rates—originated from the set of reactions that overlapped between all three annotation databases used ($P = 3.52 \times 10^{-9}$, Fisher's exact test). Inhibition of the hydrogenase reaction was found to have a strong effect on butanol formation—as experimentally observed.

We also conducted a single-nucleotide resolution analysis of the *C. beijerinckii* NCIMB 8052 transcriptome using

high-throughput RNA-Seq technology. We identified the transcription start sites and operon structure throughout the genome. We confirmed the structure of important gene operons involved in metabolic pathways for acid and solvent production in *C. beijerinckii* 8052, including *pta-ack*, *ptb-buk*, *hbd-ETF A-ETF B-crt* (*bcs*) and *ald-ctfA-ctfB-adc* (*sol*) operons; we also defined important operons related to chemotaxis/motility, transcriptional regulation, stress response and fatty acids biosynthesis along with others. We discovered 20 previously non-annotated regions with significant transcriptional activities and 15 genes whose translation start codons were likely mis-annotated. As a consequence, the accuracy of existing genome annotation was significantly enhanced. Furthermore, we identified 78 putative silent genes and 177 putative housekeeping genes based on normalized transcription measurement with the sequence data. We also observed that more than 30% of pseudogenes had significant transcriptional activities during the fermentation process. Strong correlations exist between the expression values derived from RNA-Seq analysis and microarray data or qRT-PCR results.

References

1. Milne, C.B., et al. 2011. Metabolic network reconstruction and genome-scale model of butanol-producing strain *Clostridium beijerinckii* NCIMB 8052. *BMC Systems Biology*, 5:130.
2. Wang, Yi., et al. 2011. Single-nucleotide resolution analysis of the transcriptome structure of *Clostridium beijerinckii* NCIMB 8052 using RNA-Seq. *BMC Genomics*, 12:479.

This work was funded by the DOE Office of Science through FOA: DE_FOA-0000368 and an NSF CAREER grant.

Early Career Award

84

Applying the Biology of Brown Rot Fungi to Consolidated Bioprocessing—Early Career Program

Robert A. Blanchette,¹ Shona M. Duncan,¹ Timothy Filley,² Joel Jurgens,¹ Justin Kaffenberger^{1*} (kaffe002@umn.edu), and Jonathan S. Schilling¹

¹University of Minnesota, Saint Paul and ²Purdue University, West Lafayette, Ind.

<http://schillinglab.cfans.umn.edu/projects-pubs>

Project Goals: *Postia placenta* is a wood-degrading brown rot fungus with a sequenced and annotated genome^[1]. Like other brown rot fungi, *P. placenta* is capable of an oxidative hydroxyl radical pretreatment that occurs concurrently with enzymatic saccharification of woody carbohydrates. This consolidation of otherwise incompatible reactions is fundamentally interesting and has great implication on the potential to consolidate harsh pretreat-

ments with saccharification in a single processing step. Therefore, our research goals are as follows:

1. physically sample wood degraded by the brown rot fungus *P. placenta* in order to map coincident pretreatment and saccharification reactions and to correlate relevant lignocellulose chemistry,
2. image pH and porosity at the fungus-plant interface and layer this data with images showing cellulase ingress, and
3. map, along the active hyphal front, the co-occurring expression of iron reductases associated with pretreatment and of cellulases used in saccharification.

Enzymatic bioconversion of lignocellulose plant tissues generally requires an initial pretreatment step, followed by saccharification and then fermentation or other downstream processing approaches. Consolidated bioprocessing (CBP) of lignocellulose combines enzymatic sugar release (saccharification) with fermentation, but pretreatments typically remain separate and costly. In nature, lignocellulose-degrading brown rot fungi consolidate pretreatment and saccharification, likely using spatial gradients to partition these incompatible reactions. Our goal is to characterize how this is achieved, in order to better understand the fungus and to potentially apply this approach in a mimicked consolidated approach.

The goal of this research is characterizing this relevant biological system, with objectives (stated above) to 1) physically sample wood degraded by the brown rot fungus *Postia placenta* to map reactions spatially and to correlate with cell wall modifications, 2) produce images of the environmental variables (pH and porosity) affecting cellulase ingress over time during brown rot, and 3) map, along the active hyphal front, the co-occurring expression of iron reductases associated with pretreatment and of cellulase involved in saccharification. These are spatially-focused goals. Therefore, my respective approaches involve either small-scale, spatially resolved characterization (Obj. 1), or appropriately resolved microscopy (Obj. 2 and 3).

Small-scale physical sample analysis includes traditional wet chemical characterization, coupled both with spin-trap adduct recovery of hydroxyl radicals produced by the fungus and with C13-labeled tetramethyl ammonium hydroxide thermochemolysis for specific brown rot lignin modifications. For microscopy, I am utilizing fluorescence lifetime imaging (FLIM) with confocal detection for pH measurements, cryo-transmission electron microscopy (TEM) with electron tomography for porosity measures, a complementary scanning transmission x-ray approach for porosity, and traditional TEM with immunolabeling to track cellulase ingress. I am also planning to co-localize chitinous fungal biomass, imaged using a traditional WGA-FITC dye, with fluorescence in-situ hybridization to measure mRNA transcribed from iron reductase and endoglucanase DNA sequences, made possible by the recent JGI-funded annotation of the *P. placenta* sequence. Overall, this research will help resolve how brown rot fungi consolidate oxidative pretreatments with enzyme-based saccharification, so that we might better understand and exploit natural synergies

between bioconversion steps that we currently approach as separate, distinct steps.

Reference

1. Martinez et al. (40+ authors) (2009) Genome, transcriptome, and secretome analysis of wood decay fungus *Postia placenta* supports unique mechanisms of lignocellulose conversion. *PNAS* 106:1954-1959.

85

Development of Quantum Dot Probes for Studies of Synergy Between Components of the Wood-Degrading Fungal Enzymes

Haw Yang,¹ Kevin Welsher,¹ Chih-Hao Hsia,¹ Daniel Montiel,¹ Robin Rohwer^{2*} (rrr24@psu.edu), B. Tracy Nixon^{2*} (btn1@psu.edu), and Ming Tien²

¹Princeton University, Princeton and ²The Pennsylvania State University, University Park

Project Goals: The overarching goal of this research program is to understand the cellular and molecular mechanisms that underlie the fungus-enzyme-lignocellulose gestalt that will enable the efficient and economical production of liquid fuels from plant biomass. The current project aims to develop quantum dot-based tagging and imaging technologies to simultaneously monitor several lignocellulosic enzymes in real-time and in the natural fungal/lignocellulose environment. The specific goals are: 1) Synthesis of commercially unavailable multiplexed luminescent quantum dots optimized for the study of lignocellulosic degradation. 2) Development of bio-orthogonal tags and protocols that can be generalized to other wood-degrading organisms for labeling secreted enzymes under realistic enzyme-biomass conditions. 3) Advancement of single-molecule imaging technology and the development of assays for studying individual lignocellulosic enzymes.

To achieve these objectives a team with highly complementary expertise was assembled: Tien of the Pennsylvania State University (co-PI) is the co-discoverer of the lignin-degrading enzyme lignin peroxidase, and is a world expert in fungal lignin biodegradation. He has extensive experience and knowledge of the enzymology, genetics, and biogenesis of lignocellulosic enzymes, as well as the culturing and genetic manipulation of white-rot fungi. Nixon of The Pennsylvania State University (co-PI) is the co-discoverer of the ubiquitous two-component signaling system, a central signal transduction pathway in bacterial, fungal and plant cells involved in response to environmental stimuli. He has experience using genetic, molecular, biochemical and structural approaches to link two-component signal transduction to the gene regulation machinery. Yang of Princeton University (PI) is an expert in the development of single-molecule assays and optical probes, and in using them to solve biological and biochemical problems. One of the instruments developed in his group allows real-time tracking

of individual nanoscale optical probes in three dimensions with microsecond (μ s) time resolution and nanometer localization resolution and will be used as the primary tool for this work. This team is uniquely able to discover how fungal enzymes synergistically degrade lignocelluloses, and the tools developed to accomplish this will be broadly applicable to study other systems in their natural microenvironments.

