

Systems Biology for DOE Energy Missions: Bioenergy

Bioenergy Research Centers

Great Lakes Bioenergy Research Center (GLBRC)

1

Discovery of Novel Genes Regulating Polysaccharide Biosynthesis and Secretion in Plants

Michael Held^{1,2*} (heldm@msu.edu), Starla Zemelis,^{1,2} Saunia Withers,^{1,3} Curtis Wilkerson,^{1,3} and **Federica Brandizzi**^{1,2,3}

¹Great Lakes Bioenergy Research Center, ²MSU-DOE Plant Research Laboratory, and ³Dept. of Plant Biology, Michigan State University, East Lansing

<http://www.msu.edu/~brandizz/index.html>

Project Goals: The long term goals of our project are to identify novel factors that facilitate the biosynthesis and/or secretion of hemicellulosic and pectic cell wall polymers. These advances will lay groundwork for designing new strategies for improving plant biomass yield and digestibility.

During cotton ovule development there is a rapid burst in the secretion and deposition of non-cellulosic cell wall polymers between 4 and 6 days post-anthesis (dpa). Within this developmental time period, the cotton Golgi swell and produce a large number of secretory vesicles to accommodate the increase in secretion. Differential proteomics of cotton Golgi at 4 and 6 dpa has generated an extensive list of ~800 proteins that increase in abundance and are hypothetically involved in non-cellulosic cell wall biosynthesis and/or secretion. While many of these 6dpa abundant proteins are already known to be involved in polysaccharide biosynthesis and secretion (i.e. nucleotide sugar interconverting enzymes, glycosyltransferases, and transport related proteins), most have been annotated as genes of unknown function. To identify novel proteins which facilitate and/or modulate polysaccharide transport, we have selected a subset of these 6dpa abundant unknown proteins. Currently, we have cloned seventeen *Arabidopsis* orthologs of these cotton candidates as cerulean cyan fluorescent protein fusions; fifteen of which have been localized to endomembranes by transient expression analyses followed by live cell imaging. t-DNA insertion lines have been identified for many of these candidates, and preliminary analyses show several lines having altered cell wall and/or growth phenotypes. This research puts sound foundations toward establishing the nature of factors that not only control cell wall biosynthesis, but also the

traffic of cell wall components through and from the Golgi apparatus.

2

Utilizing Biochemical Adaptations of Plants and Next-Generation Sequencing Instruments to Discover Enzymes and Transcription Factors Involved in Plant Cell Wall Biosynthesis

Jacob Kruger Jensen,^{1,2} Jean-Christophe Cocuron,^{1,2} and Curtis G. Wilkerson^{1,2,3*} (Wilker13@msu.edu)

¹DOE Great Lakes Bioenergy Research Center, ²Dept. of Plant Biology, ³Dept. of Biochemistry and Molecular Biology, Michigan State University, East Lansing

Project Goals: We are attempting to define the proteins required by plants to synthesize arabinoxylan. We are also attempting to discover the transcription factors that regulate the synthesis of these proteins. The ultimate aim is to alter the ratio of hexose to pentose sugars in the plant cell wall to increase the production of biofuels from these feedstocks by fermentation.

One way to discover enzymes and regulatory proteins involved in a biological pathway is to examine the mRNAs of cells actively engaged in the pathway of interest, ideally when the pathway is a major activity for the tissue. There are many such plant tissues that produce very large amounts of product from a single biochemical pathway. As an example, the mucilaginous layer of psyllium (*Plantago ovata* Forsk) seed contains about 60% arabinoxylan by weight (Fischer et al. 2004) and so is a good tissue to use for transcriptional profiling to discover genes involved in arabinoxylan biosynthesis. Similarly, the endosperm tissue of Fenugreek seeds contains 80% galactomannan by weight and hence provides a system to study mannan and galactomannan biosynthesis. The recent development of DNA sequencing instruments that can produce millions of sequences quickly and at moderate cost now allows us to exploit such systems to discover both biosynthetic enzymes from specific pathways and the transcription factors that control their expression. We are interested in using such a strategy to gain a greater understanding of plant cell wall biosynthesis.

Our initial work has been to gain a better understanding of the enzymes required for arabinoxylan biosynthesis using the psyllium system. Arabinoxylan is a major component of grass cell walls and as grasses are likely to be important as bioenergy crops such work could have a major impact on bioenergy research. We have successfully generated 4 cDNA libraries from psyllium mucilaginous tissue at 6, 8, 10 and 12

days post anthesis (DPA) and have obtained over 850,000 DNA sequences using the Roche GS-FLX sequencer. We have developed an analysis pipeline and query software to allow us to cluster, annotate and search these large datasets. An examination of these sequences using our software revealed that enzymes involved in the biosynthesis of UDP-xylose were highly represented in those cDNA libraries suggesting that these libraries likely are enriched in genes involved in arabinoxylan biosynthesis and its regulation. We find that homologues of the putative xylan synthases IRX10 and IRX10-like are very abundant in these tissues. We find at least six genes encoding proteins in the glycosyl transferase family 61 at high abundance. We also find other genes that have been implicated in secondary cell wall biosynthesis that could be involved in xylan biosynthesis as well. One of these genes is related to the *Arabidopsis* gene At3g50220. The expression of this gene is highly correlated with the expression of IRX10 in *Arabidopsis* and so may represent a component of the xylan synthase. Since the psyllium mucilaginous layer is synthesizing almost exclusively arabinoxylan it is likely that the subset of genes present in this tissue compared to the larger set of genes found in cambial tissues defines a minimal set of genes required to synthesize arabinoxylan. We have also identified a subset of transcription factors known to be up regulated during secondary cell wall biosynthesis. Since psyllium is making only arabinoxylan and not a secondary cell wall likely the transcription factors we have identified are involved in the regulation of arabinoxylan biosynthetic enzymes. We are currently expressing the psyllium versions of these genes in various heterologous systems to establish the function of these proteins.

This research was funded by the DOE Great Lakes Bioenergy Research Center (DOE Office of Science BER DE-FC02-07ER64494).

3 Generation and Analysis of Transgenic Poplars with Altered Wall Compositions

Jae-Heung Ko, Won-Chan Kim, Hyun-Tae Kim, and
Kyung-Hwan Han* (hanky@msu.edu)

Dept. of Forestry and Great Lakes Bioenergy Research
Center, Michigan State University, East Lansing

Project Goals: (1) Identify the key regulatory genes that control xylogenesis and secondary wall biosynthesis; (2) Characterize the biological functions of cell wall biosynthesis candidate genes identified in model plants by Thrust I collaborators; and (3) Develop biotechnology tools that allow us to create woody biomass feedstocks with altered cell walls that are more easily digestible thereby releasing higher quantities of fermentable sugars.

Wood is gaining popularity as a source of fermentable sugars for liquid fuel production. However, our current knowledge on the genetic control of woody biomass formation is limited. Secondary wall of wood consists of a complex mixture

of cellulose, hemicellulose, and lignin. Proportional variability within the mixture of the three major components varies depending on the species of feedstock used, growing site, climate, age and the part of the plant harvested. The essentially uncontrolled variability of biomass properties presents process design and operating challenges for the production of bioenergy from woody feedstocks. Better understanding of the molecular mechanisms underlying its biosynthesis will help us develop biotechnological means to genetically control key pathways that determine the quantity and quality of the biomass.

In an effort to identify the transcriptional regulatory network controlling the biosynthetic process, we developed an experimental system that induces ectopic development of secondary wall in *Arabidopsis thaliana*. Using this system, we carried out Affymetrix GeneChip and Illumina Digital Gene Expression analyses to identify a battery of genes differentially expressed during secondary wall biosynthesis. These analyses allowed us to identify a group of transcription factors whose expression is coincided or preceded with the induction of secondary wall biosynthetic genes. The list includes a novel transcription factor AtC3H14 that could activate the transcription of all of the secondary wall biosynthesis genes tested, suggesting its potential role as another master regulator of secondary wall biosynthesis. Based on these results, we derived a tentative hierarchical transcriptional regulatory network leading to biosynthesis of secondary wall components. In order to confirm the relationship between transcription factors and their target genes, we are using both *in vivo* transcriptional activation assay and electrophoretic mobility shift assay. The current study tested our hypothesis that the selected transcription factors are responsible for the activation of the individual genes involved in the biosynthesis of secondary wall, and generated additional testable hypotheses. This presentation will describe (1) our strategy to identify transcription factors regulating secondary wall biosynthesis by using inducible secondary wall thickening system and time-course whole genome transcriptome profiling and (2) functional characterization of selected candidate genes in the network.

Utility promoters with 'freedom-to-operate' constitute a key enabling tool for biotechnological improvement of bioenergy crops. In this project, we are developing strong utility promoters that can drive target gene in a developing xylem-specific manner. We have obtained tissue-specific transcriptome profiles in poplar stems and identified candidate promoters. These promoters were fused to GUS reporter gene and expressed in transgenic poplars, which confirm their tissue-specific expression. We are testing the utility of these promoters with an anthocyanin biosynthesis gene in transgenic poplars.

4

Transgenic Poplars with Altered Lignins for Improved Biomass Pretreatment and Saccharification

Hoon Kim^{1*} (hoonkim@wisc.edu), Fachuang Lu,¹ Ali Azarpira,¹ Jorge Rencoret,¹ **John Ralph**,¹ Bruce E. Dale,² Curtis G. Wilkerson,² Nick Santoro,² Shane Cantu,² Cliff Foster,² Rebecca Van Acker,³ Bart Ivens,³ Wout Boerjan,³ and Shawn D. Mansfield⁴

¹DOE Great Lakes Bioenergy Research Center and Dept. of Biochemistry, University of Wisconsin, Madison; ²DOE Great Lakes Bioenergy Research Center and Michigan State University, East Lansing; ³Dept. of Plant Systems Biology, Flanders Institute for Biotechnology (VIB), and Dept. of Plant Biotechnology and Genetics, Ghent University, Gent, Belgium; and ⁴Dept. of Wood Science, University of British Columbia, Vancouver, Canada

Project Goals: The project is aimed at reducing plant cell wall recalcitrance toward enzymatic saccharification by altering lignin composition and structure to allow pretreatment methods to be more efficient.

Introduction

Several approaches are being targeted to explore biomass crop improvement for more efficient conversion (GLBRC Roadmap Goal to “Apply cutting-edge research to help create a new generation of sustainable bioenergy feedstocks, processing technologies and fuels”). One approach, recognizing that lignin is a major factor in plant cell wall (CW) recalcitrance to breakdown (of the polysaccharides to simple sugars), capitalizes on mechanistic insight gained from examining lignin-pathway transgenics in extensive collaborative studies. It is now well recognized that massive compositional changes can be achieved, particularly by misregulation of the various hydroxylase genes.^{1,2} Wild-type poplar has a guaiacyl-syringyl lignin, i.e., one comprised of guaiacyl (G) and syringyl (S) units in comparable amounts, but with only traces of *p*-hydroxyphenyl (H) units; these units are derived from coniferyl, sinapyl, and *p*-coumaryl alcohols respectively. Downregulation of C3H produces lignins rich in the normally minor H units; downregulation of F5H produces G-rich lignins, and upregulation of F5H produces S-rich lignins. It has also become clear that monomer-substitution can occur. For example, COMT-deficient plants incorporate 5-hydroxyconiferyl alcohol into their lignins (replacing some of the sinapyl alcohol), CAD-deficient plants incorporate more hydroxycinnamaldehydes, and recent evidence is that CCR-deficient plants incorporate the hydroxycinnamic acids themselves into the polymer.^{3,4} Thus lignification is considered to be particularly metabolically plastic. Altering the lignin composition/structure, allowing it to be more readily freed (by pretreatments) from the polysaccharide components, can provide enormous energy savings for biomass conversion. Transgenics are therefore being examined for their improved pulping potential, and for improved digestibility, directly or after pretreatment.

F5H-upregulated Poplar

A lignin compositional change particularly targeted for improved alkaline pulping also results in a significant improvement in CW digestibility. F5H upregulation, driven by an appropriate promoter (C4H), produces lignins derived primarily from sinapyl alcohol; the Mansfield group has generated plants strikingly rich in syringyl units.⁵ As a result, the lignin chains are rather homogeneous, being composed essentially only from one resinol unit (from initial dimerization) and β -ether-linked units. The lignins also have only a low degree of polymerization (and are therefore low-molecular-weight). These and other factors result in cell walls that saccharify more efficiently following acidic or basic pretreatment methods⁶ – see Figure 1. We are currently evaluating the effects of the Ammonia Fiber Expansion (AFEX) pretreatment.

CCR-deficient Poplar

CCR-downregulated poplars have slightly lower lignin levels and incorporate ferulic acid into the polymer.^{3,4} CCR-deficient poplar materials saccharify particularly efficiently after mild basic pretreatments – Figure 2. AFEX pretreatment is again being examined.

Conclusions

Already, results from these transgenics (and others), targeted because their lignins were expected to have ‘favorable pretreatment properties’, suggest that altering lignification in biomass crops can significantly lower recalcitrance to pretreatment and saccharification, providing improved bioconversion efficiencies.

References

1. J. Ralph, K. Lundquist, G. Brunow, F. Lu, H. Kim, P.F. Schatz, J.M. Marita, R.D. Hatfield, S.A. Ralph, J.H. Christensen and W. Boerjan. Lignins: natural polymers from oxidative coupling of 4-hydroxyphenylpropanoids. *Phytochemistry Reviews*, 2004, 3, 29-60.
2. W. Boerjan, J. Ralph and M. Baucher. Lignin biosynthesis. *Annual Reviews in Plant Biology*, 2003, 54, 519-549.
3. J. Ralph, H. Kim, F. Lu, J.H. Grabber, J.-C. Leplé, J. Berrio-Sierra, M. Mir Derikvand, L. Jouanin, W. Boerjan and C. Lapierre. Identification of the structure and origin of a thioacidolysis marker compound for ferulic acid incorporation into angiosperm lignins (and an indicator for cinnamoyl-CoA reductase deficiency). *The Plant Journal*, 2008, 53, 368-379.
4. J.-C. Leplé, R. Dauwe, K. Morreel, V. Storme, C. Lapierre, B. Pollet, A. Naumann, Gilles, K.-Y. Kang, H. Kim, K. Ruel, A. Lefebvre, J.-P. Josseleau, J. Grima-Pettenati, R. De Rycke, S. Andersson-Gunnerås, A. Erban, I. Fehrle, M. Petit-Conil, J. Kopka, A. Polle, E. Messens, B. Sundberg, S.D. Mansfield, J. Ralph, G. Pilate and W. Boerjan. Downregulation of cinnamoyl coenzyme A reductase in poplar; multiple-level phenotyping reveals effects on cell wall polymer metabolism and structure. *Plant Cell*, 2007, 19, 3669-3691.
5. J.J. Stewart, T. Akiyama, C.C.S. Chapple, J. Ralph and S.D. Mansfield. The effects on lignin structure of overexpression of ferulate 5-hydroxylase in hybrid poplar. *Plant Physiology*, 2009, 150, 621-635.
6. S.K. Huntley, D. Ellis, M. Gilbert, C. Chapple and S.D. Mansfield. Significant increases in pulping efficiency in

C4H-F5H-transformed poplars: Improved chemical savings and reduced environmental toxins. *Journal of Agricultural and Food Chemistry*, 2003, 51, 6178-6183.

This research was funded in part by the DOE Great Lakes Bioenergy Research Center (DOE Office of Science BER DE-FC02-07ER64494).

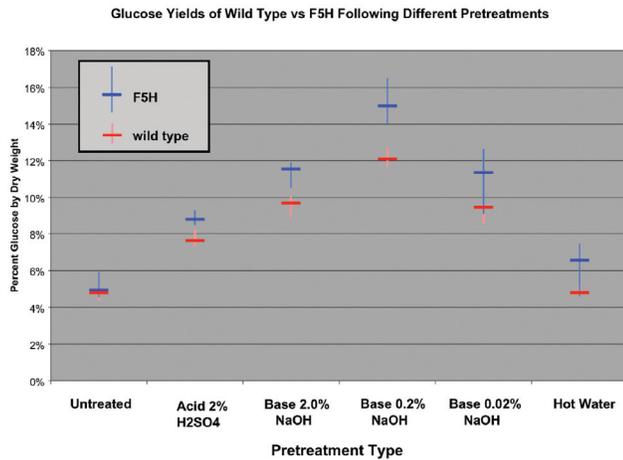


Figure 1. Comparison of saponification glucose yields following various pretreatments; except for untreated material, the F5H transgenic produced significantly higher yields.

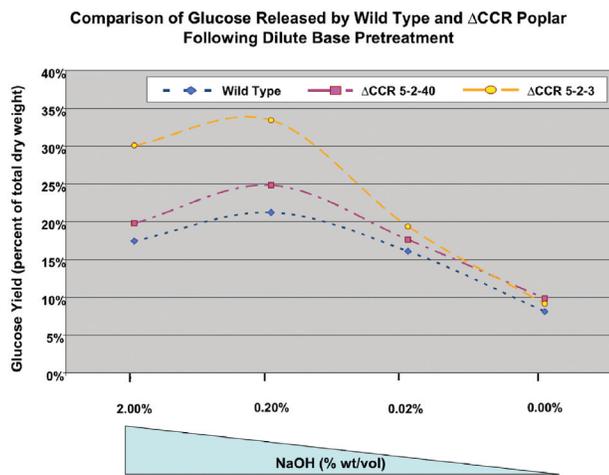


Figure 2. Glucose release for two CCR-deficient transgenics and their controls, following base pretreatment. The most CCR-deficient line has markedly elevated glucose release. Note: in both cases, non-exhaustive saccharification conditions were used allow differences to be meaningfully represented.

5 Biomass Trait Screening in *Brachypodium* Accessions and Mutant Populations

Cynthia Cass,^{1,2} Chris Schwartz,^{1,3} Laura Severs,^{1,2} Samantha Sakol,^{1,2} Michael Steinwand,⁴ John Vogel,⁴ Richard Amasino,^{1,3} and John Sedbrook^{1,2*} (jcsedbr@ilstu.edu)

¹DOE Great Lakes Bioenergy Research Center (GLBRC), Madison, Wis. (<http://www.greatlakesbioenergy.org/>); ²School of Biological Sciences, Illinois State University, Normal; ³Dept. of Biochemistry, University of Wisconsin, Madison; and ⁴Western Regional Research Center, U.S. Department of Agriculture, Albany, Calif.

Project Goals: Our short-term goal is to identify genes and gene variants affecting plant biomass quality and quantity, using *Brachypodium distachyon* as a model system. Our long-term goal is to use these data to improve bioenergy crops such as switchgrass, *Miscanthus*, and poplar through breeding and gene transformation.

The grass *Brachypodium distachyon* is emerging as an important model system for bioenergy crop grasses such as switchgrass and *Miscanthus* owing to its small genome size (~300Mbp), small stature, short generation time, transformability, and self-fertilization. The DOE Joint Genome Institute (JGI) has sequenced the *Brachypodium* genome and a large number of *Brachypodium* Expressed Sequence Tags (ESTs). These data are easily accessible and searchable at www.Brachypodium.org. We are taking a two-pronged approach to identify novel biomass trait genes and allele variants in *Brachypodium*. We are surveying a genetically diverse collection of wild type *Brachypodium* accessions for a variety of traits relevant to biomass production including cell wall hydrolytic enzyme digestibility, cell wall composition and structure, and flowering time, the last of which has a profound affect on biomass production. Not only are these data relevant for identifying gene variants that could be employed to improve bioenergy crops, they are also essential for understanding how different *Brachypodium* genetic backgrounds could modify the phenotypes of novel mutations. In that regard, we are screening through large collections of *Brachypodium* EMS mutagenized plants using both Near Infrared (NIR) spectroscopy (generates a fingerprint of cell wall composition) and an HPLC based cell wall digestibility assay (detects differences in the release of glucose and xylose from cell wall polysaccharides). We will present data on some promising mutants we have identified and discuss how the DOE Great Lakes Bioenergy Research Center (GLBRC) facilitates phenotypic characterization and gene discovery.

6

EST-SSR Markers Discriminate Switchgrass Ecotypes

M.D. Casler^{1*} (michael.casler@ars.usda.gov),
J.E. Zalapa,² S.M. Kaeppler,² N. de Leon,² and
C.M. Tobias³

¹USDA-ARS, U.S. Dairy Forage Research Center, Madison, Wisconsin; ²Dept. of Agronomy, University of Wisconsin, Madison; and ³USDA-ARS, Western Regional Research Center, Albany, Calif.

<http://www.greatlakesbioenergy.org>

Project Goals: Our objective was to use EST-SSR markers to discriminate upland from lowland ecotypes, using broader set of cultivars and individuals per cultivar than has been previously investigated. Longer-term goals of these studies are: (1) to use DNA markers to assist in identifying and selecting parents for development of heterotic gene pools and hybrid cultivars in switchgrass, and (2) to identify hybrid or backcross genotypes of mixed upland and lowland parentage, both in breeding programs and in natural populations.

Switchgrass (*Panicum virgatum*) is an important crop for bioenergy feedstock development. It is native to North America, ranging from Mexico to Canada east of the 100th meridian and adapted to a wide range of habitats (e.g., tallgrass prairie, savanna riparian habitats, etc.). Switchgrass has a range of ploidy from $2n=2x=18$ to $2n=12x=108$ and two main ecotypes: upland and lowland. The two ecotypes originate in different habitats: the upland type, originating on upland soils and the lowland type, originating in riverine and riparian habitats. Further, upland and lowland switchgrass have significantly different adaptations, with lowland types generally found south of 35°N latitude or USDA Hardiness Zones 6 and higher, while upland types tend to be more frequent at higher latitudes. Because there is a significant difference in morphology, growth pattern, and adaptation zones between the upland and lowland ecotypes, it is important to be able to easily discriminate between upland and lowland ecotypes. The presence of hybrid vigor in upland x lowland crosses and the possibility that the two ecotypes may act as natural heterotic groups creates a further need for accurate and efficient discrimination between the two ecotypes. Previous studies to identify DNA markers to discriminate between upland and lowland ecotypes have utilized a very small number of lowland ecotypes, typically only the two most common cultivars, Alamo and Kanlow.

Our objective was to use EST-SSR markers to discriminate upland from lowland ecotypes, using broader set of cultivars and individuals per cultivar than has been previously investigated. Longer-term goals of these studies are: (1) to use DNA markers to assist in identifying and selecting parents for development of heterotic gene pools and hybrid cultivars in switchgrass, and (2) to identify hybrid or backcross

genotypes of mixed upland and lowland parentage, both in breeding programs and in natural populations.

Seven lowland cultivars [Alamo (n=16), Kanlow (n=16), Miami (n=2), SG5 (n=9), Stuart (n=2), Timber (n=9), and Wabasso (n=6)] and 11 upland cultivars [Blackwell (n=16), Carthage (n=8), Cave-in-Rock (n=16), Dacotah (n=16), Forestburg (n=17), Pathfinder (n=9), Shawnee (n=4), Shelter (n=5), Summer (n=16), Sunburst (n=9), and Trailblazer (n=8)] were included in this study, with number of plants for each cultivar shown in parentheses. Additionally, in order to determine their origin, 8 unique switchgrass plants from our USDA-ARS breeding program were included in the study. We used 40 EST-SSR loci (381 alleles) from a total of 2351 that gave optimal amplification (i.e., were highly repeatable, showed correct segregation according to the expected ploidy level of each cultivar, and did not produce artifactual peaks). Relationships among the 18 switchgrass cultivars were investigated using NTSYS-pc version 2.01 based on a molecular binary data set obtained for each individual within cultivars. The resulting binary data was analyzed using the SIMQUAL routine to generate Dice similarity coefficients. Dice similarity coefficients were then used to construct a phenogram employing the SAHN procedure based on the Unweighted Pair-Group Method of the Arithmetic Average (UPGMA). The binary data were also subjected to principal component analysis (PCA) using the EIGEN routine of NTSYS-pc version 2.01.

Plants classified as upland or lowland, based on origin and phenotype, were completely separated by EST-SSR markers, with only two exceptions (Figure 1). Those two exceptions were plants that originated in the USDA-ARS breeding program at Madison, Wisconsin. The plants had been classified as lowland based on phenotype, but their phenotypic is actually intermediate between the extreme upland and lowland phenotypes: intermediate heading date, intermediate height, intermediate color, and intermediate tiller size and density. Although these two individuals were distinct from both upland and lowland phenotypes, based on marker data these clustered with the upland ecotypes. In summary, the EST-SSR markers used in this study were extremely effective at discriminating between upland and lowland ecotypes and at identifying the genetic origin for two plants of unknown origin.

There was a moderate degree of differentiation between upland 4x and upland 8x groups, with about 70% of the individuals with each group discriminated from each other. Ploidy forms an effective reproductive barrier in switchgrass, so it is not surprising to see some level of genetic differentiation between tetraploid and octaploid individuals.

Research sponsored and funded by the Great Lakes Bioenergy Research Center Grant DE-FC02-07ER64494.

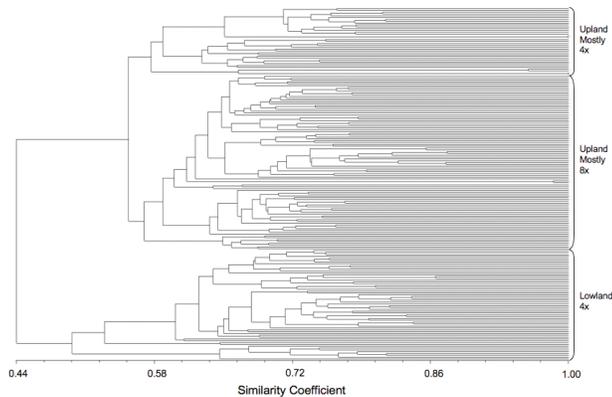


Figure 1. Cluster dendrogram of 192 switchgrass plants of known origin (upland 4x, upland 8x, and lowland 4x) grouped by similarity coefficients based on 381 EST-SSR markers.