Goal 1 is to synthesize multiplexed luminescent quantum dots specifically optimized for the study of lignocellulosic degradation which are otherwise unavailable. The design stage for synthesis of compact and water-soluble non-blinking quantum dots is finished. Doped quantum dots that have multiple emitters in a single dot were developed as an improvement to conventional quantum dots which need a thick shell to suppress blinking behavior. The blinking behavior is overcome in doped quantum dots because all of the emitters (dopants) are unlikely to stay in the dark state at the same time when the quantum dots have a high doping concentration. With the bio-orthogonal tagging technologies being developed, the compact and water-soluble non-blinking quantum dots can be used as markers to follow the translational motions of enzymes.

Goal 2 is to develop bio-orthogonal tags and protocols that can be generalized to other wood-degrading organisms for simultaneously labeling multiple secreted enzymes under realistic enzyme-biomass conditions. The designs for engineering *Trichoderma reesei* to express bio-orthogonally tagged variants of Cel7A, Cel6A and Cel7B are finished. Our strategy improves upon the work reported in the existing literature by providing forward and reverse selections for introducing any recombinant form of the cellulase genes of interest in a way that leaves absolutely no footprint other than the desired change. In accomplishing changes this way, our manipulations will be cleanly restricted to the desired changes in a targeted cellulase. This maximizes our ability in the future to examine the behavior of secreted enzymes in a wood matrix that is being normally degraded by the fungus. The strategy also streamlines the process of testing many recombinant genes by providing for quick construction of desired genes and introducing them back into *T. reesei* via a single homologous recombination event. Strain construction is underway, with preliminary results anticipated by the time of the meeting.

Goal 3 is to advance single-molecule imaging technology and develop assays for studying individual lignocellulosic enzymes. To conquer the multiple scales covered by the action of lignocellulosic enzymes, the previously reported real-time single-particle spectroscopy (RT-3D-SPTS) method has been combined with a scanning two-photon microscope to create a high-resolution, multiscale spectral imaging system. The RT-3D-SPTS is implemented in the near infrared (NIR, 650 nm - 800 nm) region of the electromagnetic spectrum with the intent to avoid intrinsic autofluorescence of biological samples, which is typically confined to shorter wavelengths. The RT-3D-SPTS system has spatial localization resolution down to 10 nm in the lateral dimensions and 20 nm in the axial dimensions, with temporal resolution of down to 10 microseconds. The

system is also fitted with a z-locking mechanism that counteracts objective drift to ensure fidelity of position measurement during tracking. The scanning-two photon microscope covers the larger length scales (up to 200 microns) allowing concurrent sampling of the substrate environment while the quantum dot tagged enzyme is investigated spectroscopically by the RT-3D-SPTS system. Crosstalk between the two modules is minimized by spectral separation, reaching further into the NIR (<800 nm) for two-photon excitation and collecting images in the visible (400 nm - 600 nm). This allows for imaging of conventional dyes, fluorescent proteins or autofluorescence. The system is further outfitted for biological samples with an onboard temperature control chamber. This multiscale imaging system has already been used to track freely diffusing nanoparticles in the presence of fluorescently labeled single cells.

This material is based upon work supported by the Department of Energy under Award Number DE-SC0006838.

86

Biodegradative Oxidant Production by Fungi in Lignocellulose: In Situ Mapping and Gene Expression

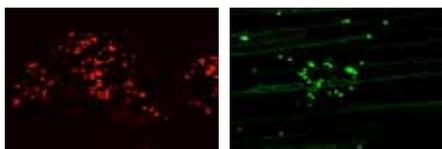
Premasagar Korripally,^{1,2} Peter Kitin,^{1,2} Christopher G. Hunt,² Carl J. Houtman,² **John Ralph**,¹ and **Kenneth E. Hammel**^{1,2*} (kehammel@wisc.edu)

¹University of Wisconsin, Madison and ²U.S. Forest Products Laboratory, Madison, Wis.

Project Goals: Naturally occurring, biological mechanisms for lignocellulose deconstruction are of interest because they may inspire the development of environmentally friendly technologies for conversion of lignocellulose into useful fuels and chemicals. The only organisms that degrade lignocellulose efficiently are certain filamentous lignin-degrading fungi, especially white rot basidiomycetes such as *Phanerochaete chrysosporium*, whose genome was recently sequenced. It is known that lignin degradation is oxidative, but the specific mechanisms remain obscure—numerous enzymes and cofactors have been proposed to have a ligninolytic role. To address this problem, it would be useful to visualize fungal ligninolytic oxidants on biodegrading lignocellulose, to map their position relative to the fungus, and to correlate their production with the expression of fungal genes.

We have developed micrometer-scale silica beads that carry an oxidant-sensitive BODIPY dye, which fluoresces red in its native reduced state but green when oxidized. The beads were placed on thin sections of wood undergoing early decay by *P. chrysosporium* and the specimens were analyzed periodically by fluorescence microscopy. Concurrently, specimens were harvested for RNA extraction and transcript analysis by quantitative RT-PCR, using primers designed to amplify sequences encoded in the genome that have a proposed ligninolytic role. Three major findings have emerged from our

study thus far: **(1)** In the first few days of biodegradation, there were marked gradients of oxidation around the fungal hyphae, i.e., beads close to the fungus were more oxidized than were more distant beads. This result is inconsistent with a ligninolytic mechanism that depends on small, freely diffusing oxidants. Instead, it indicates that enzymes associated with the fungal hyphae have a major oxidative role during the initial stage of biodegradation. **(2)** Bead oxidation was not continuous, but rather commenced abruptly several days after fungal inoculation onto the wood. This result indicates a rapid transition from nonligninolytic to ligninolytic metabolism by the fungus. **(3)** Concurrent with this oxidative burst, several fungal genes were highly up-regulated. Chief among these were several peroxidases that have been shown to cleave lignin *in vitro*. By contrast, a cellobiose dehydrogenase and two glycopeptides that have been proposed to produce ligninolytic hydroxyl radicals were not up-regulated. These results support a mechanism for incipient ligninolysis in which fungal peroxidases react directly with lignin to cleave it.



Fluorescence images of beads on spruce wood undergoing decay by *P. chrysosporium* at 40 h (left) and 96 h (right) after inoculation.

This work was supported by grant DE-FG02-11ER65271 from the Office of Science (BER), U.S. Department of Energy.

87

Multi-Species, Multi-Gene Co-Regulation: Finding DNA *Cis*-Regulatory Elements for Biofuel Pathways

Lee Ann McCue,¹ Lee A. Newberg,^{2,3*} (lee.newberg@wadsworth.org), William A. Thompson,⁴ and Charles E. Lawrence⁴

¹Pacific Northwest National Laboratory, Richland, Wash.; ²Wadsworth Center, New York State Department of Health, Albany, N.Y.; ³Rensselaer Polytechnic Institute, Troy, N.Y.; and ⁴Brown University, Providence, R.I.

Project Goals: Decreasing America's dependence on foreign energy sources and reducing the emission of greenhouse gases are important national priorities. We are undertaking research into the metabolic and regulatory networks responsible for biohydrogen and bioethanol production. In particular, among the hundreds of alpha-proteobacterial species with sequenced genomes are several species with metabolic capabilities of interest. The first long term goal of this research is to identify the ensemble of solutions that have been explored by the

alpha-proteobacteria to regulate the metabolic processes key to biofuel production. The second long term goal of this project is to develop probabilistic models to represent these multiscale processes, through Bayesian statistical inference procedures and computational methods to identify the posterior distributions of these parameters, efficient point estimates of their values, and Bayesian confidence limits for these estimates.

This poster's message: Multiple sequence alignments often contain considerable uncertainty; rather than using only one alignment of multiple nucleotide sequences or leaving the sequences unaligned in searches for *cis*-regulatory elements, we are using a Bayesian approach that searches through the joint parameter space of *cis*-regulatory elements and multiple sequence alignments. This permits information about plausible multiple sequence alignments to aid in the locating of elements and also allows the plausible locations of elements to aid in the determination of the relative quality of multiple sequence alignments. This combination enables us to better locate *cis*-regulatory elements.

The details: Key to understanding gene regulation in the biofuel pathways is the locating of the *cis*-regulatory elements that are recognized by regulatory transcription factors. Each regulatory factor energetically prefers a short sequence of nucleotides, typically 10-30 nucleotides in bacteria. There is much tolerance for variation in the recognized sequence and thus the computational search for a regulator's elements is a search for a set of similar short sequences that are present significantly more often than would be expected by chance. The search is through the genomic regions near genes that are likely to be co-regulated and across species that are likely to exhibit similar pathways of interest.

Quantification of the significance of abundance and similarity among *cis*-regulatory elements depends upon the hypothesized evolutionary relationships of the elements. For example, when input sequences are nearly identical the similarity of putative *cis*-regulatory elements is a less compelling indication of functional conservation than when the same amount of similarity is present in more distantly related sequences or in evolutionarily unrelated sequences; in the latter cases it is easier to conclude that the similarity is due to selective pressures associated with functionality rather than mere evolutionary nearness. Almost always, multiple sequence alignment (or its variant, tree sequence alignment) is the way that the orthologous relationships among nucleotides are specified. However, alignments and regulatory elements are not independent; a supposition of an alignment affects the apparent quality of a set of proposed elements and, vice-versa. As a result, a motif finding algorithm that searches for *cis*-regulatory elements using a multiple alignment built using a tool that is unaware of elements, and those that search in unaligned sequences both make sacrifices in the trade-off between sensitivity and specificity.

With a Markov chain Monte Carlo, Bayesian approach, we are simultaneously searching the space of possible sequence alignments and possible regulatory elements to find combinations that work well together to enhance the identification of unreported *cis*-regulatory elements.

Here we present preliminary results regarding the effectiveness of this Bayesian approach as applied to biofuel pathways in alpha-proteobacteria.

This research is supported by the Office of Biological and Environmental Research in the DOE Office of Science under a multi-institutional grant entitled "Bayesian computational approaches for gene regulation studies of bioethanol and biohydrogen production." (Grant No. DE-FG02-09ER64757 to C.E.L., DE-FG02-09ER64756 to L.A.N., and Grant No. FWP 55426 to L.A.M.)

88 Early Career Program Optimizing Plant-Microbial Systems for Energy—Mapping Feedstock Quality Genes in *Brachypodium distachyon*

Scott J. Lee and **Samuel P. Hazen*** (hazen@bio.umass.edu)

Biology Department, University of Massachusetts, Amherst

Project Goals: (1) Quantify relevant genetic diversity for feedstock quality and plant by microbial strain interaction effects. We will phenotype a large germplasm collection of the model grass species *Brachypodium distachyon* to determine the overall genetic diversity for the rate of ethanol production using a range of *Clostridium phytofermentans* strains. (2) Utilize advanced quantitative trait loci (QTL) mapping strategies to identify genetic loci that control biofuel feedstock quality. Genetic linkage mapping will be carried out in immortalized recombinant inbred lines to provide awareness of the number of QTL that influence the traits, and their degree of influence. (3) Identify and evaluate candidate genes for improved energy systems using a multi-tiered approach to tag several QTL. Ultimately, we are interested in identifying the genes that have an effect on plant feedstock quality.

Creation of a sustainable biofuels industry depends on new technologies that are able to release the energy stored in cellulose fibers at a cost that enables widespread use of biofuels as a competitive domestic alternative to fossil fuels. Currently, use of biomass is hampered by the economic costs associated with thermochemical treatment of biomass, enzyme production, and biomass saccharification. This limits not only the use of cellulosic ethanol, but also the advancement of third generation biofuels. The purpose of our work is to develop plant-microbe systems that will reduce the expense of these steps. This will be accomplished in part by elucidating the genetic mechanisms underlying the attributes of plants that lead to facile decomposition with an eye on the interactions between specific feedstock genotypes and microbial systems. *Clostridium phytofermentans* can directly produce ethanol from plant biomass. Using a high throughput format, we quantify ethanol of inoculated biomass as an estimate of plant feedstock quality for our genetic studies [1, 2]. To better understand genetic diversity for feedstock quality, we will concentrate effort on a genetically diverse panel of accessions of the model species *B. dis-*

tachyon. To advance from variance components to genetic mechanisms, we will take advantage of high throughput genotyping and the remarkable genome sequence resources available today to map QTLs. We will begin with the most developed populations and expand to explore others whose parents demonstrate not only striking phenotypic divergence but also possess differential interactions with the microbial strains. This will help bring the genetic landscape into focus, further elucidating the number of genes and their relative influence on plant feedstock-microbe interactions. A key advantage of *B. distachyon* as a model system is the relative ease of creating stable transgenic plants. With that, we will make a series of gain- and loss-of-function mutants to test the function of candidate genes derived from mapping experiments. The three project goals are designed to determine the plausibility of specific positive interactions between plant and microbial genotypes. Pairwise comparisons will be made, varying both plant and microbial genotypes. Similar to adapting crop varieties to different environments, these experiments will link the need for specific feedstock properties to biomass conversion processes. Leveraging the characterization of cellulose-degrading microbes and discovery of plant traits and loci associated with microbial digestibility, we seek to identify the components of a successful plant-microbe system.

References

1. Lee SJ, Warnick TA, Leschine SB, Hazen SP (*In press*) A high-throughput biological conversion assay for determining lignocellulosic quality. In J Normanly, ed, *Methods in Molecular Biology: High Throughput Phenotyping in Plants*. Springer.
2. Lee SJ, Warnick TA, Pattathil S, Alvelo-Maurosa JG, Serapiglia MJ, McCormick H, Brown V, Young NF, Schnell DJ, Smart LB, Hahn MG, Pedersen JF, Leschine SB, Hazen SP (*In review*) Biological conversion assay using *Clostridium phytofermentans* to estimate plant feedstock quality.

This work was supported by the Office of Science, Office of Biological and Environmental Research, of the U.S. Department of Energy grants DE-FG02-08ER64700DE and DE-SC0006641.

89 Early Career Program A Systems Biology, Whole-Genome Association Analysis of the Molecular Regulation of Biomass Growth and Composition in *Populus deltoides*

Leandro Gomide Neves^{1*} (gomide@ufl.edu) and **Matias Kirst**^{1,2}

¹Plant Molecular and Cellular Biology Graduate Program and ²School of Forest Resources and Conservation, University of Florida, Gainesville

Project Goals: Poplars are the principal short rotation woody crop species for providing clean, renewable and sustainable fuels in North America. Poplars are fast growing, have a perennial growth habit and wide natural distri-

bution that supports its implementation in a broad range of environments. While poplars provide the benefits of ideal bioenergy crop, the genes that regulate biomass productivity and composition are largely unknown, despite their critical relevance for efficient conversion to biofuels. Association genetics has become the primary approach for identification of genes that regulate such complex traits because it permits capturing a broad allelic range at high-resolution. Poplars are particularly suited to unveil the molecular basis of biomass productivity and composition using an association genetics approach because of low domestication, large open-pollinated native populations, and high levels of genetic and phenotypic variation. However, with few exceptions, association genetic studies in poplar and other plant species have been hampered by limited gene and polymorphism coverage.

In this project we are genotyping a large genetically unstructured population of *Populus deltoides*, one of the most widely distributed short-rotation woody crop in North America, using a combination of sequence capture and high-throughput genotyping. In parallel, the population is being phenotyped for biomass productivity and quality traits, for future discovery of marker trait associations. The overall goal is to identify the polymorphisms that accounts for differences in biomass productivity and lignocellulose composition in this species, that can be utilized for genetic improvement through transgenics or marker assisted breeding. Our specific project goals and progress towards their completion is described below:

Initially, a sequence-capture platform based on ultra-long RNA oligonucleotides has been developed based on a reference sequence of *P. trichocarpa* and a partial sequence of *P. deltoides*. The oligonucleotide library has been designed to capture sequences from over 22,000 genes. In total we have designed 220,000 probes that target coding as well as regulatory (promoter and UTR) regions of selected genes, and optimized sequence-capture procedures to multiplex 5 genotypes in each hybridization.

The second objective of this project is to sequence-capture target coding and regulatory sequences in a *P. deltoides* association population of 500 individuals. This population is composed of first generation collections made in the south and southeastern U.S. in the last two decades, and is expected to represent the genetic diversity of *P. deltoides* that exists in breeding programs. This work is currently in progress. We have also propagated the populations for trials in greenhouse and field. Greenhouse trials have been completed during the summer of 2011, when biomass data from above and below ground plant compartments were collected, as well as weekly growth data. Two field trials, in Florida and South Carolina, are being installed in the winter of 2011-2012. Initial biomass data collected from the greenhouse trials will be presented.