7

Characterizing the Microbiome of Leaf-Cutter Ant Fungus Gardens

Garret Suen^{1,2*} (gsuen@wisc.edu), Jarrod J. Scott,^{1,2,3} Frank O. Aylward,^{1,2} Sandra M. Adams,^{1,2} Susannah G. Tringe,⁴ Adrián A. Pinto-Tomás,^{5,6} Clifton E. Foster,^{1,7} Markus Pauly,^{1,7,8} Paul J. Weimer,⁹ Kerrie W. Barry,⁴ Lynne A. Goodwin,¹⁰ Pascal Bouffard,¹¹ Lewyn Li,¹¹ Jolene Osterberger,¹² Timothy T. Harkins,¹² Steven C. Slater,¹ Timothy J. Donohue,^{1,2} and **Cameron R. Currie**^{1,2,3}

¹U.S. Department of Energy (DOE) Great Lakes Bioenergy Research Center and ²Dept. of Bacteriology, University of Wisconsin, Madison; ³Smithsonian Tropical Research Institute, Balboa, Ancon, Republic of Panama; ⁴DOE Joint Genome Institute, Walnut Creek, Calif.; ⁵Departamento de Bioquímica, Facultad de Medicina and ⁶Centro de Investigaciones en Estructuras Microscópicas, Universidad de Costa Rica, Ciudad Universitaria Rodrigo Facio, San José; ⁷Dept. of Biochemistry and Molecular Biology and ⁸DOE Plant Research Laboratory, Michigan State University, East Lansing; ⁹Dairy Forage Research Center, U.S. Department of Agriculture–Agricultural Research Services, Madison, Wis.; ¹⁰DOE Joint Genome Institute, Los Alamos National Laboratory, Los Alamos, New Mexico; ¹¹454 Life Sciences, a Roche Company, Branford, Conn.; and ¹²Roche Diagnostics, Roche Applied Science, Indianapolis, Ind.

Project Goals: Understanding efficient plant biomass degradation is a critical step toward the technological goal of ethanol production. Microbes are the primary organisms on earth capable of deconstructing lignocellulose, and herbivores can gain access to organic carbon stored in plant cell walls by forming symbiotic relationships with lignocellulolytic microbes. Our project aims to characterize an insect herbivore, leaf-cutter ants, which have been farming a fungus for food for ~10 million years. These

ants, one of the most widespread and dominant herbivores in the Neotropics, are capable of harvesting a tremendous amount of leaf-material to grow their fungus. They form massive colonies consisting of hundreds of fungus garden chambers supporting millions of workers. We are utilizing a combination of culture-independent and culture-dependent approaches to characterize the microbiome of leaf-cutter ant fungus gardens, and thereby understand how this microbial community synergistically degrades plant biomass.

For ~10 million years, leaf-cutter ants have been farming fungus for food. The ant–fungus system is one of the most complex described symbioses in nature, consisting of at least four mutualists and two pathogens. These ants, which are one of the most widespread insects in the Neotropics, can have massive colonies containing hundreds of fungus garden chambers and millions of workers that forage for hundreds of Kg (dry weight) of leaf material each year. The success of the leaf-cutter ants can be attributed, in part, to their ability to convert plant biomass into nutrients through their obligate mutualistic fungus. This fungus serves as the primary food source for the entire colony, and in return, the ants provide the fungus with a constant source of nutrients, protection from competitors, and dispersal through colony founding. As a result, the leaf-cutter ants serve as an excellent model for understanding how plant biomass degradation occurs in a highly-evolved, natural system.

Interestingly, very little is known about plant biomass degradation in leaf-cutter ant fungus gardens, even though this process likely plays a critical role in these colonies reaching immense sizes. For example, it is thought that the fungus the ants cultivate for food is responsible for the majority of plant biomass degradation in the garden, despite the fact that it is incapable of deconstructing lignocellulose. To explore the possibility that a largely uncharacterized microbial community is present and responsible for biomass deconstruction in leaf-cutter ant fungus gardens, we describe the fungus garden microbiome of the leaf-cutter ant *Atta colombica*.

We employ a combination of sugar composition analysis, 16S rDNA sequencing, community metagenomics, and whole-genome sequencing to demonstrate that lignocellulose is degraded within leaf-cutter ant fungus gardens, and that this microbiome is dominated by γ -proteobacteria in the family *Enterobacteriaceae*. Our analysis also identified a diversity of microbial genes predicted to encode for enzymes involved in cellulose and hemicellulose deconstruction, suggesting that a community of microbes is likely involved in plant biomass deconstruction. Comparative metagenomic analyses with 13 other microbiomes revealed that the leaf-cutter ant fungus garden exhibits a similar carbohydrate-degrading potential as bovine rumen, which is also capable of processing large amounts of plant biomass. Finally, genomic and physiological characterization of two dominant bacteria in the fungus garden provided evidence for their capacity to degrade lignocellulose, and suggests a potential mutualism, as these bacteria are known nitrogen-fixing symbionts of leaf-cutter ants. Our analysis of the leaf-cutter ant fungus

garden microbiome provides insight into how this microbial community synergistically deconstructs plant biomass.

8

Optimization of Enzymes for Alkaline-Pretreated Biomass Conversion

Shishir Chundawat,^{1,3} Dahai Gao,^{1,3} Venkatesh Balan,^{1,3} Goutami Banerjee,^{2,3} John Scott-Craig,^{2,3} Jonathan D. Walton,^{2,3} David Mead,⁴ Phil Brumm,⁴ and **Bruce Dale**^{1,3*} (bdale@egr.msu.edu)

¹Biomass Conversion Research Laboratory, Dept. of Chemical Engineering and Materials Science, ²DOE Plant Research Laboratory, and ³DOE Great Lakes Bioenergy Research Center, Michigan State University, East Lansing; and ⁴Lucigen Corporation, Madison, Wisconsin.

Project Goals (Abstracts 8 and 9): The GLBRC's mission is grand, but simply stated: to perform the basic research that generates technology to convert cellulosic biomass to ethanol and other advanced biofuels. To accomplish its mission in a university research environment, the GLBRC will: apply cutting-edge research to help create a new generation of sustainable bioenergy feedstocks, processing technologies, and fuels; evaluate the economic and environmental impacts of these new technologies; use the results to guide research activities; bring technological advances to other academic scientists, the private sector, and the marketplace; balance mission-driven project management and evaluation with the creative milieu of its academic, private sector and national laboratory research partners; recruit broad segments of the academic, industrial, and national laboratory communities to develop and participate in relevant research programs; and provide a training program for future leaders of the biofuels industry.

Enzymes that release fermentable sugars from biomass feedstocks are one of the major costs in converting lignocellulose to ethanol. The major goal of this project is to build optimal enzyme mixtures for alkaline-pretreated biomass. This project has two stages of development. The first is to define an optimized "minimal enzyme set" composed of those enzymes that are almost certainly essential for lignocellulose degradation. This set includes exo-glucanases (cellobiohydrolases; CBHs), endo-glucanase (EG), endo-xylanase (EX), β -xylosidase (BX), and β -glucosidase (BG). The second part is to develop an "optimized enzyme mixture" in which additional enzymes, called here "accessory" enzymes, are added to the minimal set in order to create a superior mixture. By optimum we mean having the highest specific activity (lowest protein loading) to achieve a benchmark release (e.g., 85%) of glucose and xylose in a specified time (48 hr). We used Ammonia Fiber Expansion (AFEX) pretreated corn stover for this work. The enzymes for these experiments came from several sources. These include enzymes purified from commercial *Trichoderma reesei* preparations, *Trichoderma* genes expressed in *Pichia pastoris*, *Trichoderma*

proteins over expressed in *Trichoderma* itself and *A. nidulans* orthologs of the *Trichoderma* enzymes expressed in *Pichia pastoris*. Another important source of enzymes for our experiments are proteins, mainly bacterial, from other projects in the GLBRC. We have successfully completed the first stage of the project by defining the core enzymes needed to achieve the benchmark glucan and xylan conversions in 48 hr of hydrolysis. The second phase of the project using accessory enzymes along with the core enzymes is in progress. Preliminary results indicate that there is synergy operating between bacterial hemicellulases and fungal cellulases. When completed, the optimized mixture is predicted to have higher specific activity than current commercial mixtures, enabling a reduced enzyme load during hydrolysis.

9

Physicochemical Characterization of Alkali-Pretreated Lignocellulosic Biomass

Shishir Chundawat,^{1,3} Leonardo Sousa,^{1,3} James Humpala,^{1,3} Venkatesh Balan,^{1,3} **Bruce Dale**^{1,3*} (bdale@egr.msu.edu), Ramin Vismeh,^{2,3} A. Daniel Jones,^{2,3} Fachuang Lu,^{4,5} Lily Summer,^{4,5} Ali Azarpira,^{4,5} John Ralph,^{4,5} Chris Bianchetti,^{4,5} George Phillips,^{4,5} Bryon Donohoe,⁶ Michael Himmel,⁶ Umesh Agarwal,⁷ Paul Langan,⁸ S. Gnanakaran,⁹ and Albert Che¹⁰

¹Biomass Conversion Research Laboratory, Dept. of Chemical Engineering and Materials Science, ²Dept. of Biochemistry and Molecular Biology, and ³Great Lakes Bioenergy Research Center, Michigan State University, East Lansing; ⁴Dept. of Biochemistry and ⁵Great Lakes Bioenergy Research Center, University of Wisconsin, Madison; ⁶Chemical and Biosciences Center, National Renewable Energy Laboratory, Golden, Colorado; ⁷Fiber and Chemical Sciences Research, Forest Products Laboratory, Madison, Wisconsin; ⁸Biosciences Division and ⁹Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico; and ¹⁰Environmental Science and Chemistry, American University, Washington, D.C.

Project Goals: See goals for abstract 8.

The development of an economically viable and environmentally sustainable bio-based chemical industry has been impeded by the native recalcitrance of plant cell walls to thermochemical and biological processing. Alkaline pretreatments like Ammonia Fiber Expansion (AFEX) enhance cell wall digestibility through certain ultra-structural and chemical modifications that are currently poorly understood, unlike other acidic pretreatments (e.g. dilute acid, steam explosion). Understanding the physicochemical mechanisms by which alkaline based pretreatments enhance cell wall enzymatic digestibility would result in development of improved pretreatment methodologies and reduction in cellulosic ethanol production costs. The goal of this project is to identify ultra-structural and chemical modifications incorporated within lignocellulosic cell walls during

alkaline based treatments (e.g. AFEX, alkaline peroxide). High resolution microscopic (SEM, TEM) and 3D-EM-Tomographic studies indicate ultra-structural alteration of AFEX treated cell walls via formation of nanoporous (5-500 nm) tunnel-like networks. Closer analysis (ESCA, AFM and confocal fluorescence microscopy) of outer cell wall surfaces reveals presence of heterogeneous deposits rich in AFEX degradation products and other cell wall extractives (e.g. lignin, arabinoxylan based oligomers, calcium). More than 50 alkaline based degradation products have been identified and quantified using LC-MS/MS and GC-MS, with ammonolysis based by-products (acetamide and phenolic amides) being the predominant ones. Pretreated biomass was characterized by NMR to elucidate modification of various cell wall components during AFEX. Raman and XRD analyses indicate allomorphic conversion of cellulose I to III by treatment with anhydrous liquid ammonia. Importantly, the cellulose III allomorph was found to have a significantly higher rate of enzymatic hydrolysis than untreated cellulose possibly due to differences of glucan chain packing within the cellulose crystal lattice.

10

Combinatorial Discovery of Enzymes for Biomass Deconstruction

Brian G. Fox* (bgfox@biochem.wisc.edu), Abolfazi Arabshahi, Allison Riederer, Nathaniel L. Elsen, Pirom Chenprakon, David Stephenson, Paul Weimer, and Yury Bukhman

Dept. of Biochemistry, Great Lakes Bioenergy Research Center, University of Wisconsin, Madison

Project Goals: The goal of our work is to provide a new combinatorial paradigm for evaluation of enzymes from new environmental sources as well as synthesized genes and engineered enzymes.

The goal of our work is to provide a new combinatorial paradigm for evaluation of enzymes from new environmental sources as well as synthesized genes and engineered enzymes. The GLBRC bioenergy platform derives from work with cell-free translation at the NIH Protein Structure Initiative-funded Center for Eukaryotic Structural Genomics, where over 10,000 genes from various eukaryotic organisms have been cloned, tested for expression, and in the best performing cases, purified and subjected to structure determinations. An adaption of the modular vector design of this platform provides the basis for this new effort on genes and proteins contributing to cellulose destruction. Genes identified by bioinformatic evaluation of new genomes or by microarray evaluation of gene expression in cellulose utilizing organisms, and proteins identified by mass spectral proteomic studies can be targeted for high-throughput cloning and cell-free translation. Methods to prepare combinations of genes or to supplement existing mixtures of enzymes with new translation products are demonstrated to provide biomass deconstruction without need for purifica-

tion of translation products. Iterative substitution of gene variants can be used to identify protein homologs with better behavior in defined assays that can include diagnostic small molecules, model purified celluloses, or actual biomass substrates. Results of the application of these methods to characterization of the reactivity of various treated and untreated biomass materials will be presented.

This work is funded by DOE DE-FC02-07ER64494 as a subgrant from the Great Lakes Bioenergy Research Center to B.G.F.

11

Construction of Gram-Negative Consolidated Bioprocessors

David H. Keating* (dkeating@glbrc.wisc.edu), Jeffrey Gardner, Mary Tremaine, and Robert Landick

Great Lakes Bioenergy Research Center, University of Wisconsin, Madison

Project Goals: Research within the Great Lakes Bioenergy Research Center (GLBRC) aims to generate an improved understanding of the bottlenecks associated with conversion of lignocellulose to ethanol. We seek to construct consolidated bioprocessing strains capable of the degradation of lignocellulose, and fermentation of the liberated sugars to ethanol. We will then use these consolidated bioprocessing strains to better understand the current enzymatic and metabolic bottlenecks in cellulosic ethanologenesis. These studies are currently focused on the gram-negative bacterium *Escherichia coli* due to its sophisticated genetics, well-understood physiology, and widespread use as an industrial microbe.

The conversion of *E. coli* to a consolidated bioprocessor requires the introduction of genes encoding lignocellulases, as well as a mechanism for their secretion from the cell. We developed liquid and solid media assays that facilitate the rapid identification of bacteria capable of biomass degradation. Using this media, we have categorized known cellulose-degrading organisms such as *Cellvibrio japonicus* for their ability to degrade key GLBRC biomass substrates, and to identify novel gram-negative cellulolytic strains. We have introduced cellulase genes from *C. japonicus* into *E. coli* and generated first generation strains capable of cellulose degradation. Using a genetic system developed for *C. japonicus*, we have obtained evidence that the majority of cellulolytic activity produced by *C. japonicus* is secreted via the Type II secretion system. We are currently engineering *E. coli* to express the *C. japonicus* Type II secretory apparatus in order to improve secretion of cellulolytic enzymes.

We are also engineering *E. coli* to efficiently produce and tolerate ethanol. To develop efficient *E. coli* ethanologens, we have constructed deletions that inactivate alternative pathways of electron flow, and introduced the *Zymomonas mobilis* pyruvate decarboxylase and alcohol dehydrogenase genes,

which allow efficient conversion of pyruvate to ethanol. Our studies demonstrated that this ethanologen was capable of efficient ethanol production under micro-aerobic conditions, but grew very poorly under strictly anaerobic conditions. To better understand the mechanism behind the poor anaerobic growth of this strain, we subjected it to thirteen rounds of sequential subculture under anaerobic conditions. We identified 32 mutants that grew well anaerobically with glucose as a carbon source, five of which exhibited productivities greater than that of strain KO11 (the current *E. coli* benchmark). We are currently subjecting these strains to re-sequencing and multiomic analysis to understand the genetic changes responsible for restoration of anaerobic growth.

Collectively, we expect that these approaches will allow for the isolation of lead organisms that can then be subjected to additional metabolic engineering and directed evolution, with the aim of improving lignocellulolytic ability, ethanol tolerance, and ethanol production.

12 Modeling and Summarizing Growth Curves from High-Throughput Screening Data

Yury V. Bukhman* (ybukhman@glbrc.wisc.edu), Jason Shao, Trey Sato, and Sydnor Withers

DOE Great Lakes Bioenergy Research Center, University of Wisconsin, Madison

Project Goals: This project provides computational support for high-throughput screening of large collections of microbes grown on various media. The screening experiments at GLBRC aim to identify yeast strains that are most suited for growth in media that include potential feedstocks for industrial fermentation of plant biomass into ethanol. The experiments monitor cell density of yeast cultures grown in microtiter plates by measuring optical density at 600 nm at regular intervals. We aim to mathematically model the resultant growth curves and summarize them by computing a few essential characteristics of each curve.

We have developed R scripts to process and summarize our screening data. We chose to fit theoretical functions, instead of using numerical differentiation or local regression methods, in order to be able to work with limited number of data points, which maintains maximal robotic screening throughput. We have tested several known growth curve models and found that the Gompertz function^{1,2} gives the best results with our data. We have also developed methods for automatic generation of initial guesses of curve parameters and for dealing with growth curves that have anomalous shapes. Upon processing the screening data, our software generates a table of biologically meaningful growth curve characteristics, such as the fastest intrinsic growth rate, lag time, and total growth. These characteristics were used to select yeast strains with optimal growth properties in the examined

media. Although the development of this methodology was motivated by the needs of the yeast screening program at GLBRC, it has general utility for high-throughput screening studies of microbial strain collections.

References

1. Gompertz, B. "On the Nature of the Function Expressive of the Law of Human Mortality, and on a New Mode of Determining the Value of Life Contingencies." *Phil. Trans. Roy. Soc. London* 123, 513-585, 1832.
2. Zwietering, M.H., et al. "Modeling of the Bacterial Growth Curve." *Applied and Environmental Microbiology* 56 (6), 1875-1881, 1990.

13 Exploiting Natural Diversity in Wild Yeast Strains to Dissect the Mechanisms of Ethanol and Thermotolerance

Jeffrey A. Lewis^{1,2,*} (jalewis4@wisc.edu), Shau Chee C. Chung,² and **Audrey P. Gasch**^{1,2}

¹Great Lakes Bioenergy Research Center, Madison, Wis.; and ²Laboratory of Genetics, University of Wisconsin, Madison

Project Goals: see below

Pretreatment of cellulosic biomass for enzymatic saccharification generates degradation products that impair microbial fermentation. Additional stressors that affect the final ethanol yield include osmotic stress, oxidative stress, heat, and ethanol itself. This project specifically focuses on ethanol and heat tolerance in *Saccharomyces cerevisiae*—stressors that are especially relevant for simultaneous saccharification and fermentation.

Acquired stress resistance is the phenomenon where cells exposed to a mild dose of a primary stress can survive an otherwise lethal dose of a subsequent stress. In the case of ethanol, acquired resistance likely reflects the adaptation required to survive the increasing ethanol concentrations that accumulate during fermentation. While studying acquired stress resistance in *S. cerevisiae*, we made a surprising discovery—our lab strain (S288c) could not acquire resistance to ethanol. Intriguingly, our lab strain could acquire thermotolerance, even though the mechanisms of ethanol tolerance and thermotolerance are thought to be shared.

By testing a wide panel of wild and industrial yeast isolates, we verified that the lab strain was indeed an outlier; acquired ethanol resistance is wide-spread in nature. We have compared the global transcriptional response to both ethanol and heat, in both the lab strain and two wild strains (with a natural ability to acquire ethanol tolerance). Stark differences existed in the transcriptional profile between the lab strain and the two wild strains. In particular, genes known to function in both ethanol and thermotolerance (i.e. genes involved in membrane metabolism, chaperones, and

trehalose metabolism) were differentially expressed. These differences are valuable clues for elucidating the regulatory circuits for both acquired ethanol resistance and acquired thermotolerance. Screening of mutants (chosen on the basis of the transcriptome data) has led to the discovery of novel genes of involved in ethanol resistance, highlighting the power of this approach.

14

Utilization of Directed Evolution, Resequencing, and Multiomics to Improve Ethanol Tolerance in *Escherichia coli*

Michael S. Schwalbach* (mschwalbach@glbrc.wisc.edu), Frances Tran, Mary T. Tremaine, Patricia J. Kiley, Robert Landick and **David H. Keating**

Great Lakes Bioenergy Research Center, University of Wisconsin, Madison

Project Goals: The Great Lakes Bioenergy Research Center mission is grand, but simply stated; to perform the basic research that generates technology to convert cellulosic biomass to ethanol and other advanced biofuels. While the larger mission is pursued on multiple fronts by many researchers, the focus of this particular research project is to understand the bottlenecks associated with conversion of lignocellulose to ethanol. Ethanol toxicity represents one such bottleneck since elevated concentrations of ethanol inhibit fermentation yields and ultimately growth in ethanologenic bacteria.

Research within the Great Lakes Bioenergy Research Center aims to understand the bottlenecks associated with conversion of lignocellulose to ethanol. Ethanol toxicity represents one such bottleneck since it inhibits fermentation yields and ultimately growth in ethanologenic bacteria. However, ethanol tolerance capabilities vary amongst ethanologenic bacteria, with bacteria such as *Zymomonas mobilis* and *Lactobacillus buchneri* able to survive external ethanol concentrations approaching 15% (v/v) while others, such as *Escherichia coli*, are inhibited at concentrations greater than 3% (v/v). Presently the molecular mechanisms underlying differences in ethanol sensitivity are not wholly characterized nor are the cellular responses that mitigate the toxic effects of ethanol well understood. To understand better which traits contribute to improved ethanol tolerance, we conducted directed evolution experiments to incrementally increase ethanol tolerance of *E. coli* strain MG1655. Ethanol tolerant mutants and wild type cells were then exposed to 4% ethanol and the response was monitored via multiomics. From these growth experiments we collected transcriptomics, metabolomics, lipidomics and proteomic data from three time points: mid-log, ten minutes post ethanol addition, and one hour post ethanol addition. Global transcriptional analysis from cells collected ten minutes after ethanol addition indicated that wild type and the ethanol tolerant mutant exhibited mechanistically similar responses to initial ethanol challenge. Conversely, examination of the tran-

scriptional response of cells collected one hour after ethanol addition determined the initial response had diminished in the ethanol tolerant mutant, whereas wild type maintained a transcriptional profile similar to what was observed ten minutes after ethanol addition. Our results suggest that the ethanol tolerant mutant was able to mitigate the adverse effects of the ethanol challenge quickly whereas wild-type remained adversely affected by ethanol for a longer period of time. By comparing the complete multiomic responses and genomic differences in multiple strains evolved to tolerate varying concentrations of ethanol, we will identify the key cellular processes associated with ethanol tolerance.

15

Generation of a Computational Metabolic Network Representing the Pangenome of *Escherichia coli* and Construction and Validation of Six Strain-Specific Models

David J. Baumler^{1,2*} (dbaumler@wisc.edu), Roman G. Peplinski,² Jeremy D. Glasner,² Jennifer L. Reed,^{1,3} and Nicole T. Perna^{1,2,4}

¹Bringing Advanced Computational Techniques to Energy Research (BACTER) Institute; ²Genome Center of Wisconsin; ³Dept. of Chemical and Biological Engineering, and ⁴Dept. of Genetics, University of Wisconsin, Madison

Project Goals: BACTER is a research training initiative funded by the U.S. Department of Energy to promote the development of computational biology and bioinformatics techniques that are of direct value to energy science.

Within the genus *Escherichia*, only one strain, the laboratory strain *E. coli* K-12 MG1655, has had a genome-scale metabolic computational model constructed. This single model has proven useful for many applications, such as guiding the bioengineering of strains for increased production of desired end-products. We sought to enhance these efforts by constructing metabolic models for additional *E. coli* strains using a strategy based on analysis of the collective contents of all *E. coli* genomes, or the pangenome. We used 17 complete genome sequences (Table 1) to generate an *E. coli* pangenomic metabolic network consisting of the collective information from 76,080 ORFs. These ORFs were clustered into 17,647 orthologous groups. The 1,260 orthologous groups containing the ORFs used in the most recent metabolic network for *E. coli* K-12 were identified, and the gene to protein to reaction associations were propagated to the other *E. coli* strains. All remaining orthologous groups were surveyed for new metabolic reactions to add to the pangenomic network. This allowed us to update the metabolic model for *E. coli* K-12 MG1655 to account for 1,322 ORFs, and now includes a pathway for phenylacetate metabolism. A model was constructed for the very similar *E. coli* K-12 strain W3110 and differs by one metabolic reaction for galactitol transport/utilization. Genome-scale metabolic models were also constructed for enterohemorrhagic *E. coli*

O157:H7 strains EDL933 and Sakai, and uropathogenic *E. coli* strains CFT073 and UTI89 (Table 2). The metabolic networks for the pathogens contained numerous lineage-specific ORFs when compared to the K-12 models. All six *E. coli* models were used to simulate growth in different conditions. The results were compared to experimental data we collected for each strain including tests for the utilization of 76 different carbon sources in conditions with or without oxygen, and growth in batch culture. The experiments reveal metabolic differences between strains and the *in silico* results accurately predict some of these differences. Our findings demonstrate that use of the pangenomic metabolic network allows rapid construction of additional *E. coli* strain-specific models that can accurately predict strain-specific phenotypes and offers a larger suite of metabolic capabilities for engineering new *E. coli* strains.