During the next year we anticipate the completion of sequencing and polymorphisms detection, to begin the analysis of association between markers and traits. In addition, we are exploring the possibility of using the sequence capture data to uncover gene copy number

variation (CNV) in the population of *P. deltoides*. If the presence of variation in copy number becomes evident, we will assess possible associations with trait phenotypes.

90 Next Generation Protein Interactomes for Plant Systems Biology and Biomass Feedstocks Research

Mary Galli, Shelly Wanamaker, and **Joseph R. Ecker***
(ecker@salk.edu)

Salk Institute for Biological Studies, La Jolla, Calif.

Project Goals: Proteins execute most biological processes, but rarely do so alone. Therefore, large-scale efforts to interrogate the composition of the macromolecular complexes in which these proteins operate and their dynamic behavior are essential for a genome-wide understanding of cellular networks. To facilitate the identification of such interactions on a proteome-wide scale in plants, our project aims to create a high throughput protein-protein interaction system in which interacting pairs can be captured in the form of a nucleic acid readout and identified using next-generation sequencing.

Large-scale protein interaction maps are currently available for several model organisms including yeast, worm, fly, and human. In addition, we have recently reported the first large-scale plant interactome containing ~6200 novel interactions for nearly 3000 *Arabidopsis* proteins. While this effort resulted in the generation of the largest plant interaction dataset to date, we estimate it covers only ~2% of all protein interactions that exist in *Arabidopsis*. This observation is consistent with previous mapping efforts in other model organisms, which also suffered from low coverage. Therefore new technologies with the capability to significantly expand assay coverage are critical to enable comprehensive interactome mapping. This is especially pertinent for plants (in particular non-model plant systems with biomass-production potential) which contain a significantly higher number of protein coding genes relative to other eukaryotes.

We aim to develop an experimental approach that harnesses the unparalleled depth of next-generation sequencing and allows for the efficient and cost-effective interrogation of all pairwise interactions for all proteins encoded in a genome. To accomplish this goal, we have proposed to adapt the yeast 2-hybrid (Y2H) system, currently widely used in both small and large-scale protein interaction studies, to a format that is amenable with a high-throughput sequencing based output. Specifically, our strategy exploits the ability of CRE recombinase from the bacteriophage P1 to catalyze the physical linkage of two pieces of lox-containing DNA in vivo. We have introduced non-revertible lox sites into a set of Gateway-compatible Y2H vectors containing bait and prey ORF-fusion cassettes. Upon transformation into an induc-

ible CRE-containing yeast strain, an interaction between bait and prey fusion proteins results in the reconstitution of a functional yeast GAL4 transcription factor which then drives expression of a growth-based reporter gene while simultaneously driving expression of CRE recombinase. The expression of CRE triggers the physical linkage of the bait and prey plasmids that encode the interacting proteins. Because the interacting pairs are locked together, large numbers of colonies that survive selection can be pooled *en masse* and processed as a single sample using next-generation sequencing technologies to identify the interacting pairs. A simple shearing strategy of the pooled, fused interacting DNAs followed by ligation with Illumina adaptors simplifies the sequencing library preparation process. Our current Illumina HiSeq platform yields over 600 Gigabases of DNA sequencing output per run and thus provides sufficient coverage for the identification of millions of interaction pairs.

Following initial proof of concept experiments using a preexisting set of *Arabidopsis* positive and random reference set clones, we plan to re-query a subset of the *Arabidopsis* Interactome pairwise space. This will allow us to obtain a comparative estimate of the increase in assay coverage and quality of our next-generation interactome approach. Ultimately, such a system has the potential to move beyond traditional time-consuming pairwise tests using well-defined ORF collections and accommodate library vs. library screens using 'cDNA shatter libraries'. Flexible formats such as this will pave the way for the generation of interactome maps in organisms, such as potential biomass fuel stocks, for which no pre-existing ORF collections are available.

We gratefully acknowledge the U.S. DOE Office of Biological and Environmental Research in the U.S. DOE Office of Science for funding this project; DOE-DE-SC0007078.

91 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 Park¹

¹North Carolina State University, Raleigh and ²University 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 ¹³C-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 Coreregulated 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 coreregulated 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.

92

Regulation of Neutral Lipid Accumulation in Vegetative Tissues of Plants

Kent D. Chapman^{1*} (chapman@unt.edu), John M. Dyer,² and Robert T. Mullen³

¹Center for Plant Lipid Research, University of North Texas, Denton; ²USDA-ARS-ALARC, Maricopa, Ariz.; and ³University of Guelph, Ontario, Canada

http://www.biol.unt.edu/~chapman/investigators/chapman/Projects/projects_frames/projectframeset.html (select Lipid Droplet)

Project Goals: Our research project aims to examine the cellular machinery that influences the accumulation of lipid droplets in non-seed tissues of plants. Our initial efforts are focused on a homolog of a mammalian hydrolase, *CGI-58*, that in humans is the causative gene for a

neutral lipid storage disease. A loss-of-function mutation in this gene in *Arabidopsis* results in accumulation of triacylglycerols (TAGs) in leaf tissues that normally do not store neutral lipids. We proposed to 1) examine the role of *AtCGI58* gene products in lipid droplet formation in *Arabidopsis* vegetative tissues 2) characterize the biochemical and cellular properties of *CGI58* in the context of lipid metabolism. An improved understanding of the factors that regulate the accumulation of neutral lipids in vegetative tissues of plants may help to identify new strategies to enhance the energy content of crop plants.

Confocal laser scanning fluorescence imaging experiments indicated that *Arabidopsis cgi-58* loss-of-function mutants had significantly more lipid droplets in leaf mesophyll cells compared with wild-type cells. These lipid droplets were determined to be cytosolic and isolated lipid droplets were shown to contain triacylglycerols (TAG) with leaf-type fatty acids. While the numbers of lipid droplets were somewhat variable in leaf tissues of mature plants, there was a ten-fold increase in TAG levels in *cgi-58* mutants compared with wild-type plants. This lipid droplet/TAG phenotype was similar to that reported for peroxisomal mutants deficient in the uptake or catabolism of fatty acids (e.g., acylCoA oxidase, *acx1*, *acx2*; FA transporter, *cts1*, *pxa1*). There were no statistical differences in TAG levels of *cgi-58* seeds suggesting that the CGI-58 protein may act to influence lipid droplet abundance in non-seed tissues only, unlike the known peroxisomal mutants that confer a sugar-dependent germination phenotype. TAGs of *cgi-58* mutants were specifically enriched in molecular species that contained 18:3 (linolenic acid) fatty acid moieties. The enrichment in 18:3-containing lipids also was observed in major membrane galactolipids and phospholipids, suggesting that CGI-58 may participate specifically in the turnover of 18:3-containing glycerolipid species.

In animal systems, the CGI-58 protein is known to activate triacylglycerol lipases, and to transiently interact with the lipid droplet surface to mediate its functions. While the details of lipase activation in animal cells are not clear, we asked if the *AtCGI-58* protein interacts with partner proteins to accomplish its function in lipid droplet biogenesis in plants. We implemented an unbiased approach to screen for interacting partners of CGI-58 using a yeast two-hybrid system to screen an *Arabidopsis* cDNA library. Several candidate interacting proteins were identified and we are currently following up one in detail—*At4g39850* (*PXA1*, encoding peroxisomal fatty acid transporter 1, previously named *comatose*, *cts*, above). The interaction of CGI-58 with *PXA1* might not be entirely unexpected, since the phenotype of the loss of *pxa1* (*cts*) function mutant shows an increase in cytosolic lipid droplets in leaves, similar to the *cgi-58* knockouts. Our interaction assays revealed that only the C-terminus of *PXA1* interacts with CGI-58, specifically the so-called Walker B motif of nucleotide binding domain 2 (NBD2). This domain has been hypothesized by others to be involved in the regulation of the *PXA1* transporter activity. Interaction between CGI-58 and *PXA1* *in planta* was demonstrated through nuclear re-localization assays. In these experiments, the so-called Walker B motif of NBD2

(but not a mutated form of the Walker B motif, the Walker A motif nor the NBD1 domain of PXA1) containing a nuclear localization sequence (NLS) was able to redirect AtCGI-58-GFP protein to the nucleus when these two proteins were co-expressed in tobacco BY-2 cells, indicating that the Walker B motif of the PXA1 NBD2 domain is sufficient for CGI-58 interaction and that this interaction indeed occurs in the plant cytosolic environment. Additional transient expression assays in leaves with prolonged incubation have shown that in addition to a cytosolic location, CGI-58-GFP co-localizes to peroxisomes, consistent with a PXA1 interaction *in vivo*.

To test the functional association of CGI-58 and peroxisomal PXA1 beyond TAG accumulation, we turned to another “reporter” of PXA1 function, which is the uptake of 12-*oxo*-phytodienoic acid (oPDA) for its conversion to jasmonic acid. Indeed, whole seedlings or mature leaves of *cgi-58* knockout plants, when wounded, produced significantly less jasmonic acid compared with wild type plants. The reduction in wound-inducible jasmonate was to about 60% of wild-type levels, and there were no significant differences in jasmonate levels in unwounded tissues, suggesting that plants may have mechanisms in place to generate jasmonate independent of CGI-58. However, CGI-58 clearly influences the efficiency of jasmonate formation, likely through facilitating the uptake of oPDA by PXA1. These results corroborate our protein-protein interaction findings, and broaden the importance of CGI-58 function in plants to include modulation of signaling lipid metabolism in addition to regulating neutral lipid accumulation.

References (supported by this project)

1. James CN, Horn PJ, Case CR, Gidda SK, Zhang D, Mullen RT, Dyer JM, Anderson RG, Chapman KD (2010) Disruption of the *Arabidopsis* CGI-58 homologue produces Chanarin-Dorfman-like lipid droplet accumulation in plants. *Proceeding of the National Academy of Sciences (USA)* 107(41):17833-8.
2. Horn PJ, Ledbetter NR, James CN, Hoffman WD, Case CR, Verbeck GF, Chapman KD (2011) Visualization of lipid droplet composition by direct organelle mass spectrometry. *Journal of Biological Chemistry*, 286(5):3298-306.
3. Chapman KD, Ohlrogge JB (2011) Compartmentation of triacylglycerol accumulation in plants. *Journal Biological Chemistry*- Minireview- Thematic series Nov 16. [Epub ahead of print] PMID:22090025.
4. Chapman KD, Dyer JM, Mullen RT (2011) Biogenesis and functions of lipid droplets in plants. *Journal of Lipid Research*- Thematic Review Series Nov 1. [Epub ahead of print] PMID:22045929.
5. Horn PJ, Chapman KD (2011) Lipidomics in tissues, cells, and subcellular compartments. *Plant Journal* Nov 24. doi: 10.1111/j.1365-313X.2011.04868.x. [Epub ahead of print]. PMID: 22117762. Special Issue on High-Resolution Analysis of Plants.
6. Ohlrogge JB, Chapman KD (2011) The seeds of green energy—expanding the contribution of plant oils as biofuels. *The Biochemist* 33:34-38.
7. Dyer J, Mullen R, Chapman K (2012) Oil in biomass: a step-change for bioenergy production? *Inform (in press, scheduled for April 2012)*.

Supported by the Office of Biological and Environmental Research in the DOE Office of Science (DE-SC0000797).

93

Investigations of Clusters of Cellobiose-Acid-(H₂O)_n

Madeleine Pincus^{1*} (mpincus@uci.edu) and Robert B. Gerber^{1,2}

¹Department of Chemistry, University of Irvine, Irvine, Calif. and ²The Fritz Haber Research Center, The Hebrew University, Jerusalem, Israel

The interactions between cellobiose and acids of different strengths, from very strong (H₂SO₄) to very weak (HCOOH), are studied in clusters of the type cellobiose-acid-(H₂O)_n. Ab initio molecular dynamics (DFT / BLYP with dispersion) is used to analyze the interaction between cellobiose and the micro-hydrated acids, in the gas phase, at room temperature. We compare these in terms of the probability for proton transfer (determined from Mulliken charges), to the active sugar site, as well as in terms of complete acid ionization. Initial findings show that in the case of cellobiose-HCl-(H₂O)_n, the sugar does not facilitate complete acid ionization as compared with the acid in pure water. The sugar, however does affect the pH of the acid. Detailed information on the preferred protonation site of cellobiose and the competition between the sugar and water protonation are presented. Differences in the ionization behaviour of different acids in complex with the sugar are discussed.

94

Production and NMR Analysis of Deuterated Cellulose and Lignocellulosic Biomass and Its Utilization for Neutron Scattering Studies

Barbara R. Evans^{1*} (evansb@ornl.gov), Marcus Foston,² Art Ragauskas,² Hugh M. O'Neill,¹ Juhong He,¹ Sai Venkatesh Pingali,¹ Volker Urban,¹ Paul A. Langan,¹ and Brian H. Davison¹

¹Oak Ridge National Laboratory, Oak Ridge, Tenn. and ²Institute of Paper Science and Technology, Georgia Institute of Technology, Atlanta

<http://www.ornl.gov>

Project Goals: Lignocellulosic biomass from plant cell walls has the potential to play a major role in generation of renewable biofuels if cost-effective conversion can be achieved. However, it is a complex composite material that shows significant recalcitrance towards the structural deconstruction necessary for production of bioethanol. A fundamental understanding of the structural changes and

associations that occur between its component polymers at the molecular level during deconstruction is essential for enabling cost-effective lignocellulose-based fuels production. This Scientific Focus Area in Biofuels seeks to develop and demonstrate the “Dynamic Visualization of Lignocellulose Degradation by Integration of Neutron Scattering Imaging and Computer Simulation” to provide this fundamental information about the structure and deconstruction of plant cell walls by integration of neutron science and computational simulation with physical and chemical characterization by NMR techniques.

In support of neutron science experiments, NMR methods were developed and demonstrated for analysis of deuterium incorporation in cellulose and plant biomass produced by culture of target species under deuterating conditions. The methods provide crucial supporting information for visualization of lignocellulose structure by neutron techniques and computational simulation.

Small angle neutron scattering (SANS) provides useful tools for investigation of lignocellulosic structural complexity and its changes during pretreatment and hydrolysis to sugars. The use of deuterated materials in which the non-exchangeable hydrogen atoms are replaced with deuterium can greatly increase the value of these studies, as the scattering patterns of the different components can be separated by phase contrast, enabling simultaneous observations of each component. Partially deuterated biomass samples were produced by cultivation of duckweed (*Lemna minor*), and the grasses annual rye (*Secale cereale*) and switchgrass (*Panicum virgatum*) in D₂O/H₂O mixtures. Samples of cellulose with specific levels of deuterium incorporation were obtained by cultivating *Acetobacter xylinus* in a defined deuterium medium. Using these samples, NMR methods were developed for quantifying the degree of deuterium substitution in the complete and individual components of the cell wall. A second NMR methodology that requires no sample preparation and is non-destructive was specifically designed and demonstrated to analyze deuterium incorporation in whole biomass samples. It was possible to measure the ratio of the integrals resulting from the subsequent ¹H and ²H spectra. These values were correlated with known molar ratios of protons to deuterons from a calibration curve that were generated by taking several calibration standards formed from mixtures of glucose/glucose-d₇. The first SANS experiments have been carried out using the deuterated cellulose and plant biomass materials. Combined with ¹³C-NMR, X-ray diffraction (XRD), and gel permeation chromatography (GPC), the new NMR techniques provide crucial chemical and structural information needed to leverage the full potential of SANS and building of real-time computational models of biomass deconstruction.

The Scientific Focus Area Biofuels is supported under FWP ERKP752 by the Office of Biological and Environmental Research in the U. S. Department of Energy Office of Science.