Table 1. *E. coli* genomes used to construct the pangenome metabolic network

| Strain | ORFs |
|---|-------|
| <i>E. coli</i> K-12 MG1655 | 4,141 |
| <i>E. coli</i> K-12 W3110 | 4,171 |
| <i>E. coli</i> EDL933 (EHEC) ^a | 5,196 |
| <i>E. coli</i> Sakai (EHEC) ^a | 5,253 |
| <i>E. coli</i> CFT073 (UPEC) ^b | 4,889 |
| <i>E. coli</i> UTI89 (UPEC) ^b | 4,944 |
| <i>E. coli</i> 536 (UPEC) ^b | 4,599 |
| <i>E. coli</i> 53638 (EIEC) ^c | 5,172 |
| <i>E. coli</i> APEC O1 (APEC) ^d | 5,045 |
| <i>E. coli</i> ATCC 8739 | 4,236 |
| <i>E. coli</i> E2348/69 (EPEC) ^e | 4,652 |
| <i>E. coli</i> E24377A (ETEC) ^f | 4,953 |
| <i>E. coli</i> EC4115 (EHEC) ^a | 5,467 |
| <i>E. coli</i> HS | 4,393 |
| <i>E. coli</i> K-12 DH10B | 4,136 |
| <i>E. coli</i> SE11 | 4,973 |
| <i>E. coli</i> SMS-3-5 | 4,906 |

^aEnterohemorrhagic *E. coli* (EHEC)

^bUropathogenic *E. coli* (UPEC)

^cEnteroinvasive *E. coli* (EIEC)

^dAvian pathogenic *E. coli* (APEC)

^eEnteropathogenic *E. coli* (EPEC)

^fEnterotoxigenic *E. coli* (ETEC)

Table 2. *E. coli* strain-specific metabolic model information

| Strain | Additions | | | Deletions | | Essential Reactions ^a | Total in model | |
|----------------|-----------|-----------|----------|-----------|-----------|----------------------------------|----------------|-----------|
| | ORFs | Reactions | Isozymes | ORFs | Reactions | | ORFs | Reactions |
| K-12 MG1655 | - | - | - | - | - | - | 1,322 | 2,282 |
| K-12 W3110 | 0 | 0 | 0 | 1 | 1 | 0 | 1,321 | 2,281 |
| O157:H7 EDL933 | 38 | 9 | 19 | 51 | 56 | 8 | 1,328 | 2,235 |
| O157:H7 Sakai | 36 | 8 | 23 | 52 | 57 | 8 | 1,329 | 2,333 |
| UPEC CFT073 | 9 | 9 | 25 | 87 | 55 | 10 | 1,269 | 2,236 |
| UPEC UTI89 | 8 | 8 | 26 | 71 | 55 | 6 | 1,285 | 2,235 |

^aReactions with no orthologous ORF (s), but are essential to the *in silico* model

This work was supported through the BACTER training program (PI: Julie C. Mitchell University of Wisconsin-Madison [http://www.bacter.wisc.edu/BACTER/Welcome.html]), funded by the DOE:GTL #: DE-FG02-04ER25627, BACTER is a research training initiative funded by the U.S. Department of Energy to promote the development of computational biology and bioinformatics techniques that are of direct value to energy science.

16 Constraint-Based Analysis of Microbial Regulatory and Metabolic Networks for Ethanol Production

Joonhoon Kim^{1,2*} (jkim92@wisc.edu), Dipak Barua,^{1,2} and Jennifer L. Reed^{1,2}

¹Dept. of Chemical and Biological Engineering and ²DOE Great Lakes Bioenergy Research Center, University of Wisconsin, Madison

Project Goals: Design microbial strains for improved ethanol production using computational models of transcriptional regulation and metabolism.

Computational modeling and analysis of metabolic networks has been successful in metabolic engineering of microbial strains for valuable biochemical production. Limitations of currently available methods are that they are often based on reaction deletions rather than gene deletions and that they do not consider the regulatory networks that control metabolism. Thus, such methods may result in strategies that are not genetically feasible, or designed strains might not be able to grow due to the regulatory restrictions. To overcome these limitations, we developed an effective method to systematically integrate transcriptional regulatory networks and metabolic networks, which allows for the simulation of gene deletion and overexpression.

Using integrated transcriptional regulatory and metabolic models, we developed an automated approach (*Gene-Force*) for refining transcriptional regulatory rules against high-throughput growth phenotypic data. The developed approach was applied to well-curated transcriptional regulatory and metabolic models of *Escherichia coli* (Covert et al, 2003), and resulted in an overall ~ 10% improvement in model prediction accuracy for a large collection of mutant growth phenotypes (Glasner et al, 2003; Ito et al, 2005). An advantage of using an integrated model of metabolism and regulation is that an integrated model is better at predicting essential genes under a given condition, and hence it prevents gene deletions which are lethal from being included in the strategies. Accordingly, strains that are designed with regulatory considerations should grow better initially and may achieve the desired phenotype faster.

We have subsequently developed a new approach (OptORF) for identifying metabolic engineering strategies based on metabolic and transcription factor gene deletions, as well as gene overexpressions. This new approach uses integrated models of metabolism and regulation, and searches for the minimal metabolic and/or regulatory perturbations that couple biomass and biochemical production, thus proposing adaptive evolutionary strain designs. Using genome-scale models of *E. coli* (Covert et al, Nature 2003), we have implemented OptORF and compared its metabolic engineering strategies for ethanol production to those found using OptKnock (Burgard et al, 2003). The developed OptORF approach is general and can be applied to the production of different compounds in other biological systems.

This work was funded by the DOE Great Lakes Bioenergy Research Center (DOE BER Office of Science DE-FC02-07ER64494).

17

Identification of Stress-Tolerant *Saccharomyces cerevisiae* Strains for Fermentation of Lignocellulosic Feedstocks by High-Throughput Phenotypic Screening

Sydnor Withers,¹ Jason Shao,¹ Benjamin Bice,¹ Miguel Dominguez,¹ Venkatesh Balan,² Supratim Datta,⁴ Mark Davis,⁵ Tom Jeffries,^{5,6} Blake Simmons,⁴ Bruce Dale,² Wes Marner,¹ Yury Bukhman,¹ and **Trey Sato**^{1*} (tksato@glbrc.wisc.edu)

¹Great Lakes Bioenergy Research Center, Madison, Wis.; ²Michigan State University, East Lansing; ³Joint BioEnergy Institute, Emeryville, Calif.; ⁴BioEnergy Science Center and National Renewable Energy Laboratory, Golden, Colo.; ⁵University of Wisconsin, Madison; and ⁶USDA Forest Products Laboratory, Madison, Wis.

Project Goals: A major bottleneck in the conversion of lignocellulosic biomass into ethanol by the brewer's yeast *Saccharomyces cerevisiae* is the inhibition of fermentation resulting from cellular stress, which can be caused by degradation products generated during feedstock pretreatment. At the Great Lakes Bioenergy Research Center (GLBRC), we have sought to identify environmental and industrial *S. cerevisiae* isolates with greater tolerance to these stresses than laboratory strains, and understand the genetic, biochemical, and molecular traits contributing to these phenotypic differences. Results from these studies will provide insight for improving existing production strains, or developing new stress-tolerant strains for cellulosic ethanol production.

Although commonly used for the industrial production of grain ethanol, *S. cerevisiae* has a number of biological hurdles that currently prevent it from being widely utilized in the generation of fuel ethanol from lignocellulosic biomass.

Cellular and physiological stresses imposed by degradation products from feedstock pretreatment are known to limit the yield and efficiency of fermentation. Well-known degradation products include acetic acid, furfurals and lignin-derived phenolic compounds, all of which can vary in structure and concentration with different pretreatment processes.

Ongoing research at the GLBRC has compared environmental stress tolerance between laboratory and wild yeast strains, and discovered that genetic background is a significant determinant in the ability of strains to tolerate environmental stress. This also suggested that *S. cerevisiae* strains isolated from diverse natural or industrial environments, opposed to domesticated lab strains, may have traits that may allow for greater tolerance to the toxins that result from pretreatment of lignocellulosic biomass. To identify strains that may be tolerant to stresses imposed during cellulosic fermentation, we examined the growth properties of hundreds of diverse *S. cerevisiae* strains cultured in various lignocellulosic hydrolysates. These various hydrolysates were prepared from biomass pretreated by ammonium fiber expansion (AFEX), ionic liquid (IL), oxalic acid or dilute sulfuric acid. High throughput robotic screening and computational methods were developed and applied to identify the top performing strains in the various media conditions. One of the top strains included an environmental isolate that, in contrast to most other strains, grows well in AFEX-pretreated corn stover hydrolysate at elevated temperatures of 37 and 40°C. Characterization of these top performing strains for ethanol production and ability to ferment xylose is currently in progress and will be presented.

18

Efforts to Enhance Solar Hydrogen Production by Heterocyst-Forming Cyanobacteria

Eric L. Hegg^{1*} (EricHegg@msu.edu), Katrin Gaertner,¹ Adam Cornish,¹ Sigal Lechno-Yossef,² Xiaowen Kong,² C. Peter Wolk,² Hajime Masukawa,³ Robert P. Hausinger,^{1,3} Jeong-Jin Park,³ and Claire Vieille^{1,3}

DOE Great Lakes Bioenergy Research Center, ¹Dept. of Biochemistry and Molecular Biology, ²MSU-DOE Plant Research Laboratory, and ³Dept. of Microbiology and Molecular Genetics, Michigan State University, East Lansing

<http://www.greatlakesbioenergy.org/>

Project Goals: The overall goal of our project is to engineer *Anabaena* to enhance photobiological H₂ production to a commercially practical level. To accomplish this task we are using a number of different strategies including mutating nitrogenase to produce more H₂, expressing a native bidirectional [NiFe]-hydrogenase in heterocysts, and heterologously expressing an [FeFe]-hydrogenase in heterocysts. In addition, to guide genetic engineering and to enhance further H₂ production, we are elucidating the

pathways by which electrons are channeled to H₂ and the major pathways that compete for those electrons.

H₂ has the potential to become an important clean and renewable energy commodity, especially if it is generated by organisms such as cyanobacteria that use sunlight as the sole energy source and water as the ultimate electron donor. Hydrogenases (H₂ases) and nitrogenases (N₂ases) are the enzymes that produce H₂. Because these enzymes are also O₂-sensitive, oxygenic photosynthesis and H₂ production are normally separated temporally or spatially. *Anabaena* spp. and related cyanobacteria form specialized cells known as heterocysts in which N₂ases and H₂ases are protected from O₂ by inactivation of O₂-producing PSII, accelerated respiration, and synthesis of a thick envelope of glycolipids and polysaccharides that impedes O₂ penetration. Reductant required for N₂ fixation and H₂ production is generated by photosynthesis in vegetative cells and is transported to heterocysts as sugar. Our ultimate goal is to engineer a Hup⁻ strain of *Anabaena* (in which the uptake hydrogenase is inactive) to increase H₂ production to a commercially practicable level.

To increase H₂ production by N₂ase in *Anabaena*, we mutated residues near the active site, generating a total of 49 single variants. Several mutants exhibited significantly higher H₂ production rates in the presence of N₂ than did the parental Hup⁻ strain. Nineteen additional single and double site-directed mutants designed to obstruct a putative channel connecting the active site to the protein surface did not exhibit significantly higher H₂ production rates than did the parental Hup⁻ strain.

Two different strategies are being pursued to express H₂ases in *Anabaena* heterocysts. In the first strategy, efforts are under way to overexpress the native, bidirectional [NiFe]-H₂ase (Hox) genes in heterocysts using the strong *nif* promoter. Genes required for the maturation of Hox are being expressed on a replicating plasmid from the heterocyst-specific promoters of *coxBII* and *patB*. In the second strategy, [FeFe]-H₂ases and the proteins needed for their maturation are being heterologously expressed in *Anabaena* heterocysts using both integrating and replicating plasmids driven by heterocyst-specific promoters. To obtain optimal expression and activity, we are testing different combinations of [FeFe]-H₂ases and maturation proteins from a variety of organisms including *Chlamydomonas reinhardtii*, *Clostridium acetobutylicum*, and *Shewanella oneidensis*.

To elucidate how electrons are channeled to H₂ production, we compared gene expression in phototrophic, photoheterotrophic, and heterotrophic *Anabaena* cultures using RNA extracted separately from vegetative cells, heterocysts, and whole filaments. Principal component analysis of gene expression data confirmed that gene expression patterns in the vegetative cells differ from those in the heterocysts. 19%, 17%, and 16% of the genes are differentially expressed (at least 2X difference, p<0.01) between vegetative cells and heterocysts in phototrophic, photoheterotrophic, and heterotrophic growth conditions, respectively. In addition, gene expression patterns in vegetative cells vary in the different growth conditions, consistent with the fact that vegetative

cells are solely responsible for carbon uptake. Surprisingly, Hup genes and several N₂ase maturation genes are expressed in the vegetative cells in both phototrophic and photoheterotrophic cultures. These and other results will be discussed.

19 Genome-Enabled Analysis of Partitioning of Reducing Power During Fuel Production by the Photosynthetic Bacterium *Rhodobacter sphaeroides*

Wayne S. Kontur^{1*} (wskontur@wisc.edu), Eva C. Ziegelhoffer,¹ Melanie A. Spero,¹ L. Safak Yilmaz,² Daniel R. Noguera^{2*} (noguera@engr.wisc.edu), and Timothy J. Donohue^{1*} (tdonohue@bact.wisc.edu)

¹Dept. of Bacteriology, ²Dept. of Civil and Environmental Engineering, University of Wisconsin, Madison

Project Goals: We seek to understand and improve light- and feedstock-powered production of renewable fuels by the photosynthetic bacterium *Rhodobacter sphaeroides*. We have initially chosen to investigate hydrogen (H₂) production, both to optimize the production of H₂ itself and to serve as a model for partitioning of intracellular reducing potential (high-energy electron equivalents) utilized in microbial biofuel production in general. We are characterizing the intracellular pathways involved in distributing electrons throughout the cell via genome-enabled assays, such as microarray and proteomics analyses, coupled with biomass, polyhydroxybutyrate (PHB) and chemical oxygen demand (COD) analyses. We are using mutant strains to determine the effects of deleting genes predicted by the global gene expression assays to impact the intracellular reducing potential available for fuel production.

Rhodobacter sphaeroides is a photosynthetic purple non-sulfur bacterium that can accumulate a large pool of intracellular reducing potential (high energy electrons) during photoheterotrophic growth on various carbon sources, such as organic acids and biomass-derived sugars. We are studying the distribution of this reducing potential in the interest of increasing the fraction that goes toward fuel production. We are initially focusing on H₂ production, for its own utility as a fuel and in industrial purposes, and as a model to understand how electrons needed for production of other fuels will be portioned to other pathways.

We have obtained global transcript levels in cultures with various levels of H₂-production and compared them to those in non-H₂ producing control cultures to determine candidate gene products that contribute electrons to or siphon reductant from H₂ production. We also compared transcript levels in H₂-producing cultures grown on various carbon sources to characterize the relative expression levels of the genes involved in cellular pathways that act as electron sinks. In addition, we are assaying these same cultures for the

distribution of electrons between the cellular end-products, PHB, other biomass, H₂, and soluble microbial products.

From prior knowledge, for reasons outlined below, we expected four systems would impact the amount of intracellular reducing potential available for fuel production: nitrogenase, hydrogenase, the carbon dioxide fixation pathways, and PHB synthesis.

- Nitrogenase is the primary (or sole) source of H₂, produced as a byproduct of nitrogen fixation. As expected, expression of the nitrogenase structural genes is increased in all cultures that produce detectable H₂. We are testing whether expression level correlates to total H₂ production amounts; preliminary evidence suggests that cellular nitrogenase enzymatic activity does correlate with the amount of H₂ produced.
- Hydrogenase is expected to oxidize H₂ and recycle reducing potential back into cellular metabolism. We find that expression of the hydrogenase structural genes varies inversely with net H₂ production; cultures that produce relatively low amounts of H₂ (those using xylose, glucose or glycerol) have relatively high expression of hydrogenase genes, while cultures that produce relatively high amounts of H₂ (those using lactate or succinate) have relatively low expression of hydrogenase genes, which suggests that the presence of functional hydrogenase may contribute to the disparities in the amounts of H₂ produced from different carbon sources. Experiments are in progress to determine the level of functional hydrogenase enzyme made in these cultures and how the reducing potential is distributed in defined hydrogenase mutants.
- Carbon dioxide fixation consumes reducing power, so it has the potential to siphon electrons from fuel production. We find that expression of carbon fixation genes is also inversely proportional to H₂ production (as is the case for hydrogenase genes), which suggests that carbon dioxide fixation may contribute to the disparities in H₂ amounts from cultures grown on different carbon sources. *R. sphaeroides* has two sets of carbon fixation genes, and we are investigating mutants which only contain one gene from each set (*cfxA*- and *cfxB*-) to determine the relative importance of the individual pathways on H₂ production.
- *R. sphaeroides* can accumulate the polymer PHB as an energy and carbon storage compound. We find little difference in expression of PHB synthase genes in H₂-producing cultures grown on different carbon sources. However, we find that the amounts of PHB produced differ between such cultures, and we are testing the effect of deleting the PHB synthase gene on H₂ production.

20

Use of Proteomics Technologies for the Characterization of Proteins, Microbes, and Microbial Communities Important for Bioenergy Production

Mary S. Lipton* (mary.lipton@pnl.gov), Stephen J. Callister,¹ Thomas Taverner,¹ Kim K. Hixson,² Samuel O. Purvine,² Kristin E. Burnum,¹ Roslyn N. Brown,¹ Angela D. Norbeck,¹ Janani Shutthanandan,¹ Matthew E. Monroe,¹ Ronald J. Moore,¹ Therese R. Clauss,¹ Carrie D. Nicora,¹ Shishir Chundawat,³ Dahai Gao,³ Venkatesh Balan,³ Bruce Dale,³ Tim Donohue,⁴ Eva Ziegelhoffer,⁴ Audrey Gasch,⁴ Jeff Lewis,⁴ Robert Landick,⁴ Patricia Kiley,⁴ Daniel Noguera,⁴ Cameron R. Currie,⁴ Garrett Suen,⁴ and Jarrod J. Scott⁴

¹Biological Sciences Division and ²Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, Wash.; ³Michigan State University, East Lansing; and ⁴University of Wisconsin, Madison

Project Goals: This project employs the state of the art proteomics facilities resident at the Pacific Northwest National Laboratory for the rapid global determination of protein identification, and protein expression patterns in microbes or plants. Ranging from protein preparations purified from fungal and bacterial sources that demonstrate the ability to degrade lignocellulosic material to the quantitative proteomic profiling of microbes and microbial communities, the proteomics capability at PNNL is employed at producing data to further the understanding of systems important to the GLBRC.

Microbial processing for substrates to biofuels, whether the conversion of lignocellulosic material to ethanol or microbial biorefineries to produce hydrogen or electricity is a central part of the GLBRC mission. Inherent in the use of microbes for these purposes is the characterization of the fundamental machines of the cell, the proteins, and how these proteins dictate microbial function. The quantitative determination of protein expression patterns and how these patterns change with changing cell state is critical for the GLBRC to remove bottlenecks in the biofuels pipeline. Additionally, accurate measurements of protein levels and modifications will provide more extensive insights into both the plants and the microbes in the bioenergy pipeline. These analyses include temporal profiling through the analysis of time course studies, characterization of posttranslational modifications, and determination of subcellular localization of proteins. Extension of proteomic capabilities to community profiling will enable deeper understandings of how microbes interact with each other in environmental settings.

In the past year, the proteomics facility has supported the GLBRC in four aspects. The characterization of cellulolytic and hemicellulolytic rich enzyme cocktails found that the dominant cellulases were CBH I, Xyloglucanase, CBH

II, EG I, EG II, EG III, β -glucosidase. The dominant hemicellulases found in most enzyme cocktails were Endoxylanase (GH 11), β -xylosidase, arabinofuranosidase (GH 62 & 54) and Glucuronidase (GH 67). This data will help determine critical classes of cellulases and hemicellulases necessary for hydrolyzing lignocellulosic biomass and are currently absent in commercially available mixtures. Supplementation of a minimalist and optimal enzyme set for hydrolyzing ammonia fiber expansion (AFEX) treated biomass (i.e. corn stover) will help reduce the total number and amount (mg protein/gm substrate) of enzymes required for hydrolysis.

Quantitative analysis of protein abundance in cells is one method for illustrating the manner in which cells perform function or adapt to their environment. The proteomics facility is applying label free proteomics for the quantitative characterization of many bacterial and fungal systems to understand ethanol tolerance, ethanol production and hydrogen production. In *E. coli*, analysis of aerobic and anaerobic cell cultures yielded a combined total of 1697 proteins identified from the two cultures using strict cross correlation and cutoff values for the peptides and the requirement of two peptides per proteins. Of the 1697 proteins identified, 46 proteins were found in the anaerobic cultures only, 30 peptides were found in the aerobic cultures only, and 1621 proteins were identified in both cultures. Quantitative analysis of 1254 proteins between both of the samples showed 86 proteins showed at least a two-fold increase in abundance in the aerobic sample and 56 showed at least a two-fold increase in abundance in the anaerobic sample. The qualitative and quantitative characterization of *Rhodobacter sphaeroides* proteome showed an increase in abundance of nitrogen fixation genes when the organism was grown to stationary phase when compared with log phase. In yeast, differential protein expression patterns are being used to determine the mechanism behind ethanol tolerance.

Expansion of the proteomics from microbes into microbial communities can provide insights into many of the biological mechanisms present in the communities as well as information about strain heterogeneity depending on the size and quality of the sequenced genome. We are employing proteomic analyses to gain insight into the symbiotic relationship between leafcutter ants, fungi, and bacteria. This fungus digests the cellulose in the leaves and uses it to fuel its own growth. The leaf-cutter ants then feed on the fungus. The bacterial community found in this system is believed to protect the fungus from parasites and aid in cellulose digestion. The metagenome of the fungus garden, the genome of the leafcutter ant, and the genome of the fungus have been sequenced by the DOE Joint Genome Institute and the proteomics capabilities at PNNL will be used to identify protein expression within this system. Following high sensitivity and mass accuracy capillary LC-MS/MS measurements of trypsin-digested proteomes, uninterpreted tandem MS spectra will be compared to potential bacterial, fungus and ant protein sequences using Sequest. We believe our metaproteomic analyses could identify novel proteins with applications toward biofuel and antibiotic development.

This research is supported by the Office of Biological and Environmental Research of the U.S. Department of Energy. Portions of this research were performed in the Environmental Molecular Sciences Laboratory, a national scientific user facility sponsored by the DOE's Office of Biological and Environmental Research. Pacific Northwest National Laboratory is operated for the U.S. Department of Energy by Battelle Memorial Institute through Contract No. DE-AC05-76RLO 1830.

21

Production of Lower Viscosity Oils as Biofuels in Transgenic Plants: Deep Transcriptional Profiling Reveals a Novel Acetyl-CoA Diacylglycerol Acetyltransferase from *Euonymus alatus*

John Ohlrogge^{1*} (Ohlrogge@msu.edu), Timothy Durrett,¹ Adrian Troncosco,¹ Jillian Fan,¹ Xia Cao,¹ Nick Thrower,¹ Erika Lindquist,² Christa Pennacchio,² Curtis Wilkerson,¹ and Mike Pollard¹

¹Dept. of Plant Biology and DOE Great Lakes Bioenergy Research Center, Michigan State University, East Lansing and ²DOE Joint Genome Institute, Walnut Creek, Calif.

<http://www.greatlakesbioenergy.org/>

Project Goals: Design of improved plant oil structure for fuel use. The high viscosity of plant oils causes problems with standard diesel engines. Therefore, most biodiesel is produced by conversion of triacylglycerol to methyl or ethyl esters. Acetyl-glycerides are abundant in some plant species and represent a form of triacylglycerol with acetyl rather than acyl groups at sn-3. This structure results in oil with predicted lower viscosity that should directly replace diesel fuel #4 without need for transesterification. The goal of this study is to identify genes involved in biosynthesis of this valuable type of plant oil.

Unlike microarrays, EST sequencing provides a method of transcript analysis that allows quantitative comparisons between genes and between different plant species. In order to identify genes involved in plant oil biosynthesis and the transcription factors and other regulatory systems that control oil accumulation, Michigan State, together with JGI has sequenced over 10 million ESTs from a variety of oil-seeds and other oil rich tissues. Why do we need millions of ESTs? Key enzymes of lipid metabolism (e.g. acyltransferases, phospholipases, thioesterases) are very low abundance and can be difficult to detect by conventional EST sequencing. Deep EST sequencing using 454 pyrosequencing provides a large increase in EST sequence information and allows us to accurately quantify low level expression. By sequencing libraries from multiple species we obtain information on what similarities and differences distinguish oil synthesis in seeds producing unusual fatty acids and in seeds compared to other tissues such as mesocarp that produce high oil levels. Replicate analysis of samples (including cDNA synthesis and PCR) gave a 0.99 correlation coefficient between #

reads per gene. Therefore, 454 sequencing is technically and biologically reproducible and provides an accurate measure of gene expression. We have observed that core enzymes of fatty acid biosynthesis are, in general, expressed in consistent stoichiometric ratios in a number of different oilseeds and tissues. Therefore, those genes that fall outside the usual stoichiometry offer insight into unique metabolism. For example, we observe very low expression of the FatB thioesterase that controls saturated fatty acid production in castor, which agrees with the fact that castor is an oilseed with extremely low saturated fatty acid content.

Endosperm tissue from *Euonymus alatus* (Burning Bush) accumulates high levels of 3-acetyl-1,2-diacyl-*sn*-glycerols (ac-TAGs) as the major storage lipids. Ac-TAGs are unusual triacylglycerols (TAGs) with an *sn*-3 acetate group instead of a long-chain fatty acid and have added value applications in direct use as biodiesel and lubricant oil feedstocks. In addition to producing ac-TAGs, *Euonymus* fruit also synthesizes normal, long-chain TAGs (lc-TAGs) in their aril tissue. The close developmental coordination and spatial proximity of two tissues with the ability to produce different TAGs presents a unique opportunity to understand the accumulation of unusual TAGs in plants. By sequencing ESTs from these tissues we have identified candidate genes involved in ac-TAG biosynthesis. One such candidate, subsequently named EaDacT (*Euonymus alatus* diacylglycerol acetyl-transferase) was highly expressed in the endosperm and absent from the aril. Expression of EaDacT in yeast resulted in the accumulation of ac-TAGs, but not lc-TAGs. *In vitro* assays with microsomes from yeast expressing EaDacT demonstrated that the enzyme possesses acetyl-CoA diacylglycerol acetyltransferase activity, but not long chain acyl-CoA diacylglycerol acyltransferase activity. Expression of EaDacT in *Arabidopsis* seed caused the accumulation of up approximately 40% of ac-TAGs in the seed oil. These results demonstrate that EaDacT synthesizes ac-TAGs in *Euonymus* endosperm and illustrate the utility of deep transcriptional profiling as a gene discovery platform for modifying the seed oil properties of plants.