95

In situ Small-Angle Neutron Scattering and Computer Simulation Investigate Lignin Aggregation During Biomass Pretreatment

Loukas Petridis^{1*} (petridisl@ornl.gov), Sai Venkatesh Pingali,¹ Volker Urban,¹ William T. Heller,¹ Hugh M. O’Neill,¹ Marcus Foston,² Art Ragauskas,² Barbara R. Evans,¹ Paul A. Langan,¹ Jeremy C. Smith,¹ and **Brian H. Davison¹**

¹Oak Ridge National Laboratory, Oak Ridge, Tenn. and ²Institute of Paper Science and Technology, Georgia Institute of Technology, Atlanta

Program: SFA Biofuels

Project Goals: Lignocellulosic biomass comprises the vast majority of biomass on Earth and has the potential to play a major role in generation of renewable biofuels if cost-effective conversion can be achieved. Largely composed of plant cell walls, lignocellulosic biomass is a complex biological composite material that shows significant recalcitrance towards the structural deconstruction and enzymatic hydrolysis into sugars that is necessary for fermentation to bioethanol. This Scientific Focus Area in Biofuels seeks to develop and demonstrate the “Dynamic Visualization of Lignocellulose Degradation by Integration of Neutron Scattering Imaging and Computer Simulation” for multiple-length scale, real-time imaging of biomass during pretreatment and enzymatic hydrolysis. This will provide fundamental information about the structure and deconstruction of plant cell walls that is needed to drive improvements in the conversion of renewable lignocellulosic biomass to biofuels.

***In situ* small-angle neutron scattering (SANS) and extensive molecular dynamics (MD) computer simulation were used in a combined approach to examine real-time breakdown of biomass and the temperature dependence of lignin structure and dynamics.**

Lignin, a major polymeric component of plant cell walls, forms aggregates in vivo during pretreatment of lignocellulosic biomass for ethanol production. The aggregates are thought to reduce ethanol yields by inhibiting enzymatic hydrolysis of cellulose. Here, we report on real-time SANS experiments during dilute acid pretreatment of switchgrass and poplar biomass using a temperature-pressure reaction cell we developed. The temperature of the cell was ramped-up from 20 to 180 °C, then maintained at 180 °C for 60 min and finally the cell was cooled down to room temperature. A very clear characteristic structural feature appeared in the SANS data for $Q < 0.06 \text{ \AA}^{-1}$ between 120–130 °C. This structural feature progressively moves to smaller- Q with increasing temperature and residence time at 180 °C, indicating a growth in the particle size. We have identified the particle that appears at 120 °C as being lignin aggregates. To understand the temperature-dependent structure and dynamics of individual lignin polymers in aqueous solution we performed extensive (17 μ s) MD simulations. Between

150 °C and 210 °C, the lignin transitions from a glassy, compact state to a mobile, extended state. In all cases, the polymers were found to be globular particles, inside of which the R_g of a polymer segment is a power-law function of the number of monomers comprising it. In the low temperature states studied, the blobs are inter-permeable. In contrast, the particles at high temperatures become spatially separated, leading to a fractal crumpled globule form. Our simulations showed that the low-temperature collapse is thermodynamically driven by the increase of the translational entropy and density fluctuations of water molecules removed from the hydration shell. This combination of effects distinguishes lignin collapse from enthalpically driven coil-globule polymer transitions and provides a thermodynamic role for hydration water density fluctuations in driving hydrophobic polymer collapse. The detailed characterization obtained here provides insight at atomic detail into processes relevant to biomass pretreatment for cellulosic ethanol production. Specifically, our combined approach indicates that lignin aggregation occurs rapidly during heating—not during the cool-down phase, as was previously assumed.

96

Study of Plant Cell Wall Polymers Affected by Metal Accumulation Using Stimulated Raman Scattering Microscopy

Shi-You Ding^{1*} (Co-PI: Shi.you.Ding@nrel.gov), and **X. Sunney Xie**² (PI: xie@chemistry.harvard.edu)

¹Biosciences Center, BioEnergy Science Center, National Renewable Energy Laboratory, Golden, Colo. and

²Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Mass.

Project Goals: This project employs newly-developed chemical imaging techniques to measure, in real-time, the concentration, dynamics and spatial distribution of plant cell wall polymers during biomass growth with inoculation of transgenic symbiotic fungi. The proposed new pathway of biomass production will: 1) benefit metal and radionuclide contaminant mobility in subsurface environments, and 2) potentially improve biomass production and process for bioenergy. The goal of this project will explore a new pathway of delivering detoxified metal to plant apoplast using transgenic symbiotic fungi, which will enhance metal accumulation from soil, and potentially these metals may in turn be used as catalysts to improve the efficiency of biomass conversion to biofuels. We further develop chemical imaging tools to quantitatively analyze the key plant cell wall polymers such as cellulose and lignin.

In our previous project funded by BER lignocellulose imaging program, we have demonstrated that stimulated Raman scattering (SRS) microscopy, a new imaging method, which allows real time observation of biomass conversion processes (1-4). SRS microscopy offers chemical contrast based on the intrinsic Raman vibrational frequencies in a

sample with much shorter imaging time and easier spectral identification. The SRS imaging technique for studying the conversion process *in situ* offers chemical specificity without exogenous labels, non-invasiveness, high spatial resolution, and real-time monitoring capability. We propose to further the recently emerging SRS techniques to imaging the distribution of plant cell wall polymers (e.g., cellulose and lignin) and cell wall architecture with different growth conditions including in heavy metal added soils and inoculated with symbiotic fungi. Specifically, we propose to:

- Establish a model system of plant-symbiotic fungi that produce metal-bound biomass by inoculating the poplar plants with transgenic fungus *Trichoderma harzianum* that carries secretory metal-binding protein.
- Evaluate fungal expression system growing in plant, and the efficiency of inoculation of fungi in intact plant cell walls. Specifically, we will examine if a closer contact between metals and cellulosic and lignin polymers caused by the entrapment of metals within plant cell, will alter plant cell wall structure and lead to a higher efficiency for cell wall degradation.
- Develop imaging tools to quantitatively analyze the distribution and deconstruction kinetics of plant cell wall polymers.

References

1. Freudiger, Christian W.; Min, Wei; Saar, Brian G.; Lu, Sijia; Holtom, Gary R.; He, Chengwei; Tsai, Jason C.; Kang, Jing X.; Xie, X. Sunney "Label-Free Biomedical Imaging with High Sensitivity by Stimulated Raman Scattering Microscopy," *Science*, 322, 1857-1861 (2008).
2. Saar, Brian G.; Freudiger, Christian W.; Reichman, Jay; Stanley, C. Michael; Holtom, Gary R.; Xie, X. Sunney "Video-Rate Molecular Imaging in Vivo with Stimulated Raman Scattering" *Science*, 330, 1368-1370 (2010).
3. Zeng, Yining; Saar, Brian G.; Friedrich, Marcel G.; Chen, Fang; Liu, Yu-San; Dixon, Richard A.; Himmel, Michael E.; Xie, X. Sunney; Ding, Shi-You "Imaging Lignin-Downregulated Alfalfa Using Coherent Anti-Stokes Raman Scattering Microscopy" *Bioenerg. Res.* (2010).
4. Saar, Brian G.; Zeng, Yining; Freudiger, Christian W.; Liu, Yu-San; Himmel, Michael E.; Xie, X. Sunney; Ding, Shi-You "Label-Free, Real-Time Monitoring of Biomass Processing with Stimulated Raman Scattering Microscopy" *Angew. Chem. Int. Ed.* 49, 5476-5479 (2010).

This work is funded by the U.S. Department of Energy, the Office of Biological and Environmental Research (BER).

97

Defining Determinants and Dynamics of Cellulose Microfibril Biosynthesis, Assembly, and Degradation

Jocelyn Rose,^{1*} (jr286@cornell.edu) Thomas Brutnell,² Seth DeBolt,³ Detlef Smilgies,¹ and Larry Walker¹

¹Cornell University, Ithaca, New York; ²Donald Danforth Plant Science Center, St. Louis, Missouri; and ³University of Kentucky

Project Goals: The goals of this multidisciplinary project are to: (1) Establish platforms through reverse and forward genetics to identify and manipulate candidate genes that influence cellulose microfibril synthesis and structure; (2) Characterize the effects of altered candidate gene expression on cellulose microfibril synthesis and structure, and develop a mechanistic model for microfibril crystallization; (3) Determine the consequences of altering microfibril architecture on digestibility and integrate this information with nano-scale observations of enzymatic hydrolysis.