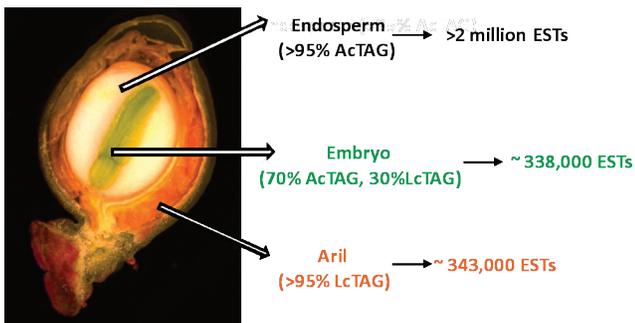


Fig 1. *Euonymus* fruit produce ac-TAGs in a tissue specific manner. The endosperm and embryo tissues of the *Euonymus* seed accumulate high levels of ac-TAGs whereas the aril tissue surrounding the seed synthesizes only lc-TAGs. Transcript profiles were obtained of these different tissues to isolate the enzyme(s) responsible for the synthesis of ac-TAGs.

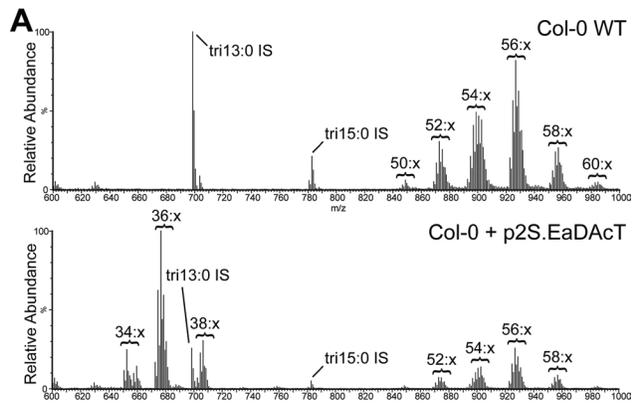


Fig 2. *Arabidopsis* seeds expressing EaDacT accumulate ac-TAGs. Positive-ion ESI mass spectra of neutral lipid extracts from Col-0 wildtype seed or T₃ seed from a representative Col-0 plant expressing EaDacT. Peaks correspond to *m/z* values of the [M + NH₄]⁺ adduct. Tritridecanoin (tri13:0) and tripentadecanoin (tri15:0) were added as internal standards. The number of acyl carbons in each series of TAG molecules is indicated.

22 Sustainable Production of Fatty Acid Derived Fuels and Chemicals in Engineered Microorganisms

Rebecca M. Lennen,^{1,2} Spencer Hoover,^{1,2} Tyler Youngquist,^{1,2} and Brian F. Pfleger^{1,2*} (pfleger@engr.wisc.edu)

¹Dept. of Chemical and Biological Engineering, University of Wisconsin, Madison; ²U.S. Department of Energy Great Lakes Bioenergy Research Center, Madison, Wisconsin

Project Goals: (1) Develop and demonstrate metabolic engineering strategies for assembling fatty acid overproducing microorganisms. (2) Identify barriers that limit the production of fatty acids using processed biomass as a substrate. (3) Develop strategies to convert fatty acids into useful products, including fuels and chemicals.

The development of renewable alternatives to diesel and jet fuels is highly desirable for the heavy transportation sector, and would offer benefits over the production and use of short-chain alcohols for personal transportation. Here we report the development of a metabolically engineered strain of *Escherichia coli* that overproduces medium-chain length fatty acids via three basic modifications: elimination of β-oxidation, overexpression of the four subunits of acetyl-CoA carboxylase, and expression of a plant acyl-acyl carrier protein (ACP) thioesterase from *Umbellularia californica* (BTE). The expression level of BTE was optimized by comparing fatty acid production from strains harboring BTE on plasmids with four different copy numbers. Expression of BTE from low copy number plasmids resulted in the highest fatty acid production. Up to a seven-fold increase in total

fatty acid production was observed in engineered strains over a negative control strain (lacking β -oxidation), with a composition dominated by C12 and C14 saturated and unsaturated fatty acids. Next, a strategy for producing undecane via a combination of biotechnology and heterogeneous catalysis is demonstrated. Fatty acids were extracted from a culture of an overproducing strain into an alkane phase and fed to a Pd/C plug flow reactor, where the extracted fatty acids were decarboxylated to saturated alkanes. The result is an enriched alkane stream that can be recycled for continuous extractions. Complete conversion of C12 fatty acids extracted from culture to alkanes has been demonstrated yielding a concentration of 0.44 g L⁻¹ (culture volume) undecane.

23

Sustainably Filling the Field to Fuel Pipeline: A GLBRC Research Priority

P. Jasrotia,¹ G.P. Robertson¹ (robertson@kbs.msu.edu), J.M. Tiedje^{2*}, K.D. Thelen,² S.K. Hamilton,¹ R. Jackson,³ D.A. Landis,² D.W. Schemske,² B.E. Dale,² S.M. Swinton,² R.C. Izaurralde,⁴ and P.J. Meier³

¹W.K. Kellogg Biological Station (KBS), Hickory Corners, Michigan; ²Michigan State University, East Lansing; ³University of Wisconsin, Madison; and ⁴University of Maryland, College Park

Project Goals: GLBRC research to improve biofuels sustainability overall objective: Support the biomass-to-bioenergy pipeline by developing ecological, agricultural, and life cycle practices that are economically viable and environmentally responsive. Evaluate and improve for different biofuel crops. 1. carbon neutrality and net greenhouse gas mitigation across the entire biofuel life cycle at multiple scales. 2. ecosystem services in biofuel landscapes (e.g., water quality, biodiversity, pest suppression).

Biofuels are attractive for economic, environmental, and strategic reasons. Reducing our dependency on foreign oil is a key national security issue. Although corn and soybeans are excellent sources of biofuels, cellulosic feedstocks clearly are the wave of the future. Poplar, switchgrass, *Miscanthus*, and even managed prairie ecosystems are potential sources of cellulosic biofuels. If we can solve the problems of effectively transforming this biomass into cellulosic ethanol or other fuels, there will be significant environmental positive gains in reducing emissions of greenhouse gases and other pollutants. However, concerns have been raised about the sustainability of cellulosic cropping systems. Biofuel production systems based on annual grains are the most straightforward for growers but may not be the most productive or sustainable in the long term. Understanding the basis for sustainable biofuel production systems is crucial for the long-term success of these systems. The rapid growth, low mineral content and high biomass yield of cellulosic crops make them a favored feedstock choice. But the important question remains: can we produce enough biomass to sustain

the biofuel industry without compromising environmental security. The biomass production potential of these crops is directly linked to concerns related to land-use change, environmental degradation, and food security. To answer these questions, field experiments are underway at several scales in Michigan and Wisconsin. Eight model cropping systems are studied for productivity, carbon balance, and biodiversity impacts. A major goal of this initiative is to test and develop biofuel crops that provide both high yields and environmental benefits such as greenhouse gas mitigation, clean water, and pest protection. Our aim is to provide a comprehensive portrait of the production potential of different cellulosic biofuel crops as well as their environmental impacts and benefits. We use field results to parameterize and test quantitative models that can then be used to simulate crop growth across larger regions.

24

Metagenomics of Bacterial Communities from the Rhizosphere of Switchgrass

Ederson da Conceição Jesus, Endang Susilawati, Benli Chai, Qiong Wang, Jorge Rodrigues, and James M. Tiedje* (tiedje@msu.edu)

Michigan State University, East Lansing

Project Goals: The specific goal of our group is to assess the structure and functional diversity of the rhizosphere community associated with biofuel energy crops by applying metagenomic approaches.

Switchgrass is one of the main plants being studied for biofuel production in the United States. Besides being indigenous and producing a large amount of biomass, it is able to grow in marginal lands, a feature that may be associated not only with plant genotype, but also with their association to beneficial microbes. Managing these microbes can help to improve biomass production and decrease production costs. For this reason, we are studying bacterial communities in the rhizosphere of switchgrass through a metagenomic approach, which allows us to assess not only the taxonomic composition of the bacterial communities in the rhizosphere, but also their functional composition. Additionally, we aimed to validate the application of metagenomics to study the rhizosphere of switchgrass. Roots of the cultivar Cave-in-Rock grown at the MSU experimental farm were sampled in 2008 from the 0-20 cm depth. After removing the excess soil, the rhizosphere soil was recovered and used for total DNA extraction with the method developed by Zhou et al. (1996). The DNA was analyzed by 454-Titanium pyrosequencing at JGI, generating 291 Mb of sequence in 642441 reads with an average length of 453 bp. Potential artificial duplicates were removed (Gomez-Alvarez et al., 2009) and the assembled data set was used for analysis with the MG-RAST pipeline. The data set without duplicates contains 487,660 contigs totaling 237,422,518 basepairs with an average fragment length of 486.86 bp. A total of 297,143 sequences (60.93%) could be

matched to proteins in SEED subsystems (using an e-value cut-off of $1e^{-5}$). Most of the reads were assigned to essential functions for cell maintenance such as amino acids and carbohydrate metabolism. The reads were assigned mainly to Proteobacteria (74%), followed by Acidobacteria (4%), Actinobacteria (3%) and Bacteroidetes (2%). Pseudomonadaceae assignments accounted for 54% of all reads. In fact, 140,697 fragments of the metagenome map to 5,233 of 5,858 features from the *Pseudomonas fluorescens* PfO-1 genome and the total base pair length of sequences matching this genome results in approximately 2.6X coverage. One lane of Illumina sequence has also recently been provided by JGI for the same DNA sample, and is being used to evaluate both the enhancements by this technology, as well as deeper insight into the gene biology of the rhizosphere. Functions commonly associated with *Pseudomonas* such as stress resistance and iron scavenging were also found. *Pseudomonas* was also found in large numbers by cultivation from the same rhizosphere sample, providing cultured models for further study. We conclude that potentially beneficial microbes, such as *Pseudomonas* are present in the rhizosphere of switchgrass and that the metagenomic approach is allowing us to obtain information about the functionality of rhizosphere communities.

25

Education and Outreach Opportunities Linked with the Research and Development of Sustainable Cellulosic Biofuels

John M. Greenler^{1*} (jgreenler@glbrc.wisc.edu), Sara J. Krauskopf,¹ Steven C. Slater,¹ Timothy J. Donohue,^{1,2} Richard M. Amasino^{1,3}

¹U.S. Department of Energy (DOE) Great Lakes Bioenergy Research Center (GLBRC), ²Dept. of Bacteriology, and ³Dept. of Biochemistry, University of Wisconsin, Madison

<http://www.greatlakesbioenergy.org/education/>

Project Goals: Develop education modules and outreach programs for K-16 communities on energy, carbon chemistry, sustainability (environmental, economic and social) issues and other areas related to biofuel production and use; offer bioenergy-focused programming for students and educators; develop informative materials to raise awareness of biofuels and related sustainability issues among the general public, extension staff, policy makers and industry representatives; provide a venue to collaborate on the development of bioenergy education and outreach materials; and present general information about the GLBRC's mission and accomplishments to interested parties.

There is currently a significant research effort to develop sustainable biofuels from cellulosic plant materials. To be broadly sustainable, this research utilizes a diverse array of modern scientific methods, and is very interdisciplinary

and collaborative in nature. Given that this work is in the public eye, there are significant opportunities to engage learners in both the details of bioenergy; and importantly, into the underlying scientific principles of biogeochemical and energy systems. Assessments of student understanding, however, reveal many consistent misconceptions that hinder students' ability to comprehend these systems.

A significant focus of our work is to develop 1) a further categorization of the range of understandings related to carbon cycles and energy flow, and 2) K-16 educational materials that will use biofuels as an entrée to engage learners in a more accurate comprehension of these basic scientific concepts. Specific content at the base of these materials includes life-cycle assessments and systems thinking. The effects of biofuel production on biodiversity, ecosystem services, climate change, and global energy dynamics are considered as well. A number of persistent difficulties are seen in assessments of student understanding at all levels. Many students, for instance, do not connect the decomposition of plants and animals to CO₂ in the atmosphere. In response to this and related patterns, we are designing a series of activities to make connections to fundamental biological concepts with which they are more familiar. Associated with the development of these educational materials are summer research experiences for undergraduates and teachers.

submitted post-press

Functional Annotation of *Fibrobacter succinogenes* Carbohydrate Active Enzymes

David Mead (dmead@lucigen.com), Julie Boyum, Colleen Drinkwater, Krishne Gowda, David Stevenson, Paul Weimer, and Phil Brumm

Lucigen and C56 Technologies, Middleton, Wisc.; Great Lakes Bioenergy Research Center, University of Wisconsin, Madison; and USDA-Agricultural Research Service, Madison, Wisc.

Project Goals: Eliminate bottlenecks in plant cell wall deconstruction.

Fibrobacter succinogenes is a predominant cellulolytic bacterium that degrades plant cell wall biomass in ruminant animals, and is among the most rapidly fibrolytic of all mesophilic bacteria. This gram-negative, strictly anaerobic bacterium is also of interest because it does not utilize either freely secreted enzymes or cellulosomal-type structures to digest cellulose. A dozen cellulolytic enzymes have been expressed and characterized previously, and an outdated partial genome sequence indicates that there are at least 33 unique glycosyl hydrolases encoded by *F. succinogenes*. In order to better understand plant cell wall degradation we have developed new tools to capture, express and identify many of the carbohydrate active enzymes (CAZymes) from this microbe. The complete genome sequence of *Fsu* was finished by the DOE Joint Genome Institute in late 2009, contributing to the growing database of cellulolytic

microbes. Preliminary analysis indicates that *F. succinogenes* contains ~ 133 glycosyl hydrolase and 63 CBM-containing genes, the most of any microbe when expressed as a percent of the total gene number.

Based on the genomic sequencing results, the number of *F. succinogenes* genes annotated as CAZymes far exceed those that have been experimentally determined by conventional enzymatic approaches. One of the goals of this work is to functionally characterize all the putative glycosyl hydrolase genes from *Fsu*, as bioinformatic analysis is an inadequate proxy for actual activity results. Before the genome sequence was available we developed a robust method to enzymatically capture functionally active CAZymes in *E. coli*. Using new expression tools developed at Lucigen and C5-6 Technologies and a multi-substrate screen for xylosidase, xylanase, β -glucosidase and cellulase activities, we generated and screened 5760 random shotgun expression clones for these activities. This represents ~ 2 X genome expression coverage. 169 positive hits were recorded and 33 were unambiguously identified by sequence analysis of the inserts. Eliminating duplicates, 24 unique CAZyme genes were found by functional screening, or 40% of the ~60 genes present in this genome potentially detectable by the multiplex assay. Several previously uncharacterized enzymes were discovered using this approach. With the full genome sequence available we will attempt to express and characterize all of the recognizable CAZymes, as well as the CBM-containing genes for actual enzyme activity. The active enzymes will also be sent to other partners in the GLBRC to assess their ability to deconstruct plant biomass.

Joint BioEnergy Institute (JBEI)

26

The Joint BioEnergy Institute: Addressing the Challenges of Converting Biomass to Fuels

J.D. Keasling* (JDKeasling@lbl.gov), H.W. Blanch, P.D. Adams, P.C. Ronald, and B.A. Simmons

Joint BioEnergy Institute, Emeryville, Calif.

Project Goals (Abstracts 26-54): In the San Francisco Bay Area, three national laboratories, major public and private universities, industry, and federal agencies have joined together to create the Joint BioEnergy Institute (JBEI). This institute will develop the basic science and technology to create an array of environmentally friendly biofuels using plant biomass and microbes. JBEI will focus its scientific effort in three key areas: feedstock production, deconstruction, and fuels synthesis. JBEI will employ an opportunistic "start-up company" approach, partnering with industry, to develop new science and technologies that address the most challenging steps in industrial

bioenergy processing. Crosscutting technologies in computational tools, systems and synthetic biology tools, and advanced imaging will be applied in a multi-pronged approach for biomass-to-biofuel solutions in addition to discovery-driven benefits for biohydrogen research, solar-to-fuel initiatives, and broader DOE programs.

In the San Francisco Bay Area, three national laboratories, major public and private universities, industry, and federal agencies have joined together to create the Joint BioEnergy Institute (JBEI). This institute is designed to address the mission of the DOE Bioenergy Research Center program: "to produce fundamental scientific discoveries and major technological advances to enable the development of cost-effective, energy-efficient, and commercially viable processes for large-scale conversion of lignocellulosic biomass into fuels." This institute is developing the basic science and technology to produce fuels from plant biomass by microbial routes. JBEI is focusing its scientific effort in three key areas: elucidating cell wall biosynthesis, lignocellulose deconstruction, and fuels biosynthesis, while employing cross-cutting technologies throughout all of its research.

There are key challenges in each of the institute's scientific divisions. The Feedstocks Division is developing an understanding of hemicellulose biosynthesis and analyzing the recalcitrance of plant cell walls to deconstruction. By modifying lignin and reducing acetate and ferulate content, crops will be better suited to biofuels production. In the Deconstruction Division, new approaches to biomass pretreatment are being developed that result in solubilization and separation of plant cell wall components. Enzymatic hydrolysis of pretreated biomass represents a significant cost in formation of saccharide monomers; efforts are underway to identify and produce new enzymes from unique environments that are more effective.

In the Fuels Synthesis, Division, a challenge is to produce advanced biofuels needed for diesel and jet engines. Many fuel-producing organisms incompletely metabolize C₅ (xylose, arabinose) sugars, while inhibitors (e.g., acetate) released from biomass pretreatment can limit fuels production. Introducing new metabolic routes to advanced biofuels relies on Synthetic Biology approaches. Few tools are available for biomass conversion research. The Technologies Division is developing high-throughput 'omics and microfluidic approaches for many aspects of bioenergy research, as well as high-throughput biochemical, spectroscopic, and imaging methods for the rapid characterization of biomass.

This work was part of the DOE Joint BioEnergy Institute (<http://www.jbei.org>) supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U.S. Department of Energy.

27

Exploiting Natural Variation in *Arabidopsis thaliana* to Understand Cell Wall Biosynthesis and Composition

Katy M. Christiansen* (KMChristiansen@lbl.gov),
A. Michelle Smith, Dominique Loque, and **Joshua L. Heazlewood**

Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Berkeley, Calif.

Project Goals: See goals for Abstract 26.

Genetic variation in different accessions of *Arabidopsis thaliana* has occurred through thousands of years of geographic spread and adaptation. As a result of this diversity, disparate accessions have phenotypic differences that can be used to identify genes that contribute to the biosynthesis and composition of cell walls. Forward and reverse genetic screens to identify single mutants often prove difficult for detection of genetic differences that may lead to more subtle phenotypes. Exploiting the inherent genetic variation in *Arabidopsis* accessions through quantitative trait analysis will allow for the detection of variation in cell wall biosynthesis and composition. We selected two methods for screening parental accessions: measuring monosaccharide composition by HPAEC and structural changes by Near Infrared (NIR) spectroscopy. From these analyses, we selected Ri-0 as the most different accession from the reference accession Col-0. Recombinant inbred lines (RILs) derived from Col-0 and Ri-0 parents were used to determine quantitative trait loci (QTL) that contribute to the differences observed in monosaccharide and pectin content and NIR spectroscopy. Two putative candidate genes encoding enzymes involved in nucleotide sugar conversion have been identified. We are currently determining differences between the parental versions of the genes and how any differences contribute to the observed phenotypes. Additionally, we have re-sequenced two *Arabidopsis* accessions, Bay-0 and Shahdara, in collaboration with the Joint Genome Institute. This effort has resulted in a collection of SNPs between these accessions and the reference *Arabidopsis* accession, Col-0. Resequencing efforts are being expanded to other accessions, beginning with Ri-0. Information obtained from resequencing will aid in QTL analysis and be of service to the general *Arabidopsis* community.

This work was part of the DOE Joint BioEnergy Institute (<http://www.jbei.org>) supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U.S. Department of Energy.

28

Towards the Plant Golgi Proteome

Harriet T. Parsons,¹ Christopher J. Petzold,¹ Alyssa Redding-Johanson,¹ Tanveer Batth,¹ and **Joshua L. Heazlewood*** (JLHeazlewood@lbl.gov)

¹Joint BioEnergy Institute, Emeryville, Calif.

Project Goals: See goals for Abstract 26.

The plant cell wall is comprised of complex sugar polymers including cellulose, hemicellulose and pectin. The Golgi apparatus within the plant cell produces a large proportion of these polysaccharides prior to their incorporation into the cell wall. The details of how these complex sugars are synthesized and delivered are currently poorly understood. Our current knowledge of the protein constituent that comprise the plant Golgi is relatively poor when compared to other subcellular components within the cell. In order to further understand the role of this organelle in cell wall biosynthesis we are characterizing this subcellular compartment using proteomics. We are employing an orthogonal approach which utilizes density centrifugation followed by charge based separation of the organelle on a Free Flow Electrophoresis system. Analysis of Golgi purified fractions from *Arabidopsis* cell culture by mass spectrometry after FFE separation indicates the method is suitable for isolation of this organelle from plants. We have identified 300 to 400 proteins from these fractions and found approximately 50 glycosyl transferases likely involved in matrix polysaccharide biosynthesis. Overall approximately 50% of the proteins in this list are of known or likely Golgi in origin; approximately 35% are unknown or are derived from the endosomal system and 15% appear to be contaminants from other organelles and membranes. Fluorescent protein constructs are being designed to confirm localizations of novel and ambiguous proteins. This technique will enable us to commence in-depth comparative cell wall proteomics focusing on protein function and changes and in the Golgi apparatus.

This work was part of the DOE Joint BioEnergy Institute (<http://www.jbei.org>) supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U.S. Department of Energy.

29

Synthesis of Phenylpropanoid-Esters and -Amides in *Arabidopsis thaliana* to Engineer a Cleavable Lignin

Aymerick Eudes, Prajakta Pradhan, Fan Yang, and
Dominique Loqué* (DLoque@lbl.gov)

Feedstocks Division, Joint BioEnergy Institute,
 Emeryville, Calif.

Project Goals: See goals for Abstract 26.

The development of alternative transportation fuels that can meet future demands while reducing global warming is critical to the national, environmental, and economic security of the United States. Currently, biofuels are produced largely from starch, but there is a large, untapped resource (more than a billion tons per year) of plant biomass that could be utilized as a renewable, domestic source of carbon-neutral, liquid fuels. However, significant roadblocks hamper the development of cost-effective and energy-efficient processes to convert lignocellulose biomass into fuels. Lignin is a very strong phenolic polymer, which embeds cellulose and hemicellulose, and its recalcitrance to chemical and biological degradations inhibits the conversion of cell wall polysaccharides (cellulose and hemicellulose) into fermentable sugars. Unfortunately, lignin provides such compressive resistance to plant cells that it cannot simply be genetically removed without incurring deleterious consequences on plant productivity. Alternative strategies to significantly reduce lignin recalcitrance would be modifying its composition and deposition. We are currently developing an alternative strategy, which is focusing on the partial replacement of the “hard bonds” (e.g. ether, carbon bonds) in the lignin polymer with “easily cleavable” ones (e.g., amide or ester bonds). For this propose, we are rerouting part of the lignin biosynthesis towards the synthesis of phenylpropanoid-derived molecules such as hydroxycinnamic acid amides and esters in order to partially replace conventional lignin monomers in the cell wall. Biosynthetic pathways and preliminary data for de novo synthesis in *Arabidopsis* of selected phenylpropanoid-derived compounds are presented.

This work was part of the DOE Joint BioEnergy Institute (<http://www.jbei.org>) supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U.S. Department of Energy.

30

Elucidating Switchgrass Genome Structure and Function of Cell Wall-Related Enzymes

M.K. Sharma, R. Sharma* (RASharma@lbl.gov),
 L.E. Bartley, M.D. Zorn, P. Cao, E. Marvinney, and
P.C. Ronald

Joint BioEnergy Institute, Lawrence Berkeley National
 Laboratory, Berkeley, Calif.

Project Goals: See goals for Abstract 26.

JBEI, in collaboration with other DOE centers, is organizing and supporting the sequencing the genome of switchgrass, *Panicum virgatum* L. clone AP13. Our goal is to elucidate the genome content to facilitate comparative functional genomics studies of this promising biofuel feedstock. Whole genome shotgun sequencing using 454-Titanium technology is underway at the DOE-Joint Genome Institute (JGI). The first phase of switchgrass sequencing will be used to produce approximately 4x sampling of the AP13 genome in 400 bp 454 reads and approximately 30x sampling in 75 bp Illumina reads. We are also using various approaches that include sequencing of hypomethylated restriction libraries, and BAC (Bacterial Artificial Chromosome) and fosmid end sequencing to efficiently cover gene space and establish long-range connectivity among assembled shotgun sequence contigs, respectively. One AP13 BAC library has been synthesized and characterized for mitochondrial/chloroplast DNA contamination as well as coverage by high-density filter hybridizations. We are using available information about genes associated with cell wall biosynthesis, degradation, biomass production and stress tolerance from other grass species to identify homologous ESTs in switchgrass and screen corresponding BACs for in depth characterization. The first ten BACs have been selected and are being sequenced at JGI. We have also selected BACs containing single copy genes to make direct comparisons between homeologous chromosomes. Due to the high level of colinearity among grass genomes, we can use sorghum and foxtail millet as reference genomes in assembling the switchgrass sequence data. With the completion of whole genome sequence, cell wall related genes will be computationally identified and comprehensive phylogenomic analysis with other grasses and dicots will be carried out to identify grass-specific and switchgrass-specific genes involved in cell wall metabolism. Already, Initial phylogenomic analysis with switchgrass ESTs has shown that the CslG family previously considered as dicot-specific, are present in the switchgrass genome. In coordination with JGI we are developing a high quality annotation pipeline and databases of grass cellwall-osome sequences to support comparative functional genomics in grasses. Our own and others' 454 sequence ESTs (400 bp reads) and approximately 500,000 Sanger ESTs (800 bp reads) in public repositories will assist us in producing gene inventories and gene annotation.

This work was part of the DOE Joint BioEnergy Institute (<http://www.jbei.org>) supported by the U.S. Department of Energy,

Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U.S. Department of Energy.

31

Identification of Genes Involved in Acetylation of Cell Wall Polysaccharides in *Arabidopsis thaliana*

Yuzuki Manabe^{1*} (YManabe@lbl.gov), Yves Verherbruggen,¹ Majse Nafisi,² Chithra Manisseri,¹ Peter Benke,¹ Aindrila Mukhopadhyay,¹ Caroline Orfila,² Yumiko Sakuragi,² and **Henrik Vibe Scheller**¹

¹Joint BioEnergy Institute, Emeryville, Calif. and

²Dept. of Plant Biology and Biotechnology, University of Copenhagen, Denmark

Project Goals: See goals for Abstract 26.

Acetylation of cell wall polysaccharides has long been observed in various plant species; however, the enzymes involved in the acetylation have thus far not been identified. Both pectins and hemicelluloses are acetylated to various degrees. While the *in vivo* role of polysaccharide acetylation is still unclear, it is known to affect biofuel yield from lignocellulosic biomass due to inhibition of enzymatic degradation of the polysaccharides and fermentation by microorganisms. Therefore, a decreased level of acetate esters in lignocellulosic biomass may increase the efficiency of biofuel production. JBEI is therefore investigating the mechanism of polysaccharide acetylation and assessing the possibility of modifying acetylation level *in planta*.

We have analyzed four *Arabidopsis* homologues of a protein known to be involved in polysaccharide acetylation in a fungus. *Arabidopsis* mutants with insertional mutagenesis in the respective genes were identified, and we found that at least one of the mutants, designated *reduced wall acetylation* (*rwa1*, *rwa2*, *rwa3* and *rwa4*) had decreased levels of acetylated cell wall polymers. Two independent alleles of *rwa2* mutants were examined by analyzing alcohol insoluble residues extracted from leaves. Extracts treated with 0.1M NaOH released about 20% lower amounts of acetic acid when compared to wildtype. Interestingly, the monosaccharide composition of the cell wall polysaccharides in *rwa2* was not altered. Current efforts are aimed at determining which acetylated polysaccharides are affected in the *rwa2* mutants. There was no apparent visible difference Scheller observed between wildtype and either allele of mutants at any developmental stages. However, both alleles of *rwa2* have displayed increased resistance toward *Botrytis cinerea*, a necrotrophic fungus. The other mutants, *rwa1*, *rwa3*, and *rwa4*, did not have detectable changes in acetylation, presumably due to genetic redundancy. Double, triple and quadruple mutants are currently being investigated.

This work was part of the DOE Joint BioEnergy Institute (<http://www.jbei.org>) supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U.S. Department of Energy.

32

Forward Genetic Screen to Identify Rice Mutants with Changes in Cell Wall Composition and Saccharification Efficiency

Miguel Vega-Sánchez* (MEVegasanchez@ucdavis.edu), Mawsheng Chern, Wing Hoi Sze-To, Laura Bartley, A. Michelle Smith, Bradley Holmes, Rajiv Bharadwaj, **Blake Simmons**, and **Pamela Ronald**

Feedstocks, Deconstruction, and Technology Divisions, Joint BioEnergy Institute, Emeryville, Calif.

Project Goals: See goals for Abstract 26.

Understanding plant cell wall biosynthesis is crucial for the development of the next generation of biofuels derived from lignocellulosic material. Current limitations in the harvest of fermentable sugars from cellulose derive from the inherent recalcitrance of plant cell walls. Basic knowledge of how the structure and composition of the cell wall can be modified to obtain biomass suitable for efficient and economically viable biofuel production is needed. We are using a forward genetics approach to identify genes responsible for cell wall characteristics affecting cell wall composition and deconstruction. By means of fast neutron mutagenesis, we have generated a rice mutant population consisting of 6,500 M0 lines and harvested more than 100,000 M1 seed from approximately 4,000 M0 plants. Leaves and stems from these lines have been collected and we are in the process of screening them for alterations in saccharification efficiency and cell wall composition. To screen for changes in fermentable sugar release from biomass, we have optimized a protocol using either hot water or dilute acid pre-treatment followed by enzymatic saccharification for adaptation into a 96 well format. In addition, we are standardizing a high throughput microfluidics platform for analyzing alterations in the C5/C6 monosaccharide ratios of total sugar extracts from leaves to identify mutants with changes in cell wall composition. We have also validated a method for pre-screening intact dried leaf tissue using Near Infrared spectroscopy to identify outliers in the mutant population that will be then analyzed using the microfluidics system. Once cell wall mutants are confirmed, we will extract DNA from wild type and highly prioritized mutant candidates and then carry out whole genome comparative hybridization on rice tiling arrays. This approach will allow us to identify genes in deleted region responsible for the mutant phenotypes. Mutants will be complemented with candidate genes using transgenic analysis and assayed for restoration of the cell wall phenotypes.

This work was part of the DOE Joint BioEnergy Institute (<http://www.jbei.org>) supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U.S. Department of Energy.

33

Glycosyltransferases (GTs) from the *Arabidopsis* CAZy Family: High-Throughput Cloning of a Library of GT and GT-Related Genes

Peter McInerney,¹ Stephanie Morrison,¹ Hou Cheng Chu,¹ Ai Oikawa,² Henrik Scheller,² **Pam Ronald,² Paul Adams,¹** and Masood Hadi^{1*} (MZHadi@lbl.gov)

¹Technology and ²Feedstocks Divisions, Joint BioEnergy Institute, Emeryville, Calif.

Project Goals: See goals for Abstract 26.

The CAZy (Carbohydrate Active EnZyme) family of enzymes includes glycosyltransferases (GTs), glycosylhydrolases (GHs), polysaccharide lyases (PLs), carbohydrate esterase (CEs) and carbohydrate binding modules (CBMs). Many enzymes in this family are involved in various aspects of plant cell wall metabolism. The GTs represent one of the most diverse CAZy groups, with 91 separate protein families (not including non-classified sequences) that are assigned based on 3D protein structure, catalytic mechanism and donor/acceptor substrate requirements. In the simplest terms, GTs catalyze the transfer of sugar molecules from a donor molecule to an acceptor. However, the seemingly limitless combination of specific sugar, donor and acceptor molecules underscores the necessity for a large number of enzymes of this family. Our group is undertaking an effort to clone all 455 GTs in the CAZy database from *Arabidopsis thaliana* as well as the 90 GT-like proteins identified from other bioinformatic analyses. This library of GT clones will be a valuable resource at JBEI for a wide range of applications. At the level of biofuels research, results applicable to cell wall engineering are expected, based on the observation that a significant proportion of GT genes play roles (or are proposed to play roles) in cell wall metabolism. Furthermore, from a basic science standpoint, a great deal of new information should result from the study of these genes, since many of these genes are hypothetical or have unknown functions. Our approach relies heavily on automation, for informatics steps such as PCR primer generation and DNA sequence analysis, in addition to laboratory robotics, for assembly of enzymatic reactions and purification steps. To date clones for 80% of the targets have gone through the pipeline and are being sequence verified. Current efforts are centered on maximizing the number of clones that perfectly match the target DNA sequence by optimizing our cloning workflow. Once production of sequence-validated clones is complete, we will transfer genes to vectors suited to specific needs (e.g. expression for biochemical analyses or crystallography trials, fluorescence localization studies, etc.).

This work was part of the DOE Joint BioEnergy Institute (<http://www.jbei.org>) supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U.S. Department of Energy.

34

Screening Glycosyltransferases for Enzymatic Activity

Anongpat Suttangkakul^{1*} (ASuttangkakul@lbl.gov), Ai Oikawa,¹ Carsten Rautengarten,¹ Jesper Harholt,² Yves Verherbruggen,¹ Michelle Truong,¹ Peter McInerney,¹ Masood Hadi,¹ Peter Benke,¹ Naomi Geshi,² Jacob K. Jensen,² Eva Knoch,^{1,2} and **Henrik Vibe Scheller¹**

¹Joint BioEnergy Institute, Emeryville, Calif. and ²Dept. of Plant Biology and Biotechnology, University of Copenhagen, Denmark

Project Goals: See goals for Abstract 26.

Declining sources of fossil fuels, global warming and political instability in oil producing regions have led many countries to develop strategies for alternative energy. Plant biomass is a convenient way to harness solar energy and photosynthesis, and biomass is already an important supplement to fossil fuels. However, the energy efficiency of biofuel production is low, and environmental impact can be high. There is a great need to develop new technologies that can provide fuels, especially liquid fuels for transportation, in an efficient and environmentally friendly way.

Plant cell walls are composed mainly of polysaccharides and production of biofuels from biomass requires decomposition of the polymers. Many of the polymers are recalcitrant to degradation and some degradation products cannot be converted efficiently into fuels or may even be inhibitory. Better understanding of the biosynthesis of the cell wall polysaccharides may enable development of crops with improved properties as biofuels feedstocks. Despite rather detailed information on the structure of the cell wall polysaccharides, little is known about their biosynthesis. The key enzymes are glycosyltransferases (GTs) and plants need a large number of GTs to synthesize the complex polysaccharides present in the walls. In *Arabidopsis*, approximately 450 GT genes have been identified; however, in spite of a significant effort, only few GTs have had their activity determined.

We have expressed many of the *Arabidopsis* GTs in tobacco and *E. coli* and developed assays to determine their activity. The activity of a given GT can be determined by assumption-free assays where a range of different substrates are tested, e.g. combinations of nucleotide sugars and monosaccharide. In other cases, mutant analysis or phylogeny strongly suggests a particular role for a GT, which can then

be tested with more specific assays. Examples of the use of these strategies to determine activity of GTs involved in pectin, hemicellulose, and arabinogalactan biosynthesis will be presented.

This work was part of the DOE Joint BioEnergy Institute (<http://www.jbei.org>) supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U.S. Department of Energy.

35

Microfluidic Technology for Biofuels Applications

Rajiv Bharadwaj* (Rbharadwaj@lbl.gov), Aarthi Chandrasekaran, April Wong, **Blake A. Simmons**, **Paul D. Adams**, and **Anup K. Singh**

Joint BioEnergy Institute, Emeryville, Calif.

Project Goals: See goals for Abstract 26.

We are developing a suite of microfluidic technologies to address the throughput limitations of conventional approaches for biomass analysis. We have developed a microfluidic electrophoretic assay for rapid (< 1 min) and multiplexed analysis of lignocellulosic biomass samples. To address the low throughput of conventional enzyme screening approaches, we have developed a rapid (1-2 hr) integrated microscale platform for cell-free expression and activity analysis of thermophilic cellulases. We have also developed a high-throughput approach for saccharification studies of ionic-liquid pretreated solid biomass. Currently, we are integrating the above technologies to develop optimized cellulase cocktails for cost-effective production of biofuels from lignocellulosic feedstocks.

This work was part of the DOE Joint BioEnergy Institute (<http://www.jbei.org>) supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U.S. Department of Energy.

36

A Microscale Platform for Integrated Cell-Free Expression and Screening of Cellulase Activity

Aarthi Chandrasekaran,^{1*} Rajiv Bharadwaj,¹ Joshua I. Park,² Masood Z. Hadi,¹ Rajat Sapra,² **Blake A. Simmons**² (BASimmons@lbl.gov), **Paul D. Adams**,¹ and **Anup K. Singh**¹

¹Technology Division, ²Deconstruction Division, Joint BioEnergy Institute, Emeryville, Calif.

Project Goals: See goals for Abstract 26.

A critical step in the efficient conversion of lignocellulosic biomass to fuel is the deconstruction of the biomass to fermentable sugars. Several efforts are therefore focused on identification, expression and characterization of novel enzymes that hydrolyze lignocellulosic biomass. High-throughput enzyme assays that enable rapid screening of these enzymes can greatly accelerate the current enzyme engineering efforts for biofuels development. In this study, we have developed a miniaturized high-throughput, fluorescence-based screening platform for rapid activity profiling of thermophilic cellulases at elevated temperatures (>80°C). This platform integrates cell-free expression and functional characterization of the cellulases in microwell arrays with volumes as low as two microliters. Herein, we demonstrate the use of this approach to express and screen a panel of thermophilic β -glucosidases and cellobiohydrolases.

This work was part of the DOE Joint BioEnergy Institute (<http://www.jbei.org>) supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U.S. Department of Energy.

37

Ionic Liquid Pretreatment of Biomass: Dynamic studies with Light Scattering, GC-MS and FTIR

Patanjali Varanasi* (Pvaranasi@lbl.gov), Manfred Auer, **Blake Simmons**, and Seema Singh

Deconstruction and Technology Division, Joint BioEnergy Institute, Emeryville, Calif.

Project Goals: See goals for Abstract 26.

Among the many choices for alternative energy, biofuels promise to have the most potential for clean and renewable energy. Cellulose and hemi-cellulose present in the biomass can be converted to simple sugars through enzymatic hydrolysis and further to advanced biofuels downstream. But lignin present in the biomass hinders the enzyme accessibility to cellulose and thus enzyme efficiency and total

biofuel yield. At JBEI, research efforts are focused on Ionic liquid (IL) pretreatment to overcome biomass recalcitrance. However, biomass degradation during IL pretreatment and its effect on microbial growth is not understood. Hence, for efficient biofuel production it is important to fundamentally understand the deconstruction of biomass and the compounds produced from biomass de-polymerization during IL pretreatment process for process optimization. To understand the depolymerization of biomass, avicel (model cellulose), model lignin monomers, dimmers and lignin polymer (kraft lignin and low sulfatealkali lignin), and different biomass (switchgrass, pine and eucalyptus) were pretreated with IL at 120°C and 160°C for different time periods (1, 3, 6, 12h). The resultant compounds from the pretreatment process were analyzed using FTIR and GC-MS. Light scattering was used to find if the pretreatment caused lignin dissolution or is effective in any depolymerization of lignin. The extent of lignin depolymerization was found to be temperature dependent. The results from these experiments show that treating biomass at 160°C for 12 hrs may be the best route to degrade biomass if depolymerization is desired.

This work was part of the DOE Joint BioEnergy Institute (<http://www.jbei.org>) supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U.S. Department of Energy.

38

Understanding Ionic Liquid Pretreatment of Lignocellulosic Biomass by Hyperspectral Raman Imaging

Lan Sun* (LanSun@lbl.gov), Chenlin Li, **Blake A. Simmons**, and Seema Singh

Pretreatment Group, Deconstruction Division, Joint BioEnergy Institute, Emeryville, Calif.

Project Goals: See goals for Abstract 26.

Pretreatment of lignocellulosic biomass is essential for breaking down the highly interwoven matrix of cellulose, lignin and hemicellulose, which facilitates enzyme accessibility and adsorption to cellulose for efficient saccharification. Ionic liquids are a new class of non-volatile solvents exhibiting excellent solvating properties and have shown great promise for lignocellulosic biomass pretreatment with easy recovery of cellulose by rapid precipitation with anti-solvents. Ionic liquids have been demonstrated to be very effective in cellulose solubilization in bulk, and have shown to swell cell walls perhaps by breaking inter and intra chain hydrogen bonding. However, to date, molecular level understanding of ionic liquid pretreatment on lignin and its impact on different tissue and cell types of biomasses is lacking. The aim of this research is to develop a fundamental understanding of ionic liquid pretreatment by monitoring the compositional changes during the pretreatment

process. Raman microscopy based on molecular vibrational spectroscopy is a label-free imaging technique capable of real-time and noninvasive examination of plant cell walls with chemical selectivity. In this research, we employed hyperspectral Raman imaging to study tissue and cell type specific distribution of cell wall components and the impact of ionic liquid pretreatment on various cell types of corn stover to identify signatures for predicting deconstruction-ability and understand pretreatment dynamics. The Raman mapping results have shown that the distribution of lignin and cellulose varies significantly across different tissue and cell types in the following order: sclerenchyma cells and tracheids > epidermal cells > bundle sheath cells > parenchyma cells. Lignin content decreases rapidly in tracheids and sclerenchyma cells and slowly in parenchyma cells during ionic liquid pre treatment. Significant cell wall swelling of various cell types during ionic liquid pretreatment was revealed by confocal fluorescence microscopy.

This work was part of the DOE Joint BioEnergy Institute (<http://www.jbei.org>) supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U.S. Department of Energy.

39

Characterization of a Hyperthermophilic Cellobiohydrolase from *Caldicellulosiruptor saccharolyticus*: Enzymatic Hydrolysis of Cellulose Mediated by Substrate Binding

Joshua I. Park^{1,2*} (JIPark@lbl.gov), Michael S. Kent,^{1,3} Supratim Datta,^{1,2} Bradley M. Holmes,^{1,2} Dean C. Dibble,^{1,2} Zhaohua Huang,⁴ **Blake A. Simmons**,^{1,2} and Rajat Sapra^{1,2}

¹Deconstruction Division, Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Emeryville, California; ²Dept. of Biomass Science and Conversion Technology, Sandia National Laboratories, Livermore, Calif.; ³Biodefense and Materials Science Center, Sandia National Laboratories, Albuquerque, N.M.; ⁴Depts. of Bioengineering, Therapeutic Sciences, and Pharmaceutical Chemistry, University of California, San Francisco

Project Goals: See goals for Abstract 26.

We cloned, expressed, purified, and characterized a recombinant cellobiohydrolase (EC 3.2.1.91) domain from celB, a modular cellulolytic gene from *Caldicellulosiruptor saccharolyticus* that contains glycoside hydrolase family 10 (GH10), carbohydrate binding module family 3 (CBM3), and GH5 domains. The deletion analysis of *celB* confirmed that the constructs containing the GH5 domain were able to hydrolyze the soluble substrates carboxymethyl-cellulose (CMC) and *p*-nitrophenyl-β-D-cellobioside (*p*NPC). Therefore, we focused our study on the recombinant CBM3-GH5 and GH5. The recombinant proteins were expressed in *E. coli*,

and purified to homogeneity by affinity and ion-exchange chromatography methods. The functional stability and melting temperature measurements demonstrated that both CBM3-GH5 and GH5 are highly stable up to 80°C at pH 5.5. CBM3-GH5 and GH5 were also able to hydrolyze microcrystalline cellulose (Avicel), ionic liquid (IL)-pretreated cellulose, and IL-pretreated corn stover to produce cellobiose; CBM3-GH5 produced more cellobiose than GH5 did from these insoluble substrates. We employed fluorescence confocal microscopy and total internal reflective fluorescence (TIRF) methods to investigate whether the binding interaction between the enzyme and substrate was attributed to the product yield from the insoluble substrates. We observed stronger binding interaction between CBM3-GH5 and cellulose (both microcrystalline and amorphous) than that between GH5 and cellulose. Thus, the higher product yields from the enzymatic hydrolysis of microcrystalline cellulose, IL-pretreated cellulose, and IL-pretreated corn stover by CBM3-GH5 were possibly mediated by the interaction between the CBM3 domain and the substrates. The recombinant CBM3-GH5 is a thermostable and active cellobiohydrolase that could be used with other types of cellulolytic enzymes for degradation of IL-pretreated biomass to produce fermentable sugars.

This work was part of the DOE Joint BioEnergy Institute (<http://www.jbei.org>) supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U.S. Department of Energy.

40 Targeted Enzyme Discovery in Feedstock-Adapted Microbial Communities

John M. Gladden* (JMGladden@lbl.gov), Martin Allgaier, Amitha M. Reddy, Jean S. VanderGheynst, Terry C. Hazen, **Blake A. Simmons**, Phillip Hugenholtz, and **Steven W. Singer**

Microbial Communities Group, Deconstruction Division, Joint BioEnergy Institute, Emeryville, Calif.

Project Goals: See goals for Abstract 26.

Efficient saccharification of pre-treated feedstocks is essential to produce inexpensive biofuels derived from biomass. The enzyme cocktails used for this step need to be robust and able to withstand inhibitors produced during pretreatment. To identify enzymes suitable for saccharification on an industrial scale, we have focused on adapting microbial communities to specific feedstocks and pretreatment conditions (i.e. high temperatures, ionic liquid contamination, etc.), followed by functional characterization of secreted glycosyl hydrolases. To select for thermophilic enzymes, microbial communities derived from compost were adapted to biomass feedstocks at 60°C. Phylogenetic profiling of these communities show that each consists of a simple mixed con-

sortia with just two or three species in high abundance. The secretomes obtained from these consortia have biomass-deconstructing enzymatic activity that is both thermostable and active in high concentrations of ionic liquid, two highly desirable characteristics for industrial enzymes. Zymography was used to investigate the complement of glycosyl hydrolase enzymes expressed by the consortia, detecting at least a dozen active enzymes for multiple polysaccharide substrates. Currently, proteins within individual zymogram bands are being studied by MS-based proteomics. Candidate glycosyl hydrolases will be identified by comparing measured peptide masses to predicted protein sequences from Carbohydrate Active enZYme (CAZy) database and genome sequences of reference organisms related to consortial members. Future work will utilize metagenomic and single-cell genomic sequencing to document comprehensively the glycosyl hydrolases secreted by these highly active microbial communities. Cataloguing the glycosyl hydrolases in these secretomes will enable us to design thermophilic enzyme cocktails for biomass deconstruction that function under the conditions required for industrial conversion of biomass to biofuels.

This work was part of the DOE Joint BioEnergy Institute (<http://www.jbei.org>) supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U.S. Department of Energy.

41 Ionic Liquid Cation Influence on the Dissolution of Isolated Lignins and Biomass for Ethanol Production

Anthe George^{1*} (AGeorge@lbl.gov), Peter Benke,² Kim Tran,¹ **Blake Simmons**,¹ and Bradley Holmes¹

¹Pretreatment Group, Deconstruction Division and ²Technology Division, Joint BioEnergy Institute, Emeryville, Calif.

Project Goals: See goals for Abstract 26.

For economically viable fermentation-based lignocellulosic biofuel production, pretreatment of the biomass is required. Ionic liquids are proving a promising pretreatment medium. However, debate exists as to the extent of biomass dissolution by ionic liquids and the mechanisms by which this solubility occurs. In this study the degree of biomass dissolution as a function of ionic liquid cation was investigated. Four chloride anion based ionic liquids were studied for their ability to dissolve either organosolv, klason, indulin AT, and milled-wood lignins along with eucalyptus and pine biomass. Imidazolium, phosphonium, ammonium and pyridinium anions were screened. Size exclusion chromatography and MALDI-TOF methods were developed to study the degree of dissolution and polydispersity as a function of solvent cation. The comparison between isolated lignins

and biomasses deduced not only the degree of dissolution, but mechanistic information on the component of biomass dissolved by ionic liquids.

This work was part of the DOE Joint BioEnergy Institute (<http://www.jbei.org>) supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U.S. Department of Energy.

42

Trapping Lignin Degrading Microbes in Tropical Forest Soil

Kristen M. DeAngelis^{1,2*} (KMDeangelis@lbl.gov), Martin Allgaier,^{2,3} Whendee L. Silver,^{1,4} Yaucin Chavarria,¹ Julian Fortney,¹ Phillip Hugenholtz,^{1,3} **Blake Simmons**,^{2,5} Kerry Sublette,⁶ and Terry C. Hazen^{1,2}

¹Ecology Dept., Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, Calif.; ²Microbial Communities Group, Deconstruction Division, Joint BioEnergy Institute, Emeryville Calif.; ³DOE Joint Genome Institute, Walnut Creek, Calif.; ⁴Dept. of Environmental Science, Policy and Management, University of California, Berkeley; ⁵Sandia National Laboratories; and ⁶University of Tennessee

Project Goals: See goals for Abstract 26.

Lignin in feedstock plant material represents a barrier to more efficient plant-to-biomass conversion and can also hinder enzymatic access to cellulose. For this reason, it is critical to develop a way to degrade recovered lignin for next generation feedstock-derived biofuels. While the best-known ligninases are fungal, bacteria are more amenable to emerging cellulosic biofuels technologies. Tropical rain forest soils in Puerto Rico are likely dominated by bacterial decomposers because of the frequent anoxic conditions and fluctuating redox characteristic of these soils, so we focused here to search for novel bacterial lignase producers. To do this, we buried bug traps containing lignin-amended and unamended biosep beads in the soil and incubated them for 1, 4, 13 and 30 weeks. At each time point, phenol oxidase and peroxidase enzyme activity was found to be elevated in the lignin-amended versus the unamended beads, while cellulolytic enzyme activities were significantly depressed in lignin-amended beads. Quantitative PCR of bacterial communities showed more colonization in the lignin-amended compared to the unamended beads after one and four weeks, which attenuated over the course of the incubation. The microbial community was analyzed by microarray (PhyloChip) and by pyrotag sequencing of the community 16S ribosomal RNA genes. Community trends were strongly driven by time but also lignin-amendment to the beads. These techniques also allow us to identify which taxa were increased in lignin-amended compared to unamended beads,

which included representatives from the phyla Actinobacteria, Firmicutes, Acidobacterial and Proteobacteria.

This work was part of the DOE Joint BioEnergy Institute (<http://www.jbei.org>) supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U.S. Department of Energy.

43

High-Resolution Electron Microscopy Imaging of Plants and Pretreated Biomass

B. Knierim, L. Prak, M. Zemla, D. Jorgens, P. Varanasi, C. Li, Ö. Persil Cetinkol, L. Sun, K. Tran, Y. Verhertbruggen, H.V. Scheller, D.C. Dibble, B.M. Holmes, **B.A. Simmons**, **P.D. Adams**, S. Singh, and M. Auer* (Mauer@lbl.gov)

Joint BioEnergy Institute, Emeryville, Calif.

Project Goals: See goals for Abstract 26.