The central paradigm for converting plant biomass into soluble sugars for subsequent conversion to transportation fuels involves the enzymatic depolymerization of lignocellulosic plant cell walls by microbial enzymes. Despite decades of intensive research, this is still a relatively inefficient process, due largely to the recalcitrance and enormous complexity of the substrate. A major obstacle is still insufficient understanding of the detailed structure and biosynthesis of major wall components, including cellulose. For example, although cellulose is generally depicted as rigid, insoluble, uniformly crystalline microfibrils that are resistant to enzymatic degradation, the *in vivo* structures of plant cellulose microfibrils are surprisingly complex. Crystallinity is frequently disrupted, for example by dislocations and areas containing chain ends, resulting in “amorphous” disordered regions. Importantly, microfibril structure and the relative proportions of crystalline and non-crystalline disordered surface regions vary substantially and yet the molecular mechanisms by which plants regulate microfibril crystallinity, and other aspects of microfibril architecture, are still entirely unknown. This obviously has a profound effect on susceptibility to enzymatic hydrolysis and so this is a critical area of research in order to characterize and optimize cellulosic biomass degradation.

The entire field of cell wall assembly, as distinct from polysaccharide biosynthesis, and the degree to which they are coupled, are relatively unexplored, despite the great potential for major advances in addressing the hurdle of biomass recalcitrance. Our overarching hypothesis is that identification of the molecular machinery that determine microfibril polymerization, deposition and structure will allow the design of more effective degradative systems, and the generation of cellulosic materials with enhanced and predictable bioconversion characteristics.

We believe that the most effective way to address this long standing and highly complex question is to adopt a broad ‘systems approach’. Accordingly, we have assembled a multidisciplinary collaborative team with collective expertise in plant biology and molecular genetics, polymer structure and chemistry, enzyme biochemistry and biochemical engineering. Our team will use a spectrum of cutting edge technologies, including plant functional genomics, chemical genetics, live cell imaging, advanced microscopy, high energy X-ray spectroscopy and nanotechnology, to study the molecular determinants of cellulose microfibril structure. Importantly, this research effort will be closely coupled with an analytical pipeline to characterize the effects of altering microfibril architecture on bioconversion potential, with the goal of generating predictive models to help guide the identification, development and implementation of new feedstocks. This project therefore spans core basic science and applied research, in line with the goals of this program.

This project is supported by the Office of Biological and Environmental Research in the DOE Office of Science.

98

Reduced Lignification Enhances Cellulose Accessibility in Wood Cell Walls

Catherine I. Lacayo,¹ Shi-You Ding,² Mona Hwang,¹ and Michael P. Thelen^{1*} (mthelen@llnl.gov)

¹Lawrence Livermore National Laboratory, Livermore, Calif. and ²National Renewable Energy Laboratory, Golden, Colo.

Project Goals: Effective deconstruction of plant cell wall lignocellulose is crucial to producing economical biofuels. The objective in this section of our SFA project (J. Pett-Ridge, Lead PI) is to dynamically monitor and quantitate the changes in lignin and polysaccharide content during various treatments of wood cells. We have focused in particular on investigating ways to reduce lignin content in walls, and how this influences enzyme accessibility to cellulose.

Most studies on cellulase enzymes use purified model substrates, usually either insoluble, crystalline cellulose, or soluble, chemically modified cellulose. While these substrates are good for assays such as measuring the release of glucose, they have dubious predictive value for the efficiency of chemical and enzymatic processing of plant feedstocks during biofuel production. Also, milled feedstock material such as switchgrass contain a heterogeneous mixture of cell types in different physical states, and therefore are difficult to use for statistical measurements in cell populations. A biologically relevant substrate for analyzing lignocellulose deconstruction is provided by cultured wood cells, which can be examined individually and in populations for the removal of lignin, cellulose, and hemicellulose. Toward this end, we established the *Zinnia* culture system for generating homogeneous suspensions of fully differentiated xylem cells,

which are characterized by distinctly organized features found in all wood cell walls and are easily recognized by brightfield, fluorescence, and atomic force microscopy techniques.

Using this system, we characterized the topology of primary and secondary cell walls at nanoscale resolution, providing us with an architectural model of the *Zinnia* xylem cell (Lacayo et al, 2010, *Plant Physiol*). Lignin autofluorescence was diminished by about 90% in wood cells incubated for 6 hours in acidified sodium chlorite at 70 °C. By observing these oxidized cells in an environmental slide chamber under polarized light, we could track the intrinsic cellulose birefringence in secondary walls in real time during incubation with an endoglucanase. The results indicated that most of the cellulose was removed over a 3 hour period. This semi-quantitative analysis was confirmed by similar experiments in which fluorescence intensity was measured in cells probed with a GFP linked carbohydrate binding protein, *CtCBM3*, which is specific for crystalline cellulose. When cellulose was enzymatically digested during the 3 hour incubation, GFP-*CtCBM3* binding decreased, leading to a statistically significant loss of fluorescence when compared to a control lacking cellulase. These studies provided the foundation for further experiments designed to determine the accessibility of cellulase and xylanase enzymes in secondary walls.

The amount of lignin in wood cells was also decreased during xylem development by the use of specific inhibitors of lignin biosynthesis. Culture medium was amended with reduced glutathione (GSH) or diphenyleneiodonium (DPI), and secondary wall differentiation proceeded normally except that autofluorescence and phloroglucinol staining were significantly lower than in control cells. After the treated cells were incubated with cellulase, GFP-*CtCBM3* fluorescence measurements indicated a statistically significant correlation of cellulose removal with a decrease in secondary wall lignin. Although the inhibition of lignin synthesis during differentiation was not as pronounced as removing lignin post-differentiation by chemical oxidation, both resulted in increased secondary cell wall accessibility to cellulase. Our studies confirm observations of enhanced saccharification in cellulase treated *Arabidopsis* lignin biosynthetic mutants, and provide a quantitative approach to assessing enzymatic efficiencies in a different model system.

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

99

Live Cell Identification of Redox Regulated Dithiol Sensors by Chemical Probes

Aaron T. Wright^{1*} (aaron.wright@pnnl.gov), Natalie C. Sadler,¹ Na Fu,¹ Eric A. Hill,¹ Matthew R. Melnicki,¹ Thomas C. Squier,¹ Allan E. Konopka,¹ Diana J. Bigelow,¹ Erika M. Zink,¹ Michael P. Lewis,¹ Suchitra Datta,¹ Dian

Su,¹ Wei-Jun Qian,¹ Don A. Bryant,² and Alexander S. Beliaev¹

¹Biological Sciences Division, Pacific Northwest National Laboratory, Richland, Wash. and ²Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park

Project Goals: The long-term goal of the PNNL BSFA is to develop a predictive understanding of pathways and regulatory schemes involved in solar energy conversion to biofuel precursors or products. To date, knowledge gaps are not limited to biochemical pathways of energy and carbon flux but to an even larger degree include regulatory events within and between metabolic subsystems that include intracellular signals to which transcriptional regulation is responding. We suggest that “redox sensing,” as a means to maintain redox homeostasis in photosynthesizing cells, is likely an equally important mechanism. In this project, we are investigating the central hypothesis that these sensors regulate electron flux and carbon partitioning in cyanobacteria. Specifically, we have developed and applied probe approaches for live cell capture, characterization, and imaging of redox sensors of environmental change using advanced separations, mass spectrometry and traditional biochemical techniques. It is anticipated that the identification of redox-sensitive dithiol linkages and their modulation by thioredoxin (Trx), peroxiredoxin, and other regulators will provide key inputs for understanding the control points of flux distribution.

The primary challenge in identifying redox-regulated dithiol sensors *in vivo* is that cysteine residues are highly reactive and easily oxidized to dithiol linkages following cellular lysis. This eliminates the ability to experimentally measure redox regulation within native physiological settings, and makes requisite the exogenous addition of chemical or biological reductants. Therefore, we developed two different but complementary approaches for synthesizing and employing cell permeable probe reagents that react in live cells (*in situ*) with thiols and permit real-time imaging and mass spectrometric characterization of probe targets. In the first approach we have synthesized click-chemistry enabled chemical probes for fluorescent and mass spectrometric identification of redox regulated dithiol sensors, and in the second approach we used a cell-permeable arsenic probe (TRAP-Cy) that selectively binds to reduced disulfides in close proximity to trap available dithiols in living cells prior to cell lysis. We have used our probes *in situ* to identify proteins in *Synechococcus* sp. PCC 7002 that under-go disulfide exchange in response to changes in cellular conditions.