Using sophisticated sample preparation followed by scanning and transmission electron microscopy imaging we have studied a variety of plants and pretreated biomass samples, with a focus on the effects of ionic liquid pretreatment on Switchgrass plant cell walls. We have compared ionic liquid pretreatment to other pretreatment techniques such as acid pretreatment and ammonia fiber expansion, and find ionic liquids to exhibit a much larger effect on the biomass. By imaging plant material being pretreated with ionic liquids from different time points, we found that the cell wall material shows significant changes within less than 30 minutes of exposure to ionic liquids. We utilized widefield TEM to cover a statistically significant number of cell walls in different tissues and compared their precise dimension as a function of exposure to ionic liquids. We found that only the secondary cell wall of sclerenchyma cells undergoes dramatic changes, increasing to twice its original dimension over time, whereas primary cell walls were not found to expand significantly.

Interestingly, the effect of the ionic liquid pretreatment is very different for various plant feedstocks: While switchgrass and corn stover get dissolved easily there is only a minor effect on Eucalyptus arboretum. Other efforts currently underway at JBEI include cell wall characterization of *Arabidopsis* mutants both via optical and EM imaging as well as mechanical stress testing, and in addition imaging of lignocellulose digesting microbial communities.

This work was part of the DOE Joint BioEnergy Institute (<http://www.jbei.org>) supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U.S. Department of Energy.

44

Biobrick Vectors and Datasheets: A Synthetic Biology Platform for Metabolic Engineering

Taek Soon Lee* (TSLee@lbl.gov), Rachel A. Krupa, Meghdad Hajimorad, Nilu Prasad, Sung Kuk Lee, and **Jay D. Keasling**

Joint BioEnergy Institute, Emeryville, Calif.

Project Goals: See goals for Abstract 26.

Synthetic biology has been explored as a powerful tool in metabolic engineering. Standardization of biological parts is one of the most fundamental grounds to provide the foundation of designing synthetic biological systems, and there have been several attempts of establishing standard assembly strategy. For example, researchers at MIT had established BioBrick™ standard biological part strategy using XbaI and SpeI restriction enzymes and started the Registry of Standard Biological Parts. Here, using a similar strategy, we present a new standard using more robust BamHI and BglII restriction enzymes for the construction of novel plasmids with gene expression devices. We have designed and constructed 96 biobrick-compatible plasmids with a various combination of replication origin, antibiotic resistance, and transcriptional promoter. With these plasmids, we have collected protein expression data in various culture conditions using fluorescent protein as a reporter and documented them as a format of datasheet. This biobrick vector datasheet will be a useful source of information for designing and engineering metabolic pathways toward biofuel production.

This work was part of the DOE Joint BioEnergy Institute (<http://www.jbei.org>) supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U.S. Department of Energy.

45

Flux Analysis of Biodiesel-Producing *E. coli*

H.G. Martin* (HGMartin@lbl.gov), J. Gin, E. Steen, P. Benke, H. Burd, A. Mukhopadhyay, and **J.D. Keasling**

Fuels Synthesis Division, Joint BioEnergy Institute, Emeryville, Calif.

Project Goals: See goals for Abstract 26.

Biofuel yields can be limited below their theoretical maximum because of pathway characteristics (e.g. lack of enzyme activity or unbalanced gene expression) or a lack of flux directed towards the synthesizing pathway. The goal of the Host Engineering directorate is to prevent the latter. A way to achieve this is to knock genes out in such a way that carbon flux gets channeled towards the desired pathway. In

order to do that in a rational manner it is desirable to know the internal metabolic fluxes and have a way to predict the outcome of knockout experiments. Here we present a internal metabolic flux profiles for a biodiesel-producing *E. coli* measured through ¹³C metabolic flux analysis (¹³C MFA) and knockout suggestions obtained through Flux Balance Analysis (FBA) constrained by ¹³C MFA flux measurements.

This work was part of the DOE Joint BioEnergy Institute (<http://www.jbei.org>) supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U.S. Department of Energy.

46

Identification of Genes Essential to Long-Chain Alkene Biosynthesis in *Micrococcus luteus*

Ee-Been Goh, **Jay D. Keasling**, and Harry R. Beller* (HRBeller@lbl.gov)

Fuels Synthesis Division, Joint BioEnergy Institute, Emeryville, Calif.

Project Goals: See goals for Abstract 26.

Aliphatic hydrocarbons are appealing targets for advanced cellulosic biofuels, as they are predominant components of petroleum-based gasoline and diesel fuels and thus would be compatible with existing engines and fuel distribution systems. We have studied alkene biosynthesis in *Micrococcus luteus*, a close relative of *Sarcina lutea* (now *Kocuria rhizophila*), which was previously reported to biosynthesize *iso*- and *anteiso*-branched, long-chain alkenes. The underlying biochemistry and genetics of alkene biosynthesis were not elucidated in those studies. We show here that heterologous expression of a three-gene cluster from *M. luteus* (Mlut_13230-13250) in a fatty-acid overproducing *E. coli* strain resulted in production of long-chain alkenes, predominantly 27:3 and 29:3 (no. carbon atoms: no. C=C bonds). Heterologous expression of Mlut_13230 (*oleA*) alone produced no long-chain alkenes but unsaturated aliphatic monoketones, predominantly 27:2, and *in vitro* studies with the purified Mlut_13230 protein and tetradecanoyl-CoA produced the same C₂₇ monoketone. Gas chromatography-time of flight mass spectrometry confirmed the elemental composition of all detected long-chain alkenes and monoketones (putative intermediates of alkene biosynthesis). Negative controls demonstrated that the *M. luteus* genes were responsible for production of these metabolites. Studies with wild-type *M. luteus* showed that the expression of Mlut_13230-13250 and 29:1 alkene biosynthesis both corresponded with bacterial population over time. We propose a metabolic pathway for alkene biosynthesis starting with acyl-CoA (or -ACP) thioesters and involving decarboxylative Claisen condensation as a key step, which we believe is catalyzed by OleA. Such activity is consistent with our data

and with the homology of Mlut_13230 (OleA) to FabH (β -ketoacyl-ACP synthase III), which catalyzes decarboxylative Claisen condensation during fatty acid biosynthesis.

This work was part of the DOE Joint BioEnergy Institute (<http://www.jbei.org>) supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U.S. Department of Energy.

47

Towards Automated Assembly of Biological Parts

N.J. Hillson* (NJHillson@lbl.gov), J.W. Thorne, M.Z. Hadi, and **J.D. Keasling**

Fuels Synthesis and Technology Divisions, Joint BioEnergy Institute, Emeryville, Calif.

Project Goals: See goals for Abstract 26.

The production of clean renewable biofuels from cellulosic starting material requires concerted feedstock engineering, deconstruction of plant matter into simple sugars, and microbial fermentation of the sugars into biofuel. These three efforts share significant molecular biological challenges, including the construction of large enzymatic libraries (e.g. vast collections of glycosyl transferases, cellulases, and efflux pumps), the generation of combinatorial libraries (e.g. multi-functional enzyme domain fusions; variations in copy number, promoter and ribosomal binding site strength), and the concurrent assembly of multiple biological parts (e.g. the incorporation of an entire metabolic pathway into a single target vector). With these challenges in mind, we are developing hybrid multi-part assembly methodologies and translating them to robotics-driven protocols. Given a target library to construct, our vision is that the high-throughput methodology will provide automated oligo and optimal assembly process design, and robotic control of the PCR and multi-part assembly reactions. The beneficial output of this work will include reagents and resources for, and collaborations with, members of the JBEI and larger life sciences communities, reducing the time, effort and cost of large scale cloning and assembly tasks, as well as enabling research scales otherwise not feasible without the assistance of computer-aided design tools and robotics.

This work was part of the DOE Joint BioEnergy Institute (<http://www.jbei.org>) supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U.S. Department of Energy.

48

Improving Biofuel Production by Using Efflux Pumps to Limit Solvent Toxicity

Mary J. Dunlop, Masood Hadi, Mario Ouellet, **Paul D. Adams**, **Jay D. Keasling**, and Aindrila Mukhopadhyay* (Amukhopadhyay@lbl.gov)

Fuels and Technology Divisions, Joint BioEnergy Institute, Emeryville, Calif.

Project Goals: See goals for Abstract 26.

Biofuels can be produced by microbes that break down plant matter or sugars to make fuel. However, biofuel-producing microbes are limited by the intrinsic toxicity of the solvent like biofuels they are trying to produce. The more fuel the cell produces, the more toxic the surrounding environment becomes. RND efflux pumps are a class of membrane transporters that confer resistance to a wide variety of toxins, including solvents. We focus on investigating the role of native, as well as heterologously expressed, efflux pumps in *E. coli*. Targeted studies focus on the well-characterized *E. coli* AcrAB-TolC system, and efflux pumps from solvent resistant bacteria such as *Pseudomonas putida*. Because efflux pumps are likely to be specific to certain fuel molecules and stressors, a wider range of native and heterologous efflux pump systems must be tested against different fuel compound exposure, growth conditions, and in different engineered hosts. To address our broad goal of improving solvent resistance using efflux pumps, a high-throughput approach has been initiated to create a library of expression vectors representing all efflux pumps from *E. coli* as well from other organisms known to be naturally resistant to solvents.

This work was part of the DOE Joint BioEnergy Institute (<http://www.jbei.org>) supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U.S. Department of Energy.

49

Quantitative Proteomics for Metabolically Engineered Biofuel Pathway Optimization

Christopher Petzold* (CJPetzold@lbl.gov), Alyssa Redding, Tanveer Bathth, Rossana Chan, Rachel Krupa, Taek Soon Lee, Aindrila Mukhopadhyay, **Paul Adams**, and **Jay Keasling**

Fuels Synthesis and Technology Divisions, Joint BioEnergy Institute, Emeryville, Calif.

Project Goals: See goals for Abstract 26.

Analytical methods are fundamental to establishing the success of biofuel metabolic engineering efforts. Monitoring

and quantifying pathway intermediates facilitates identification of bottlenecks and where alterations offer the greatest potential impact on titer. However, many different factors may contribute to a bottleneck. High levels of protein production or high activity of an upstream enzyme will produce substrates at rates higher than can be consumed by later steps in the pathway. Analogously, downstream enzymes may have low activity or be poorly produced, and factors such as low solubility or rapid protease degradation can limit the amount of enzyme available. Consequently, the detection and quantification of the enzymes of interest is integral to optimizing engineered pathways. Quantitative proteomics analysis, consisting of liquid chromatographic separation coupled to mass spectrometry, is a rapid method to correlate protein expression levels with metabolite titers. At JBEI, both targeted and untargeted proteomics experiments are used to quantify proteins crucial for biofuel production. With these methods we are characterizing a variety of protein expression conditions for several biofuels pathways, including various promoters and plasmid systems, to identify bottlenecks and determine optimal protein levels for high biofuel titers.

This work was part of the DOE Joint BioEnergy Institute (<http://www.jbei.org>) supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U.S. Department of Energy.

50

Strategies to Improve Resistance and Production Phenotypes of *E. coli*

Adrienne E. McKee, John Haliburton, Helcio Burd, **Jay Keasling**, and Swapnil Chhabra* (SRChhabra@lbl.gov)

Fuels Synthesis Division, Joint BioEnergy Institute, Emeryville, Calif.

Project Goals: See goals for Abstract 26.

Many of the microbial phenotypes of interest to metabolic engineers are complex in that multiple genes, pathways, and regulatory networks are involved in generating the targeted behavior. *A priori* prediction of additional changes that will further improve phenotypes can therefore prove difficult due to our incomplete understanding of the functions and connectivity of gene products far removed from the pathway of interest. One complement to rational approaches is to exploit the strength of mutation and selection or screening to obtain strains capable of improved resistance to pretreatment growth improved production titers in the case of production phenotypes. Towards these goals, we are employing and refining methods that rely on natural or augmented mutation rates or on directed protein evolution to improve *E. coli* phenotypes. Our three major avenues of investigation include selection for inhibitor resistance by continuous culture in chemostats, development of inducible

and temperature sensitive mutator plasmids, and generation and screening plasmid libraries of mutated gene regulators for enhanced phenotypic behavior. We are interested in both the genetic and regulatory alterations that underlie phenotypic improvements.

This work was part of the DOE Joint BioEnergy Institute (<http://www.jbei.org>) supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U.S. Department of Energy.

51

Microbial Production of Fatty Acid-Derived Fuels and Chemicals in *Escherichia coli*

Eric J. Steen^{1,2} * (EJSteen@lbl.gov), Yisheng Kang,¹ Mario Ouellet,¹ Helcio Burd,¹ John Haliburton,¹ Michael Lee,^{1,2} Alyssa Redding,¹ Heather Szmids,¹ Jennifer Gin,¹ Nathan Hillson,¹ Harry R. Beller,¹ and **Jay D. Keasling**^{1,2,3,4,5}

¹Joint BioEnergy Institute, Emeryville, Calif.; ²Dept. of Bioengineering, ³QB3 Institute, and ⁵Dept. of Chemical Engineering, University of California, Berkeley; and ⁴Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, Calif.

Project Goals: See goals for Abstract 26.

Increasing energy costs and environmental concerns have emphasized the need to sustainably produce renewable fuels and chemicals. Major efforts to this end are focused on the microbial production of high-energy fuels through cost-effective “consolidated bioprocesses”. Fatty acids are composed of long alkyl chains and represent nature’s “petroleum,” being a primary metabolite class used by cells for both chemical and energy storage functions. These energy rich molecules are today isolated from plant and animal oils for a diverse set of products ranging from fuels to oleochemicals. A more scalable, controllable, and economic route to this important class of chemicals would be through the microbial conversion of renewable feedstocks, such as biomass-derived carbohydrates. Here we demonstrate the engineering of *E. coli* to produce structurally tailored fatty acid ethyl esters (biodiesel), fatty alcohols, and waxes directly from simple sugars and the further engineering of the biodiesel-producing cells to secrete hemicellulases, a step toward producing these compounds directly from hemicellulose, a major component of plant-derived biomass¹. Although this complete production scheme has been demonstrated, increases in titer, productivity, and yield are necessary for industrial transition. Strategies employed for increasing yields of biodiesel include balancing the enzymes in the pathway, condensing the pathway onto a triple-operon, single-plasmid system, and subsequent chromosomal integration. These efforts guided towards understanding fermentation scalability, pathway stability, and balancing pathway enzymes for biodiesel production have resulted in higher yields.

Reference

1. Steen EJ, Kang Y, Bokinsky G, Hu Z, Schirmer A, McClure A, del Cardayre SB, Keasling JD. "Microbial Production of Fatty Acid-Derived Fuels and Chemicals from Plant Biomass," *Nature* (2009) accepted for publication.

This work was part of the DOE Joint BioEnergy Institute (<http://www.jbei.org>) supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U.S. Department of Energy.

52

JBEI Electronic Laboratory Notebook System

D. Pletcher* (DCPletcher@lbl.gov), J. Gladden, B. Holmes, J. Keasling, S. Lane, W. Morrell, A. Herrera Panganiban, J. Petersen, F.W. Townsend-Merino, J. Wildman, and **P.D. Adams**

Joint BioEnergy Institute, Emeryville, Calif.

Project Goals: See goals for Abstract 26.

JBEI is pursuing a novel approach to implement an electronic laboratory notebook (ELN) system. Ease of use, performance and cost are primary factors in the design of the system. Commercially successful, inexpensive software products are combined into a full-featured, hybrid system using integration software developed in-house. Powerful desktop note-taking software on a commodity tablet computer enables scientists to capture and organize notes, sketches, images, documents and other digital data files. All notebook information is periodically copied to a central, web-accessible repository for search and review by other authorized users. Integration with a secure digital time-stamping service ensures that intellectual property can be defended in court. Pilot users have exhibited a high level of satisfaction; new users are adopting the system willingly based on peer recommendations. The completed system will facilitate online discovery and scientific collaboration, and will serve as a core component of our GTL Knowledgebase.

This work was part of the DOE Joint BioEnergy Institute (<http://www.jbei.org>) supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U.S. Department of Energy.

53

JBEI Computational Biology Core

Dylan Chivian^{1,2,3*} (DCChivian@lbl.gov), Greg D. Friedland,^{1,2,4} John T. Bates,^{1,2,3} Paramvir S. Dehal,^{2,3} Marcin P. Joachimiak,^{2,3} Keith Keller,^{2,3} Morgan N. Price,^{2,3} Jason K. Baumohl,^{2,3} and **Adam P. Arkin**^{1,2,3,5*}

¹DOE Joint BioEnergy Institute, Berkeley, Calif.; ²Virtual Institute for Microbial Stress and Survival, Berkeley, Calif.; ³Lawrence Berkeley National Laboratory, Berkeley, Calif.; ⁴Sandia National Laboratories, Livermore, Calif.; and ⁵University of California, Berkeley

Project Goals: See goals for Abstract 26.

The Computational Biology Core Group in the Technology Division of the Joint BioEnergy Institute (JBEI) is responsible for data integration and comparative, evolutionary, and functional genomic analysis for the purpose of engineering microbes for biofuel production. Leveraging the VIMSS MicrobesOnline web resource (<http://www.microbesonline.org>) for comparative and evolutionary genomics and analysis of microarray, proteomic, and metabolomic data sets, we are extending the supported microorganisms to include bioenergy-relevant fungi and algae as well as integrating capabilities to allow for pursuit of questions specific to bioenergy challenges.

Computational analysis and biological engineering requires an understanding of the biology at several scales. The efforts of the Computational Biology Core are therefore geared towards creating tools to facilitate our research at the level of components, systems, cells, and communities. For example, biological degradation of plant cells walls is accomplished by enzymes containing multiple domains that in combination confer specificity and activity. We are studying the combinations nature has employed to allow for making our own combinations. At the atomic scale, we are computationally designing the structures of proteins for increased stability under industrial conditions. At the system level, we are working to discover metabolic pathways for biofuel production as well as the genetic factors involved in tolerance to harsh biomass pretreatment conditions and resistance to biofuel toxicity. To aid efforts in synthetic pathway engineering, we are building a framework for analyzing functional data in a metabolic network context. At the cellular level, we are using evolutionary studies with phenotype data and genetic analysis to engineer cell lines that are better suited to industrial conditions or have superior yields of the desired biofuel. Finally, at the community level, discovery of genes from environmental samples will expand the repertoire of enzymes we can engineer for biomass degradation under varying conditions. Our approach is to analyze the environmental genomic data in a phylogenetic context to allow for a higher-resolution annotation of the role of each enzyme, as well as to obtain an evolutionary picture of the key functional genes and organisms in each ecosystem.

This work was part of the DOE Joint BioEnergy Institute (<http://www.jbei.org>) supported by the U.S. Department of Energy,

Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U.S. Department of Energy.

54

Techno-Economic Modeling of Cellulosic Biorefineries

Daniel Klein-Marcuschamer* (DKlein@lbl.gov), Piotr Oleskowicz-Popiel, **Blake A. Simmons**, and **Harvey W. Blanch**

Deconstruction Division, Joint BioEnergy Institute, Emeryville, Calif.

Project Goals: See goals for Abstract 26.

The realization of biofuels at a commercial scale will depend on processes that are energetically, environmentally, and economically sustainable. One main thrust of biofuels research, therefore, has been in techno-economic analysis of biofuel alternatives, which facilitates process design, optimization, and performance evaluation. These studies usually rely on experimentally-derived or assumed parameters to estimate process performance values such as capital and operating costs, GHG emissions, biofuel yield on feedstock, among others. Naturally, the results of the simulations strongly depend on the parameter choices or scenario maps that are considered, which limits the information that the community can extract from the results if only a few instances are outlined. Here, we present a techno-economic model of lignocellulosic ethanol production that is open and transparent and that uses assumptions that reflect technologies that are currently available. Using this model, we have studied how feedstock, enzyme, and strain engineering efforts could potentially affect the economic and performance attributes of the process. The results indicate that pretreatment and hydrolysis yield, inhibition during fermentation, and energy consumption are major factors impacting the economic viability of the process.

This work was part of the DOE Joint BioEnergy Institute (<http://www.jbei.org>) supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U.S. Department of Energy.

BioEnergy Science Center (BESC)

55

The BESC Knowledgebase and Public Web Portal for Bioenergy-Related Organisms

Morey Parang,^{1,2} Denise Schmoyer,^{1,2} Tatiana Karpinets^{1,2*} (karpinetsv@ornl.gov), Mustafa Syed,^{1,2} Mike Leuze,^{1,2} Byung Park,^{1,2} Doug Hyatt,^{1,2} **Ed Uberbacher**,^{1,2} and Martin Keller^{1,2}

¹BioEnergy Science Center, Oak Ridge, Tenn. and ²Oak Ridge National Laboratory, Oak Ridge, Tenn.

<http://bioenergycenter.org>

Project Goals (Abstracts 55-56): The BioEnergy Science Center (BESC) is focused on the fundamental understanding and elimination of biomass recalcitrance. BESC's approach to improve accessibility to the sugars within biomass involves 1) designing plant cell walls for rapid deconstruction and 2) developing multitolerant microbes for converting plant biomass into biofuels in a single step (consolidated bioprocessing).

BESC researchers in characterization, modeling, education and data management seek to provide integrating data and understanding across the center. We use advanced technologies to analyze chemical and structural changes within biomass. We have developed a coordinated characterization pipeline to provide composition, pretreatment, and enzyme digestibility data from native and modified plant samples and to store this data within a data management system. We are developing knowledge environments to allow access to improved data analysis and modeling tools.

Combining experimental data with biological models will help define the genomic and physical basis of plant cell-wall recalcitrance and deconstruction. Laboratory Information Management Systems and knowledgebase environments have separate key roles.

The objective of the BESC Knowledge base project is to develop and implement an infrastructure to represent systems level knowledge, data and information about key plants, microbes, and molecules in BESC and provide views of these data to users. The Knowledgebase (KB) is designed to aid investigators in the comprehensive evaluation of organisms for bioenergy associated strategies. Specifically, the Knowledgebase provides annotations of sequenced bioenergy related organisms, their phenotypic characteristics, pathway genome databases (PGDBs) for each organism, orthology information, a set of tools for annotation and analysis of carbohydrate active enzymes (CAZymes tool kit) and a diverse set of product annotations from different public resources, like InterPro, CDD, COG, PFAM, SMART, and others. The user environment has (i) genomic tools and gene cards, (ii) BioCyc pathway tools including capabilities

to search and browse PGDBs, (iii) a comparative analysis environment for overlaying the metabolic maps with BESC or user provided experimental 'omics' data, (iv) a knowledge mining environment that allows phenotype comparison in terms of CAZy Families/ Pathways/ Enzymes and sequence annotation by CAZy families using PFAM domains, and (v) integration of investigator-derived or external data (External Experimental Data (GEO/ArrayExpress) with product annotation from diverse set of databases including CAZy. The collected information has been employed in the analysis of experimental data produced by BESC and for comparative analysis of phenotypic characteristics of the BESC targeted organisms in terms of their genomic, metabolic and cellular characteristics.

We have constructed a public portal to the BESC knowledgebase to provide the larger community with information about bioenergy organisms. The portal provides integrated views of information available from a variety of different public resources or produced by BESC (public domain) and presents this knowledge in a systematic and unified way. The portal environment allows users to search for different types of annotations for each organism or across organisms; download the annotations and results of queries in user friendly formats; compare phenotypic characteristics of organisms in terms of their genomic, metabolic and cellular characteristics; visualize and link the annotations to the experimental data for their further analysis.

The KB is linked to the BESC Laboratory Information Management System (LIMS), which is used for tracking samples and capturing experiment metadata. The combined KB and LIMS projects have developed ORNL standard data models for experimental workflow from the project level down to individual experiments, aliquots, and instrument metadata. We intend to leverage these ongoing efforts for the combined GTL data sharing effort and global GTL knowledge base developments.

The BioEnergy Science Center is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science.

56 Carbohydrate-Active Enzyme Annotation Tools (CAT) in the BESC Knowledgebase Portal

Mustafa Syed^{1,2*} (syedmh@ornl.gov), Byung-Hoon Park,^{1,2} Tatiana Karpinetz,^{1,2} Michael Leuze,^{1,2} Ed Uberbacher,^{1,2} and Martin Keller^{1,2}

¹BioEnergy Science Center, Oak Ridge, Tenn. and ²Oak Ridge National Laboratory, Oak Ridge, Tenn.

<http://bioenergycenter.org>

Project Goals: See goals for abstract 55.

The Carbohydrate-Active Enzyme (CAZy) database is a manually curated repository of experimentally characterized enzymes that build and breakdown complex carbohydrates. The enzymes, also referred to as CAZymes, implement many important functions in the cell and of great interest of the biofuel research because of their involvement in biosynthesis and the degradation of the complex polysaccharides of the plant cell wall. Despite rich and invaluable information stored in the database, software tools utilizing this information for analysis and annotation of newly sequenced genomes by CAZy families are limited, partly because of limited capabilities of the CAZy database for searching and downloading enzyme sequences.

We have addressed this problem by developing a set of tools to search the database and to annotate a new sequence or a set of sequences with CAZy families. We used Perl scripts to scrape HTML web pages for each family in the CAZy database and have organized the downloaded information in a local MySQL database. The collected information was supplemented by protein sequences downloaded from Genbank and by sequence associated information, such as Genbank accession numbers, enzyme names, EC numbers, predicted protein families from the Pfams database and the organism taxonomy. We then developed a set of tools and a user friendly web interface to search the collected information by the enzymes name or keyword, CAZy family, organism taxonomy, name or keyword. Because many CAZymes, especially from Glycosyl Hydrolase (GH) families, have a complex modular architecture and are annotated by two and more families, an interface was developed to search such associations across the entire CAZy database.

Two complementary approaches were provided in the toolkit to annotate sequences with CAZy families. The first approach is based on a similarity search of a protein sequence or a set of sequences in a given genome against the entire non-redundant sequences of the CAZy database using uni-directional or bi-directional Blast. The second approach is based on associations between protein domains and CAZy families. In addition to links between Pfam domains and CAZy families available in the CAZy database, we have derived a set of new links by applying the association rule learning algorithm to the collected data. The augmented set of Pfam-to-CAZy family associations has significantly improved sensitivity of the annotation. The evaluation of the approaches using the manually curated genomes of *Clostridium thermocellum* ATCC 27405 and *Saccharophagus degradans* 2-40 indicated that in combination they can provide a high degree of specificity and sensitivity for predicting CAZymes in the newly sequenced organisms.

The developed tools were employed to predict CAZymes in a recent submission of the *Populus trichocarpa* genome (Jul 30 2009) and in the *Escherichia coli* genome. By using the discovered confident associations between CAZy families and pfam domains we assigned functional activity, taxonomic groups and CAZy families to several unknown domains including DUF2029, DUF297, DUF303, and DUF847. We have also revealed some conserved associa-

tions between CAZy families that are characteristics of specific taxonomic groups, like plants, fungi or bacteria.

The CAZy toolkit and environment is accessible as part of the BESC public portal at <http://cricket.ornl.gov/cgi-bin/cat.cgi>.

The BioEnergy Science Center is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science.

57 Technology and Transgenics for Unparalleled Improvements in Switchgrass Biomass Quality

D.G.J. Mann^{1,5*} (dmann1@utk.edu), L.L. Abercrombie,^{1,5} H. Baxter,^{1,5} K. Berk,^{1,5} J.N. Burris,^{1,5} F. Chen,^{2,5} X. Chen,^{1,5} C. Fu,^{2,5} M. Halter,^{1,5} J. Han,^{1,5} F. Hardin,^{2,5} J.S. Hawkins,^{3,5} T. Hernandez,^{2,5} H. Hisano,^{2,5} Z. King,^{3,5} P.R. LaFayette,^{3,5} Y. Lee,^{2,5} M. Mazarei,^{1,5} J.R. Mielenz,^{4,5} R.J. Percifield,^{3,5} C. Poovaiah,^{1,5} H. Ramanna,^{2,5} M. Rudis,^{1,5} H. Shen,^{2,5} I. Torres-Jerez,^{2,5} X. Xiao,^{2,5} J. Zhang,^{2,5} J.L. Bennetzen,^{3,5} R.A. Dixon,^{2,5} R. Nelson,^{2,5} W.A. Parrott,^{3,5} M.K. Udvardi,^{2,5} Z.-Y. Wang,^{2,5} J. Zale,^{1,5} **C.N. Stewart Jr.**,^{1,5} and M. Keller^{4,5}

¹University of Tennessee, Knoxville; ²The Samuel Roberts Noble Foundation, Ardmore, Okla.; ³University of Georgia, Athens; ⁴Oak Ridge National Laboratory, Oak Ridge, Tenn.; and ⁵BioEnergy Science Center, Oak Ridge, Tenn.

<http://bioenergycenter.org>

Project Goals (Abstracts 57-59): The BioEnergy Science Center (BESC) is focused on the fundamental understanding and elimination of biomass recalcitrance. BESC's approach to improve accessibility to the sugars within biomass involves 1) designing plant cell walls for rapid deconstruction and 2) developing multitolerant microbes for converting plant biomass into biofuels in a single step (consolidated bioprocessing).

BESC biomass formation and modification research involves working directly with two potential bioenergy crops (switchgrass and *Populus*) to develop varieties that are easier to break down into fermentable sugars. We are using both testing and generation of a large number of natural and modified plant samples as well as detailed studies into poorly understood cell wall biosynthesis pathways.

The Switchgrass activity is pursuing improved transformation methods, association studies, and targeted pathways such as lignin synthesis.

Switchgrass (*Panicum virgatum*) has become a leading candidate feedstock for biofuels in the U.S. and is a crucial model feedstock component of BESC. Biotechnology of switchgrass is important in screening potential cell wall

biosynthesis genes, and is being performed by six laboratories in three institutions within BESC. Accomplishments include altering lignin biosynthesis, improved tissue culture and transformation systems, optimization of a virus-induced gene silencing (VIGS) system and a new vector set for monocot transformation. A transgene pipeline committee was established to identify genes of interest to be evaluated in stably transformed switchgrass (through overexpression or knockdown technologies). To evaluate genes with unknown or poorly understood functions prior to stable transformation we are utilizing VIGS, which more rapidly (within 2 months) provides results through transient knockdown of target gene expression. Our VIGS system uses a *Brome mosaic virus*-based vector to silence genes in foxtail millet (*Setaria italica*), a closely related species to switchgrass which serves as a simpler model system. If a recalcitrance-altering phenotype is observed through VIGS, stable switchgrass transformants will be produced for that gene. One example of success of our research path is the downregulation of a gene from the lignin biosynthesis pathway. The transgenic switchgrass for this gene yields 25% more ethanol than its non-transgenic isolate. In order to coordinate gene expression and to facilitate more rapid screening of genes, we have developed a Gateway-compatible monocot transformation vector set (pANIC) for overexpression and RNAi with visual and selectable markers. BESC has facilitated the coordination of scientific expertise and research in switchgrass biotechnology that would have been otherwise impossible in any individual laboratory.

The BioEnergy Science Center is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science.

58 Enhanced Quality, Value, Yield, Carbon Capture, and Sustainability of Switchgrass Biomass by the Improvement of Root, Microbe, and Soil Interactions

Sita Ghimire^{1,4*} (srghimire@noble.org), **K.D. Craven**,^{1,4} Srinivasa Chaluvadi,^{2,4} Jeff Bennetzen,^{2,4} Eric Worley,^{1,4} Jiading Yang,^{1,4} Michael Udvardi,^{1,4} and Martin Keller^{3,4}

¹The Samuel Roberts Noble Foundation, Ardmore, Okla.; ²University of Georgia, Athens; ³Oak Ridge National Laboratory, Oak Ridge, Tenn.; ⁴BioEnergy Science Center, Oak Ridge, Tenn.

<http://bioenergycenter.org>

Project Goals: See goals for abstract 57.

Nutrient management in biomass production systems serves to maximize yield and minimize production costs and environmental impact. Sustainable production of switchgrass for biofuel will depend, in part, on maximizing nutrient acquisition and assimilation throughout the growing season as well as minimizing nutrient loss at harvest. Nutrient acquisition

and uptake by plants can be enhanced by beneficial soil microbes as well as those existing endophytically within the roots of the switchgrass host. We have undertaken a comprehensive characterization of the microbes associated with the rhizosphere of planted switchgrass cultivars as well as those found within the healthy, surface-sterilized root systems of natural plants found in their native habitat. High levels of microbial biodiversity were detected for both fungi and bacteria, and several strains have been isolated for evaluation of fitness effects on elite switchgrass cultivars. Dramatic differences in rhizosphere and endophyte microbial populations have been found to be a function of host genetics by analysis of different switchgrass cultivars, and mapping studies are now initiated to identify the host genes that determine microbial composition in and around the root. Nutrient use efficiency in plants is a consequence of both the frugality of utilization in the field as well as the effective recycling of those nutrients from aerial tissues to the root system at or before senescence. To facilitate the breeding of varieties that are conservative in their use of soil nutrients to produce biomass, we assessed the natural variation in nutrient-use and remobilization efficiencies of 31 accessions of *Panicum virgatum* by measuring the concentration of 20 elements (N, P, K, Li, B, Na, Mg, Ca, Mn, Fe, Co, Ni, Cu, Zn, As, Se, Rb, Sr, Mo and Cd) in shoots of field-grown plants harvested at two different stages of development. We detected significant differences between accessions for elemental composition at maturity and after senescence. The accessions/cultivars with the greatest nutrient-use efficiency (smallest loss of nutrient per unit biomass) were BN-14668-65, Kanlow, and Caddo from the point of view of N content, and Kanlow, Cave-in-Rock, and Blackwell from the point of view of P content in senescent shoots. These data will allow a holistic nutrient management strategy to be employed for maximizing yield and sustainability of this important bioenergy crop.

The BioEnergy Science Center is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science.

59

Creating the Genetic and Genomic Foundations to Improve Bioenergy Production from Switchgrass

Katrien M. Devos^{1,4*} (kdevos@uga.edu), Ananta Acharya,^{1,4} Bochra Bahri,^{1,4} Wesley Dean,^{1,4} Matt Estep,^{1,4} Ansuya Jogi,^{1,4} Jennifer Hawkins,^{1,4} Yi-Ching Lee,^{2,4} Ryan Percifield,^{1,4} Ivone Torres-Jerez,^{2,4} Xueweng Wang,^{1,4} Limin Wu,^{1,4} Jiyi Zhang,^{2,4} Qihui Zhu,^{1,4} Joe Bouton,^{2,4} E. Charles Brummer,^{1,4} Malay C. Saha,^{2,4} Shavannor M. Smith,^{1,4} Yuhong Tang,^{2,4} Michael Udvardi,^{2,4} Jeffrey L. Bennetzen,^{1,4} and Martin Keller^{3,4}

¹University of Georgia, Athens; ²The Samuel Roberts Noble Foundation, Ardmore, Okla.; ³Oak Ridge National Laboratory, Oak Ridge, Tenn.; and ⁴BioEnergy Science Center, Oak Ridge, Tenn.

<http://bioenergycenter.org>

Project Goals: See goals for abstract 57.

Switchgrass (*Panicum virgatum*), a grass native to U.S. prairies, has agronomic characteristics that make it an exceptionally promising feedstock for bioenergy production. There is great potential for improving the biomass composition and biomass yield of switchgrass using genomics-based breeding. We are using a multifaceted approach to create the genetic, genomic and breeding resources that are required for building a successful switchgrass biofuel production program. A comprehensive genetic map consisting of SSR and DArT markers is under construction in a lowland Alamo AP13 x upland Summer VS16 cross. The map will be used for trait mapping, as well as anchoring of the genomic sequence that is currently being produced from genotype AP13 by our collaborators at JGI. An ~6X coverage fosmid library has been generated from nuclear DNA of switchgrass cultivar Alamo, and >30 fosmids containing genes involved in switchgrass cell wall synthesis/composition have been selected, sequenced and annotated. Seven million ESTs from AP13 and VS16 have been generated and placed in a searchable database. The ESTs are derived from mRNA isolated from roots and shoots at three different stages of development and from mRNA enriched for secondary cell wall biosynthesis using laser capture microscopy. The sequence data will be used for SNP development and linkage disequilibrium studies, and will assist in the annotation of the switchgrass genomic sequence and will help in understanding and engineering improved cell wall production in switchgrass. A large biodiversity study consisting of 384 mostly lowland genotypes is under way to assess natural variation for components that affect recalcitrance to ethanol production, biomass yield and disease resistance. Genotyping and phenotyping will be combined to identify trait-marker associations. This information is fed into the breeding program to enhance the development of switchgrass cultivars with desirable cell wall composition and increased yield for commercialization. In summary, our

project provides the essential resources to conduct structural and functional analyses in switchgrass.

The BioEnergy Science Center is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science.

60

Application of an Integrated High-Throughput Pretreatment and Enzymatic Hydrolysis (HTPPH) Screening Tool to Identify Key Biomass Features and Processing Conditions

Jaclyn DeMartini^{1,3*} (jaclyn.demartini@email.ucr.edu), Michael Studer,^{1,3} Charles E. Wyman,^{1,3} and Martin Keller^{2,3}

¹University of California, Riverside; ²Oak Ridge National Laboratory, Oak Ridge, Tenn.; and ³BioEnergy Science Center, Oak Ridge, Tenn.

<http://bioenergycenter.org>

Project Goals (Abstracts 60-64): The BioEnergy Science Center (BESC) is focused on the fundamental understanding and elimination of biomass recalcitrance. BESC's approach to improve accessibility to the sugars within biomass involves 1) designing plant cell walls for rapid deconstruction and 2) developing multitolerant microbes for converting plant biomass into biofuels in a single step (consolidated bioprocessing).

BESC researchers in characterization, modeling, education and data management seek to provide integrating data and understanding across the center. We use advanced technologies to analyze chemical and structural changes within biomass. We have developed a coordinated characterization pipeline to provide composition, pretreatment, and enzyme digestibility data from native and modified plant samples and to store this data within a data management system. We are developing knowledge environments to allow access to improved data analysis and modeling tools.

Knowledge gained by thoroughly characterizing biomass chemistry and structure will drive coordinated development of improved plant biomass and degradation microbes. Native, genetically modified, and partially deconstructed lignocellulosic samples will be analyzed.

A high throughput pretreatment and enzymatic hydrolysis (HTPPH) method has been developed as part of the BioEnergy Science Center (BESC) to screen natural and genetically modified biomass types for those with lower recalcitrance to sugar release, define pretreatment conditions, and screen enzyme formulations. The high throughput system has been shown to mimic conventional pretreatment and enzymatic hydrolysis laboratory methods but has the advantage of being able to quickly screen hundreds of

samples. Thus far, the HTPPH system has been successfully applied to a set of 47 natural *Populus trichocarpa* samples in BESC's Poplar Association Study to define trends in sugar release behavior and identify outliers that warrant further study. In addition, the HTPPH system has enabled screening of individual annual rings from a cross section of *Populus tremuloides* to investigate radial variation of sugar release and the importance of sampling technique. Results will be shown from these studies to demonstrate the power of the HTPPH system for screening biomass samples for sugar release and identifying those with reduced recalcitrance. In support of the system, a scaled-down method was also developed to determine biomass composition that speeds compositional analysis, increases accuracy, and reduces labor demands.

The BioEnergy Science Center is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science.

61

Chemical Extraction of Down-Regulated C3H and HCT Alfalfa Reveals Structural Differences in Lignin

Angela Ziebell,^{1,6} Kristen Reichel,^{1,6} Rui Katahira,¹ Fang Chen,^{3,6} Yunqiao Pu,^{3,6} David Johnson,⁴ Robert Sykes,^{1,6} Art Ragauskas,^{3,6} Rick Dixon,^{2,6} Mark Davis^{1,6*} (mark.davis@nrel.gov), and Martin Keller^{5,6}

¹National Bioenergy Center, National Renewable Energy Laboratory, Golden, Colo.; ²Samuel Robert Noble Foundation, Ardmore, Okla.; ³Institute of Paper Science and Technology, Georgia Institute of Technology, Atlanta; ⁴Chemical and Biosciences Center, National Renewable Energy Laboratory, Golden, Colo.; ⁵Oak Ridge National Laboratory, Oak Ridge, Tenn.; ⁶BioEnergy Science Center, Oak Ridge, Tenn.

<http://bioenergycenter.org>

Project Goals: See goals for abstract 60.

The composition of biomass with regards to lignin content can impact the ease and cost associated with biomass processing and lignin reduction through breeding and genetic modification therefore has potential for reducing costs in biomass processing industries. The compositional changes of two low lignin Alfalfa (*M. sativa*) lines are investigated; antisense down regulated *p*-coumarate 3-hydroxylase (C3H) and hydroxycinnamoyl transferase (HCT). Sequential base extraction readily reduced the lignin content of the transgenic lines leaving a residual H lignin component equal in all lines. We will show that the differences in the lignin extraction of the different lines can be related to the differences in the reactivity of the H monomer versus the normally dominant G and S monomers during lignin formation.

The BioEnergy Science Center is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science.

62

Structural Characterization of the Xylan Oligosaccharides by Mass Spectrometry

Koushik Mazumder^{1,3*} (koushik@ccrc.uga.edu), Ameya Kulkarni,^{1,3} **William York**,^{1,3} and Martin Keller^{2,3}

¹University of Georgia, Athens; ²Oak Ridge National Laboratory, Oak Ridge, Tenn.; and ³BioEnergy Science Center, Oak Ridge, Tenn.

<http://bioenergycenter.org>

Project Goals: See goals for abstract 60.

The plant cell wall is a highly organized composite of many different polysaccharides, protein and aromatic substances. In plants the three main polysaccharides of the cell wall are cellulose, pectin and hemicellulose. Xylans play a very important role in establishing and maintaining the structural integrity of the plant's secondary cell walls, which are the most abundant components of terrestrial biomass. It is likely that differences in the molecular features of xylans (degree of branching and spatial arrangement along the xylan backbone) can alter the properties of the cell wall, including its recalcitrance to enzyme-catalyzed saccharification.

Structural characterization of carbohydrates is achieved usually using a combination of different techniques, which include NMR, GC-MS, mass spectrometry (ESI and MALDI-MS), specific chemical reactions or specific glycosidases. A major advantage of mass spectral analysis (ESI and MALDI-MS) over the other analytical techniques is its extremely high sensitivity (i.e., at the nanogram scale). We have developed highly sensitive methods for the analysis of xylan oligosaccharides using a combination of electrospray ionization (ESI) and multiple-stage mass spectrometry (MSⁿ) with a linear ion trap spectrometer. This approach provides specific information regarding the glycosyl sequences and branching patterns of these oligosaccharides. We will describe the application of these methods for the detailed structural characterization of the neutral and acidic xylan oligosaccharides obtained from the cell walls of various plant species.

The BioEnergy Science Center is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science.

63

Understanding Cellulase Activity Using Single Molecule Spectroscopy

Yu-San Liu,^{1,2} Yining Zeng,^{1,2} Yonghua Luo,¹ Michael E. Himmel,^{1,2} **Shi-You Ding**^{1,2*} (Shi.you.Ding@nrel.gov), and Martin Keller^{2,3}

¹Biosciences Center, National Renewable Energy Laboratory, Golden, Colo.; ²BioEnergy Science Center, Oak Ridge, Tenn.; and ³Oak Ridge National Laboratory, Oak Ridge, Tenn.

<http://bioenergycenter.org>

Project Goals: See goals for abstract 60.

To develop more cost-effective approaches to liberate fermentable sugars from recalcitrant biomass, the enzyme cocktail used for saccharification must be improved. We have developed a single-molecule technique based on fluorescence imaging and atomic force microscopy to characterize the cellulose morphology changes and to track the binding orientation and the motion of cellulase components with spatial resolution at the nanometer scale. We used single molecule spectroscopy to study the surface morphology of crystalline cellulose, as well as the real-time behavior of enzymes while bound to cellulose crystals. Preliminary results have revealed a confined nanometer-scale movement of the cellulase components bound to cellulose with preferred binding orientation. Cellulose crystals have also been imaged in real-time showing surface roughness changes, sharpening, and peeling effects by enzyme hydrolysis. The single molecule approach used here offers new opportunities to guide us toward a fundamental understanding of cellulase function, especially the mechanism of the "processivity" of exoglucanase.

The BioEnergy Science Center is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science.

64

Ultrastructure of Lignocellulose "Native-Pretreated-Deconstructed" by Advanced Solid-State NMR

Marcus Foston,^{1,3} Yunqiao Pu,^{1,3} **Arthur J. Ragauskas**^{1,3*} (arthur.ragauskas@ipst.gatech.edu), and Martin Keller^{2,3}

¹Institute of Paper Science and Technology, Georgia Institute of Technology, Atlanta; ²Oak Ridge National Laboratory, Oak Ridge, Tenn.; and ³BioEnergy Science Center, Oak Ridge, Tenn.

<http://bioenergycenter.org>

Project Goals: See goals for abstract 60.

Solid state nuclear magnetic resonance (NMR) methods can provide not only chemical information but atomistic structural details that are not easily accessible by other non-destructive high-resolution structural techniques. This makes solid state NMR methodology particularly useful when studying structural problems in biological systems such as the changes occurring in the ultrastructure and supramolecular structure of biomass. Native, pretreated and enzymatically deconstructed biomass was subjected to advanced ^{13}C , ^1H and ^2H 1D and 2D solid state NMR techniques, in which those results were then complemented with carbohydrate and Klason lignin analysis, enzymatic cellulose digestibility and gel permeation chromatography. Unique information about changes in key substrate characteristics, such as crystallinity index, microfibril/microfibril aggregate dimensions, pore size distribution, pore tortuosity, and possible macromolecular connectivities were studied and used to gain insight into the nature of recalcitrance, mechanisms of pretreatment and optimization of biofuel production.

The BioEnergy Science Center is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science.

65

Elucidating the Mechanism of Xylan Biosynthesis: A Biochemical Approach

Maria J. Peña^{1,4*} (mpena@ccrc.uga.edu), Breeanna Urbanowicz,^{1,4} Jason Backe,^{1,4} Alan G. Darvill,^{1,2,4} Malcolm O'Neill,^{1,4} **William S. York**,^{1,2,4} and Martin Keller^{3,4}

¹Complex Carbohydrate Research Center and ²Dept. of Biochemistry and Molecular Biology, University of Georgia, Athens; ³Oak Ridge National Laboratory, Oak Ridge, Tenn.; and ⁴BioEnergy Science Center, Oak Ridge, Tenn.

<http://bioenergycenter.org>

Project Goals (Abstracts 65–67): The BioEnergy Science Center (BESC) is focused on the fundamental understanding and elimination of biomass recalcitrance. BESC's approach to improve accessibility to the sugars within biomass involves 1) designing plant cell walls for rapid deconstruction and 2) developing multitolerant microbes for converting plant biomass into biofuels in a single step (consolidated bioprocessing).

BESC biomass formation and modification research involves working directly with two potential bioenergy crops (switchgrass and *Populus*) to develop varieties that are easier to break down into fermentable sugars. We are using both testing and generation of large number of natural and modified plant samples as well as detailed studies into poorly understood cell wall biosynthesis pathways.

Biosynthesis research provides detailed studies in these and other plant models on key biosynthetic pathways for

lignin, pectin, xylan and cellulose. The coordinated transformation of target genes for validation is in progress.

Xylan is the most abundant hemicellulose in biomass of poplar and switchgrass. This key polymer in cell wall architecture has been proposed to bind cellulose microfibrils and lignin. Modification of xylan structure has the potential to alter cross-linking between wall components, increasing their accessibility by hydrolytic enzymes and decreasing recalcitrance. Despite the biological and economic importance of xylan, its biosynthesis and subsequent incorporation into the wall are still poorly understood.

Our previous studies of *Arabidopsis* mutants showed that several genes that encoded putative glycosyl transferases are required for normal xylan biosynthesis (Peña et al., *Plant Cell* 19:549–563 2007). However none of the proteins predicted to be involved in xylan synthesis have been biochemically characterized. We are using a multifaceted approach to provide insight into mechanisms of xylan synthesis. A series of *Arabidopsis* mutants have been isolated corresponding to genes that encode putative glycosyl transferases, which are highly expressed during secondary growth. The plants have been analyzed to determine the effects of the mutations on plant fitness as well as how the amount, structure, and extractability of xylan synthesized by these plants is altered. Through the use of this approach, we have identified several good candidate genes for the improvement of biomass recalcitrance that influence the structure and quantity of xylan, yet are not critical to plant development. We are now extending this knowledge to alter cell wall structure in *Populus*, a model energy crop.

Furthermore, we have developed an *in vitro* xylan biosynthesis assay using microsomal membranes, fluorescence-labeled oligosaccharide acceptors, and UDP-sugar donors. In-depth structural characterization by NMR spectroscopy of the reaction products allowed us to detect and confirm several glycosyltransferase activities related to xylan synthesis. We are currently using this *in vitro* assay to identify and evaluate biosynthetic chemotypes in our series of *Arabidopsis* mutants with defective secondary wall formation. This system has also been applied to microsomes isolated from *Populus* trees and switchgrass, allowing the mechanisms of xylan synthesis to be compared to what has been observed in *Arabidopsis*. Switchgrass microsomes contain enzymes that, in the presence of UDP-Xyl, also extend the labeled acceptors. Our results suggest that the xylan backbone is extended by a comparable mechanism in monocots and dicots. The results of these studies will be presented their implications for the mechanism of xylan biosynthesis will be discussed.

The BioEnergy Science Center is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science.

66

Immunolocalization of Cell Wall Carbohydrate Epitopes in Switchgrass (*Panicum virgatum*)

Utku Avci^{1,3*} (uavci@ccrc.uga.edu), Michael G. Hahn,^{1,3} and Martin Keller^{2,3}

¹Complex Carbohydrate Research Center, University of Georgia, Athens; ²Oak Ridge National Laboratory, Oak Ridge, Tenn.; and ³BioEnergy Science Center, Oak Ridge, Tenn.

<http://bioenergycenter.org>

Project Goals: See goals for abstract 65.

Plant cells are surrounded by a dynamic extracellular matrix called the cell wall that undergoes modifications in composition and configuration depending on the developmental stage and external factors such as environmental stress or interactions with pathogenic and symbiotic microbes. In order to better understand the cell wall architecture, it is necessary to visualize the components of the cell wall in situ. Cell wall-directed specific monoclonal antibodies are useful tools to examine the distribution of wall polymers in plant cell walls at the whole plant, tissue, cell, and sub-cellular levels. These antibodies can also be used to monitor changes in cell wall composition as a function of plant development and in response to mutational and environmental influences.

A toolkit of monoclonal antibodies (~170) against diverse cell wall polysaccharides structures now exists that includes antibodies recognizing epitopes present in most major classes of wall polymers. These antibodies are available to the research community largely through antibody stock centers: (http://www.crc.uga.edu/~carbosource/CSS_home.html and <http://www.plantprobes.net/>).

The monocot, switchgrass (*Panicum virgatum*), is currently of interest as a potential source of biomass for biofuel production. Here we show that the antibody toolkit can be used to localize wall polysaccharide epitopes in switchgrass leaf blade, leaf sheath, and stem. The epitope localization patterns observed switchgrass are similarly complex to those observed in dicots, such as *Arabidopsis*, although the patterns are different for several antibodies. For example, xylan-directed antibodies label a much broader diversity of cell types in switchgrass than has been observed in *Arabidopsis*. In contrast, homogalacturonan-directed antibodies label only cell corners and cell-cell interfaces in switchgrass, in contrast to *Arabidopsis* where these antibodies label whole cell walls. A detailed understanding of polysaccharide localization patterns has the potential to identify potential targets for either modification or degradation that will lead to more efficient deconstruction of biomass into fermentable monosaccharides for biofuel production.

The BioEnergy Science Center is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science.