Click-Chemistry Enabled Probes. Probes were synthesized with three chemical elements: a moiety to impart cell permeability, an iodoacetamide or maleimide group for irreversibly labeling cysteines, and a reporter tag for detection and isolation of probe-labeled proteins. We exploited the multimodal bio-compatible click chemistry (CC) reaction to create “tag-free” probes for profiling proteins in living systems. Probe-labeled proteins were visualized by addition of a complementary azido-tetramethylrhodamine (fluorescence)

or an azido-biotin tag for enrichment and mass spectrometric analysis (LC-MS).

The maleimido and iodoacetamide probes were added simultaneously *in vitro* or *in situ* to *Synechococcus* sp. PCC 7002 grown in a turbidostat under maximal growth rate conditions. Cells were also removed and placed in the dark for two hours and then labeled *in vitro* or *in situ*. Following probe-labeling cells were lysed and probe-labeled proteins attached to biotin via CC. The probe-labeled proteins were then enriched on streptavidin, digested with trypsin, and the peptides analyzed by LC-MS. Critically, in the cells that were labeled post-lysis (*in vitro*) we found no changes in redox regulation of dithiol sensors, demonstrating that lysis rapidly oxidizes biological samples. However, when we analyzed the *in situ* labeled cells we identified 153 redox-regulated proteins that were statistically different between the light and dark conditions. In a follow-up study using a carbon-limited turbidostat we were able to identify redox changes *in situ* within 30 seconds following the addition of CO₂. A time-course study revealed remarkable changes from 30 seconds to 60 minutes post CO₂ addition. Importantly, these changes cannot be measured by traditional transcriptomic or proteomic measurements because these redox events occur within a time-frame that protein and mRNA content has not changed.

Arsenic Probes. TRAP-Cy contains a cyanine dye to facilitate high-throughput gel-based screening prior to affinity purification and MS based identification of reactive dithiols. Using this trapping reagent, we have explored redox-active dithiol reactivity and their relationship to photosynthetic metabolism in *Synechococcus* sp. PCC 7002 under varied environmental conditions. Upon cell lysis, covalently bound cysteines in association with TRAP-Cy are stable, permitting reduction of all remaining thiols and their alkylation. After treatment with excess ethane dithiol, the TRAP-Cy probe is released, permitting enrichment of these cysteines using a thiol-capture affinity resin. In the vast majority of cases homology models of the structures demonstrate that identified thiols are in close proximity in the tertiary structure. Consistent with our hypothesis that dithiol “switches” (e.g., Cys- X_{xx_n} -Cys) activated by thioredoxin-dependent pathways regulate electron flux and carbon sequestration to modulate energy partitioning.

Probe-identified proteins from both probe types map well onto multi-subunit supramolecular complexes involving photosynthetic pathways associated with efficient collection of excitation energy (light harvesting), electron transfer reactions linked to formation of electrochemical gradients, carbon dioxide sequestration (dark reactions), and ATP synthesis. Additional redox-dependent pathways include those involving chaperone activity, transcriptional regulation, and antioxidant proteins linked to protein repair. Together, these results provide quantitative information regarding redox-dependent switches associated with photosynthetic regulation, and provide a systems biology tool capable to providing high-throughput information necessary for predictive metabolic modeling. Finally, both cell permeable probe approaches represent the only existing methods

for identifying and imaging live cell redox regulation, and they will be critical to informing the predictive models of metabolism needed for bioenergy applications.

This research was supported by the Genomic Science Program (GSP), Office of Biological and Environmental Research (OBER), U.S. Department of Energy (DOE), and is a contribution of the Pacific Northwest National Laboratory (PNNL) Biofuels Scientific Focus Area. Mass spectrometry-based proteomics measurements were performed in the William R. Wiley Environmental Molecular Sciences Laboratory, a national scientific user facility sponsored by OBER and located at PNNL.

100 Effects of Nutrient Limitation, Light Quantity and Quality on Growth Dynamics and Physiology of *Synechococcus* sp. PCC 7002

Allan E. Konopka^{1*} (allan.konopka@pnnl.gov), Grigoriy E. Pinchuk,¹ Eric A. Hill,¹ Leo A. Kucek,¹ Matthew R. Melnicki,¹ Trang Vu,² Jennifer L. Reed,² and Alex S. Beliaev¹

¹Pacific Northwest National Laboratory, Richland, Wash. and ²University of Wisconsin–Madison

Project Goals: The PNNL Biofuels Scientific Focus Area (BSFA) conducts fundamental research of microbial photoautotrophs with specific emphasis on photosynthetic energy conversion, reductant partitioning, and central carbon metabolism. As a model system, we utilize unicellular prokaryotic organism *Synechococcus* sp. PCC 7002, which exhibits one of the fastest growth rates known among cyanobacteria and which is also remarkably tolerant to high light intensities. Understanding the genetic and physiological bases of these properties could provide fundamental new insights that are broadly applicable to the optimization of other biological systems for biofuels development. To that end, we are interrogating fluxes through central metabolic pathways to define the major constraining factors (*i.e.* metabolic and regulatory controls) governing carbon partitioning through the metabolic subsystems of cyanobacteria that can be manipulated to increase yields of specific molecules that are either precursors or fuel molecules themselves.

The initial experimental phase, which consists of analyses of photosynthesis and carbon fixation modules, employs turbidostat cultivation technology which allows the organisms to grow at their unrestricted, maximum growth rate. This approach is suited particularly well for photoautotrophs since attenuation of light by self-shading and mass transfer of CO₂ must be taken into account. We are utilizing a custom-designed photobioreactor (PBR) to rigorously control the physiological state of *Synechococcus* 7002, and analyze the properties of both wild type (and subsequently) discrete mutant strains. PBR illumination is provided by light-emitting diodes (LED) generating 680 nm and 630 nm light for the preferential excitation of chlorophyll *a* and phycobilin pigments. The continuous culture system was

modified to control medium input to maintain constant turbidity (turbidostat); this produces a steady state in which growth rate is unrestricted by dissolved nutrient concentrations, but determined by either light irradiance, or cellular capacity (at saturating irradiances). *Synechococcus* 7002 grew when irradiated with only 680 or only 630 nm light, with the latter resulting in higher growth rates; notably, the growth rate differences were greatest at the lowest irradiances. Steady-state dissolved O₂ concentrations were linear functions of irradiance; the slope was 2.3-fold greater when a low amount of 680 nm irradiance was combined with 630 nm light than in cultures grown under 630 nm irradiances alone. When irradiance values were changed, the culture experienced either an increase (shift-up) or decrease (shift-down) in growth rate. After shift-downs, the new steady-state growth rate was usually achieved within 3 h. In contrast, several shift-up transitions required >10 h to reach the new growth rate. The physiological consequences of shifts in light quality and quantity have been investigated using PAM fluorometry to calculate the maximum capacity for photosynthetic electron transport (rETR) and quantum yield of photosystem II (F_v/F_m).

Steady-state chemostat cultures have also been generated and are undergoing analysis of cellular composition, in conjunction with constraint-based modeling and global expression analyses. Significant differences in macromolecular composition have been found among unrestricted growth and light, carbon, or nitrogen-limited cultures and are consistent with previous understanding of macromolecular composition under these conditions. Protein content varied from 33% (N limitation) to 67% (light limitation). Under N limitation, polysaccharide comprised 61% of biomass, but was 10-17% under all other conditions. These measurements will be used to formulate the organism-specific biomass equation and incorporated into developed metabolic model of *Synechococcus* 7002 (*iSyp612*). The initial model was updated with additional genome annotation from pathway/genome databases (PGDBs) and includes 672 genes and 572 reactions. The new information from PGDBs resulted in the modifications of existing gene-protein-reactions associations (GPRs) due to addition of isozymes. A few (~20) new reactions were added to the pathways involved in quinone biosynthesis, carbohydrate, amino acid and nucleotide metabolism. Using the developed model, we will estimate the growth and non-growth associated ATP requirements (GAR, and NGAR, respectively) using the data described above for cultures grown under various light intensities. Model predicted growth, uptake, and secretion rates will be compared with experimental values for cells grown under different chemostat conditions.

This research is supported by the Genomic Science Program (GSP), Office of Biological and Environmental Research (OBER), U.S. Department of Energy (DOE), and is a contribution of the Pacific Northwest National Laboratory (PNNL) Biofuels Scientific Focus Area.