67

Changes in Cell Wall Composition and Structure of Alfalfa Reduced Lignin Lines That Might Influence Biomass Recalcitrance

Sivakumar Pattathil^{1,2*} (siva@ccrc.uga.edu), Trina Saffold,^{1,2} Malcolm O'Neill,^{1,2} Jadel Wasson,^{1,2} Jeff Miller,¹ Fang Chen,^{1,3} Richard A. Dixon,^{1,3} William S. York,^{1,2} Michael G. Hahn,^{1,2} and Martin Keller^{1,4}

¹BioEnergy Science Center, Oak Ridge, Tenn.; ²University of Georgia, Athens; ³Samuel Roberts Noble Foundation, Ardmore, Oklahoma; and ⁴Oak Ridge National Laboratory, Oak Ridge, Tenn.

<http://bioenergycenter.org>

Project Goals: See goals for abstract 65.

Genetic modification of biomass crops is a widely used strategy for obtaining less recalcitrant plants with suitably altered cell walls that can be efficiently used for biofuel production. Lignin is one of the plant cell wall components that significantly contributes to biomass recalcitrance. Hence, efforts are underway to obtain biomass crops with reduced levels of lignin in their cell walls. In this regard, two alfalfa (*Medicago sativa*) lines were developed in which two important genes encoding the lignin biosynthetic enzymes, 4-coumarate 3-hydroxylase (C3H) and hydroxycinnamoyl-CoA shikimate/quinate transferase (HCT), were silenced. In depth characterization of the cell wall polysaccharides in these mutants has been carried out using chemical and immunological approaches and compared with walls from wild-type plants. The cell wall preparations were sequentially extracted with a series of solutions of increasing alkalinity and the amount of material extracted in each step was quantified by gravimetric and various colorimetric techniques. The glycosyl residue compositions of the solubilized materials were determined by GLC analysis of the alditol acetate derivatives and by HPAEC-PAD of the underivatized glycoses. The types of polysaccharide that were present in each fraction were also studied using NMR spectroscopic analysis of the per-O-acetylated material. In parallel to these chemical studies, ELISA analyses (Glycome Profiling) of the extracted materials were carried out using a library of cell wall glycan-directed monoclonal antibodies to identify glycan classes eluting in each cell wall extract. Glycome profiles showed that the cell walls from lignin-reduced lines are altered in their extractability when compared to the walls of wild-type plants. For example, chlorite did not release any glucuronoarabinoxylan epitopes from HCT-knock-down cell walls compared to wild-type walls. In contrast, oxalate extraction released glucuronoarabinoxylan epitopes from walls of both reduced-lignin lines that are not released from w.t. walls. Overall, our studies indicate that reduction in lignin biosynthesis leads to a loosening of the wall leading to more facile release of polysaccharides. Cell walls that are more loosely held together may be less recalcitrant to deconstruction for the purposes of biofuel production.

The BioEnergy Science Center is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science.

68

QTLs and Bioinformatically Identified Candidate Genes Underlying Lignin Content and Cell Wall Constituents Are Differentially Expressed in Stem and Root Tissues of *Populus*

G.A. Tuskan^{1,3*} (tuskanga@ornl.gov), R. Priya,^{1,3} T.M. Yin,^{1,3} M. Davis,^{1,2,3} U. Kalluri,^{1,3} X. Yang,^{1,3} S. Jawdy,^{1,3} L. Gunter,^{1,3} X.Y. Zhang,^{1,3} and M. Keller^{1,3}

¹Oak Ridge National Laboratory, Oak Ridge, Tenn.; ²National Renewable Energy Laboratory, Golden, Colo.; and ³BioEnergy Science Center, Oak Ridge Tenn.

<http://bioenergycenter.org>

Project Goals (Abstracts 68-70): The BioEnergy Science Center (BESC) is focused on the fundamental understanding and elimination of biomass recalcitrance. BESC's approach to improve accessibility to the sugars within biomass involves 1) designing plant cell walls for rapid deconstruction and 2) developing multitolerant microbes for converting plant biomass into biofuels in a single step (consolidated bioprocessing).

BESC biomass formation and modification research involves working directly with two potential bioenergy crops (switchgrass and *Populus*) to develop varieties that are easier to break down into fermentable sugars. We are using both testing and generation of large number of natural and modified plant samples as well as detailed studies into poorly understood cell wall biosynthesis pathways.

The *Populus* activity tests existing genetic resources (in association and activation tag studies) as well as the direct transformation of *Populus* for several hundred genes of interest.

Quantitative trait loci (QTL) studies are an integral part of plant research and are used to characterize the genetic basis of phenotypic variation observed in structured populations and inform marker-assisted breeding efforts. These QTL intervals can span large physical regions on a chromosome comprising hundreds of genes, thereby hampering candidate gene identification. Genome history, evolution, and expression evidence can be used to narrow the genes in the interval to a smaller list that is manageable for detailed downstream functional genomics characterization. As a first step in this study, we analyzed the lignin content of 29 cell wall constituents both in stem and root in an inter-specific three generation hybrid poplar pedigree. By establishing a high density genetic map for this pedigree, QTL analyses were conducted to explore the underlying genetic loci. The major findings are as follows; a) Lignin content in stem is significantly higher than that in root, 2) Cell wall constituents

can be classified into four groups, with strong correlation only observed within organs (stem or root), 3) Pleiotropic QTLs are common and 4) QTLs are differentially expressed in stem and root. Our second motivation for the present study was to address the need for a research methodology that identifies candidate genes within a broad QTL interval. Towards this end, a bioinformatics-based approach for subdividing candidate genes within QTL intervals into alternate candidate groups of decreasing probability. Application of this approach in the context of studying cell wall traits, specifically, lignin content and S/G ratios of stem and root in *Populus* plants, resulted in the identification of manageable sets of genes of both known and putative cell wall biosynthetic function. These results provide a roadmap for future experimental work leading to identification of new genes controlling cell wall recalcitrance, and ultimately, in the utility of plant biomass as an energy feedstock.

The BioEnergy Science Center is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science.

69

Transcriptome and Metabolome Profiling of *Populus* Tension Stress Response

U.C. Kalluri^{1,3*} (kalluriudayc@ornl.gov), T. Tschaplinski,^{1,3} X. Yang,^{1,3} S. Jawdy,^{1,3} G. Tuskan,^{1,3} E. Lindquist,² Z. Wang,² and M. Keller^{1,3}

¹Oak Ridge National Laboratory, Oak Ridge, Tenn.; ²DOE Joint Genome Institute, Walnut Creek, Calif.; and ³BioEnergy Science Center, Oak Ridge, Tenn.

<http://bioenergycenter.org>

Project Goals: See goals for abstract 68.

Tension wood is a special type of reaction wood that is formed on upper side of bent stems and branches of angiosperm trees. The wood type is characterized by properties, such as high cellulose content, low lignin content, higher xylem cell number and increased secondary cell wall thickness relative to normal wood. Since these characteristics also constitute desirable feedstock properties, we designed an integrated study to understand the differential molecular and phenotypic properties that underlie tension stress response in *Populus* stems. Here we report results from the GC-MS based metabolite profiling of xylem and phloem tissue samples collected from normal, tension, and opposite wood types. Our analysis reveals several significant quantitative differences in metabolites between wood types (normal, tension and opposite), although the nature of the metabolites present didn't differ appreciably. The nature of the metabolites present differed markedly between genotypes and among tissue types (xylem and phloem). We also report summaries from paired-end read transcriptome data generated using the Illumina platform. This constitutes a large dataset with up to ~5.5 million reads/ library mapped

to the annotated *Populus* transcriptome covering ~77% of predicted gene models in the *Populus* genome version 2.0. The metabolomic data is being assessed in the context of the transcriptome and LC-MS/MS proteome profiles generate plausible hypotheses on mechanisms of tension wood formation and identify new genes involved in enhanced cellulose biosynthesis.

The BioEnergy Science Center is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science.

70

The Use of Metabolomics to Characterize Extreme Phenotypes in a *Populus* Activation-Tagged Population

T.J. Tschaplinski^{1,5*} (t2t@ornl.gov), G.A. Tuskan,^{1,5} U. Kalluri,^{1,5} T. Yin,^{1,5} S. Jawdy,^{1,5} L. Gunter,^{1,5} M. Martin,^{1,5} N. Engle,^{1,5} X. Yang,^{1,5} V. Busov,² S. Strauss,^{3,5} X. Ma,^{3,5} M. Davis,^{4,5} R. Sykes,^{4,5} and M. Keller^{1,5}

¹Oak Ridge National Laboratory, Oak Ridge, Tenn.; ²Michigan Technological University, Houghton; ³Oregon State University, Corvallis; ⁴National Renewable Energy Laboratory, Golden, Colo.; ⁵BioEnergy Science Center, Oak Ridge, Tenn.

<http://bioenergycenter.org>

Project Goals: See goals for abstract 68.

Bioenergy feedstock species provide the raw materials for biochemical conversion into sugars and ultimately liquid transportation fuels. The plant cell walls of feedstock plants contain three major polymers including cellulose (a complex chain of glucose molecules and the most abundant polymer on earth), hemicellulose (a complex mixture of five and six carbon sugars) and lignin (a polyphenolic matrix that protects the cell wall sugars). These three polymers form a tightly bound cell wall matrix that causes raw feedstock material to be highly recalcitrant.

Understanding and overcoming cell wall recalcitrance to improve the yields of fermentable sugars from feedstock plants is a key goal of BESC. More specifically, the *Populus* activity has been tasked with understanding and overcoming recalcitrance in *Populus*, a potential bioenergy feedstock plant. While we understand that complete removal of lignin would be ideal for sugar yields, it would be fatal to the plant. Thus, finding an optimum phenotype with a ratio of lignin to cellulose that results high amounts of sugar *and* viability in the plant is one of our goals. To help achieve this goal we are currently utilizing existing genetic resources including a population of activation-tagged poplar trees established at a field site in Oregon. Gas chromatography-mass spectrometry (GC-MS)-based metabolomics is being used to characterize the clones that exhibit extreme wood chemistry phenotypes.

The BioEnergy Science Center is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science.

71

Visualizing Supramolecular Cell Wall Degrading Enzyme Complexes and Aggregates

Thomas Haas,^{1,3} Bryon Donohoe,^{1,3} Hui Wei,^{1,3} Yunfeng Yang,^{2,3} Martin Keller,^{2,3} **Michael E. Himmel**,^{1,3} and Shi-You Ding^{1,3*} (Shi.you.Ding@nrel.gov)

¹National Renewable Energy Laboratory, Golden, Colo.;

²Oak Ridge National Laboratory, Oak Ridge, Tenn.; and

³BioEnergy Science Center, Oak Ridge, Tenn.

<http://bioenergycenter.org>

Project Goals (Abstracts 71-74): The BioEnergy Science Center (BESC) is focused on the fundamental understanding and elimination of biomass recalcitrance. BESC's approach to improve accessibility to the sugars within biomass involves 1) designing plant cell walls for rapid deconstruction and 2) developing multitolerant microbes for converting plant biomass into biofuels in a single step (consolidated bioprocessing [CBP]).

BESC research in biomass deconstruction and conversion targets CBP by studying model organisms (e.g., *Clostridium thermocellum*) to understand novel strategies and enzyme complexes for biomass deconstruction. We are also searching natural diversity for enzymes and microbes - in particular, thermophilic anaerobes.

BESC is studying and modeling the structure and activities of multi-enzyme complexes to design or screen for variants with better cell wall deconstruction capabilities.

Among the many biomass-digesting microorganisms, a number produce structured biomass-degrading enzyme complexes. These complexes, called cellulosomes, are known to contain a variety of biomass-degrading enzymes docked to structural proteins termed "scaffoldins," which also often contain carbohydrate binding domains. Cellulosomes and their structural and enzymatic components may play important roles in bioenergy production and in future bio- and nanotechnologies. Several cellulolytic members of the bacterial genus *Clostridium* produce cellulosomes. In order to understand the structure and organization of cellulosomes the biomass-degrading properties of these organisms, we have employed electron tomography of high-pressure frozen/freez substituted *C. cellulolyticum* cultures grown on native switchgrass to examine the complex 3-D ultrastructure of the whole, intact cell wall degrading system at 3-5 nm resolution. We also employed immunolabeling techniques and transmission electron microscopy (TEM) to detect two major cellulosome components *in situ*: A processive endocellulase, Cel48F, and a scaffoldin, CipC. Our

observations show that the cellulosome allows *Clostridium cellulolyticum* to employ cell wall deconstruction mechanisms that differ from the mode of action of fungal free-enzyme digestion. Electron tomography has also revealed structural details of the tethers that anchor cellulosomes to bacterial cells.

The BioEnergy Science Center is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science.

72

High Throughput Pretreatment and Enzyme Hydrolysis: A Massively Parallel Approach

Stephen R. Decker* (steve.decker@nrel.gov), Michael J. Selig, Melvin P. Tucker, Mark Davis, Robert Sykes, Kristen Gracom, Michael E. Himmel,^{1,3} and Martin Keller^{2,3}

¹National Renewable Energy Laboratory, Golden, Colo.;

²Oak Ridge National Laboratory, Oak Ridge, Tenn.; and

³BioEnergy Science Center, Oak Ridge, Tenn.

<http://bioenergycenter.org>

Project Goals: See goals for abstract 71.

The complexities of plant cell walls exist on chemical, macrostructural, and ultrastructural levels, rendering a good understanding of recalcitrance a highly daunting undertaking. Understanding how this extensive heterogeneity contributes to cell wall recalcitrance, however, is one of the keystone efforts in the BioEnergy Science Center. Several approaches are in place to tease apart this convoluted puzzle, including efforts in plant breeding, natural variation screening, genetic mapping and sequencing, and *in planta* molecular gene manipulation. While looking for changes in a few mutants or natural variants is very straightforward, migrating the pretreatment, enzyme digestion, and analytical techniques to a massively parallel pipeline capable of handling thousands of small samples each month is exceedingly difficult. Despite the inherent challenges of this undertaking, developing and implementing a high throughput pipeline capable of screening huge libraries of plant variants is one of the key achievements of the BESC. Here we detail the problems and solutions of designing, engineering, building, and implementing the world's first massively parallel biomass pretreatment and enzyme hydrolysis pipeline to measure a realistic biomass recalcitrance phenotype.

The BioEnergy Science Center is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science.

73

Critical Enzymes for Lignin Degradation

Simo Sarkanen^{1*} (sarka001@umn.edu), Yi-ru Chen,¹ Frank Summerfield,¹ Yun-Yan Wang,^{1,3} and Martin Keller^{2,3}

¹University of Minnesota, St. Paul; ²Oak Ridge National Laboratory, Oak Ridge, Tenn.; and ³BioEnergy Science Center, Oak Ridge, Tenn.

<http://bioenergycenter.org>

Project Goals: See goals for abstract 71.

The discovery of lignin depolymerase enzymes that effectively degrade lignins under ambient conditions will contribute to dramatic reductions in pretreatment costs for the saccharification and fermentation of lignocellulose to liquid fuels.

In the early 1980's, it was thought that the initial depolymerization of lignins at the hands of white-rot fungi is primarily caused by hydroxyl radicals. Certainly, hydroxyl radicals can oxidize and cleave lignin macromolecules but, from 1983 onwards, such matters received less emphasis owing to the claim that fungal peroxidases and laccases might exhibit ligninolytic properties. However, for mechanistic reasons that are well understood, these enzymes maintain a poise between cleaving and polymerizing lignin preparations, wherein depolymerization becomes more prominent only at low substrate concentrations. Thus, it has not yet been possible to cleave polymeric lignin preparations completely by enzymatic means *in vitro*, even though extensive degradation is readily achieved with hydroxyl radicals. Consequently, it is not surprising that attention has returned in recent years to the role of reactive oxygen species as agents of fungal lignocellulose degradation.

True lignin depolymerase activity. The availability of the genomes of two white-rot fungi and one brown-rot fungus has now made it possible to consider the protein models with predicted secretion signals from three closely related basidiomycetes. Such comparisons draw attention to particular kinds of enzymes that have not previously been directly implicated in ligninolysis. In comparing the functional secretomes of the white-rot fungi, *Trametes cingulata* and *Phanerochaete chrysosporium*, with that of the brown-rot fungus, *Postia placenta*, it must be remembered how the activities of these basidiomycetes differ toward lignin degradation. White-rot fungi are capable of degrading large proportions of the lignin in lignocellulose completely; however, brown-rot fungal activity results in a marked disappearance of the aliphatic side-chain structures between the aromatic rings that are not themselves cleaved. From this perspective, some new kinds of extracellular fungal enzymes have emerged as likely agents of lignin depolymerization. They are distinguished by their inability, from a mechanistic point of view, to polymerize lignin components, and thus they may be expected to play a prominent role in lignin degradation *in vivo*.

The BioEnergy Science Center is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science.

74

Understanding the Cellulosome and Its Assembly: Towards Improving the CBP Process

Yannick J. Bomble,^{1,5*} (Yannick.Bomble@nrel.gov), **Michael F. Crowley**,^{1,5} Mark R. Nimlos,^{1,5} Markus Alahuhta,^{1,5} Qi Xu,^{1,5} Vladimir V. Lunin,^{1,5} Edward A. Bayer,² John W. Brady,^{3,5} Jiancong Xu,^{4,5} Moumita Sarahay,^{4,5} Jeremy C. Smith,^{4,5} Michael E. Himmel,^{1,5} and Martin Keller^{4,5}

¹National Renewable Energy Laboratory, Golden, Colo.; ²Weizmann Institute of Science, Rehovot, Israel; ³Cornell University, Ithaca, N.Y.; ⁴Oak Ridge National Laboratory, Oak Ridge, Tenn.; and ⁵BioEnergy Science Center, Oak Ridge, Tenn.

<http://bioenergycenter.org>

Project Goals: See goals for abstract 71.

The mechanism by which *C. thermocellum* and other cellulosomal cell wall degrading enzymes assemble on the scaffoldin is not currently known. In our study, we focused on the cellulosome-integrating protein (Cip) A of *C. thermocellum* and key cellulosomal enzymes from families 5, 9 and 48. We developed the first coarse-grained model to study the formation and function of a cellulosome assembly within CHARMM. This study aims at understanding the mechanisms involved in the sequential binding of the cellulosomal enzymes to the CipA scaffold of *C. thermocellum*. Understanding this mechanism is essential in order to design efficient engineered cellulosomes. Also, individual subdomains acting on cellulose surfaces or with individual cellulose chains were studied using molecular dynamics and normal mode analysis. These domains include: catalytic domains, carbohydrate binding domains, fibronectins, and the immunoglobulin-like domain. We have studied extensively the complex, seven-domain family 9 enzyme from *C. thermocellum*, CbhA. From this work, several new protein structures were determined, including CBM4, FNIII2, FNIII3, and CBM3b. Taken *in toto*, these 7 subdomains have enabled novel computational studies providing new insights in the role of this enzyme and the function of its individual subdomains. This new understanding of cellulosome function will be combined with novel mutational strategies to modify the *C. thermocellum* cellulosome to yield superior cell wall degrading *C. thermocellum* strains.

The BioEnergy Science Center is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science.

75

Application of Phylogenomic Techniques in Studying Glycosyltransferase and Glycoside Hydrolase Families

Yanbin Yin^{1*} (yinyb@csbl.bmb.uga.edu), Wen-Chi Chou,¹ **Ying Xu**,^{1,3} and Martin Keller^{2,3}

¹University of Georgia, Athens; ²Oak Ridge National Laboratory, Oak Ridge, Tenn.; and ³BioEnergy Science Center, Oak Ridge, Tenn.

<http://bioenergycenter.org>

Project Goals (Abstracts 75-77): The BioEnergy Science Center (BESC) is focused on the fundamental understanding and elimination of biomass recalcitrance. BESC's approach to improve accessibility to the sugars within biomass involves 1) designing plant cell walls for rapid deconstruction and 2) developing multitolerant microbes for converting plant biomass into biofuels in a single step (consolidated bioprocessing).

BESC researchers in characterization, modeling, education and data management seek to provide integrating data and understanding across the center. We use advanced technologies to analyze chemical and structural changes within biomass. We have developed a coordinated characterization pipeline to provide composition, pretreatment, and enzyme digestibility data from native and modified plant samples and to store this data within a data management system. We are developing knowledge environments to allow access to improved data analysis and modeling tools.

Combining experimental data with biological models will help define the genomic and physical basis of plant cell-wall recalcitrance and deconstruction.

Glycosyltransferase (GT) and glycoside hydrolase (GH) families are enzymes that build and degrade, respectively, carbohydrates and other glycosylated molecules. For example, genes of GT2 family are known to encode cellulose synthases and hemicellulose backbone synthases; and at least 11 GH families are cellulases or glucanases. More than 200 GT and GH families have been categorized by the carbohydrate active enzyme (CAZy) database, corresponding to ~100,000 NCBI-nr proteins. Some of the CAZy families are very huge, for example, the glycosyltransferase family 2 (GT2) consisting of more than ten thousand proteins from various organisms including animals, plants, fungi and bacteria. In addition, many GT and GH families are composed of enzymes with different biochemical functions. Thus it will be valuable to the carbohydrate research field to classify the large CAZy families into smaller subfamilies, ideally, each of them having distinct biochemical function.

We have populated the CAZy families by including homologs from metagenomes and fully sequenced plant genomes. For 211 out of 292 CAZy families which have a Pfam domain/family model to represent them, we have

identified 26,924 homologs from the JGI (Joint Genome Institute) metagenomes and 126,796 homologs from the CAMERA (Community Cyberinfrastructure for Advanced Marine Microbial Ecology Research and Analysis) marine metagenomes. These numbers together already exceed the total number CAZy proteins in the NCBI-nr database. We then combined CAZy proteins from NCBI-nr, metagenomes and fully sequenced plant genomes for each of the 211 CAZy families. The resulting data set of each family was further sub-classified based on sequence similarity and also based on phylogenetic topology; the resulting sub-families were further used to build hidden markov models (HMMs). Some key GT families that are responsible for plant cell wall polysaccharide biosynthesis were examined in further detail, namely the cellulose synthase superfamily (Csl) of GT2 which also includes hemicellulose backbone synthases, putative pectin and xylan synthases related GT8, GT43 and GT47 families etc.

Our study doubled the current CAZy database by including metagenomic CAZy homologs. Our sub-classification of CAZy families into subfamilies which are represented by HMM models provides a new tool to annotate newly sequenced genomes in terms of their CAZy compositions. Moreover numerous metagenome-specific subfamilies were found after the sub-classification, representing novel CAZy subfamilies that are not found in the NCBI-nr database but instead are found exclusively in the environmental metagenomes. These novel CAZy enzyme subfamilies may have new functions that are particularly interesting for bioenergy related researches, for instance, more robust and stronger biomass breaking hydrolases.

The BioEnergy Science Center is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science.

76

Developing Proteogenomics in Plants: Analysis of Proteomics Data Suggests Hundreds of Gene Model Corrections in *Populus*

Andrey A. Gorin^{1,2*} (agor@ornl.gov), Xinran Liu,¹ Grace M. Prazniak,¹ Robert M. Day,¹ Nikita D. Arnold,¹ Tamah Fridman,¹ Udaya C. Kalluri,^{1,2} Gregory B. Hurst,^{1,2} Richard J. Giannone,^{1,2} Robert Hettich,^{1,2} Edward Uberbacher,^{1,2} **Ying Xu**,^{2,3} and Martin Keller^{1,2}

¹Oak Ridge National Laboratory, Oak Ridge, Tenn.;

²BioEnergy Science Center, Oak Ridge, Tenn.; and

³University of Georgia, Athens

Project Goals: See goals for abstract 75.

Mass-spectrometry (MS) based proteomics provides the most direct evidence of gene product presence in specific samples, but MS data interpretation requires sophisticated computational algorithms to uncover the wealth of the encoded information. Our research aims to develop more

robust and reliable identification algorithms for shotgun tandem MS (MS/MS) experiments. More comprehensive proteome characterizations will lead to more complete assessments of functional potential for collected plant samples, as well as, to more confident reconstruction of protein co-expression subnetworks, identification of regulatory signals, etc.

Here we present a study that merges genomics and proteomics bioinformatics methods in order to identify new genes and refine existing gene models in *Populus trichocarpa*—one of the most important target plants for the Bioenergy Science Center research program. The study combined three elements (a) protein database obtained by six-frame translation of the *Populus* genome; (b) a collection of over 900,000 MS/MS spectra obtained from various plant tissues; (c) a novel computational algorithm that at least 2-times outperforms standard database search techniques to reveal 470 novel peptides, which are present in the plant tissue samples, but cannot be explained by the existing gene models. Several layers of control were implemented to ensure complete reliability of the findings: distractor database identifications were checked both at the level of unique peptides and at the level of detected open reading frames.

Confidently identified novel peptides were mapped to the current genome annotation resulting in over 80 predictions of novel genes and over 200 gene models that should have corrections of gene boundaries. Importantly, hundreds of gene model corrections were obtained during re-analysis of the existing data collected for other purposes, at no additional cost for sample preparation or MS/MS runs. Based on this study, the scale of the BESC proteomics effort (millions of spectra per year) represents an unparalleled opportunity to obtain very precise maps of gene positions and the actual proteome of *Populus* species, contributing to both the discovery of important molecular mechanisms (such as factors contributing to recalcitrance) and an improvement of gene annotation algorithms in plants.

The BioEnergy Science Center is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science.

77

Linking Genomic and Biochemical Information to Identify Cellulolytic Enzymes: The GH5 Family Test Case

Brian J. Cantwell^{1,2,4*} (cantwellbj@ornl.gov), Leonid O. Sukharnikov,^{1,2,4} Kathrine Xue,³ Alborz Bejnood,³ **Igor B. Zhulin**,^{1,2,4} and Martin Keller^{2,4}

¹University of Tennessee, Knoxville; ²Oak Ridge National Laboratory, Oak Ridge, Tenn.; ³Oak Ridge High School, Oak Ridge, Tenn.; ⁴BioEnergy Science Center, Oak Ridge, Tenn.

<http://bioenergycenter.org>

Project Goals: See goals for abstract 75.

Tremendous advances in genome sequences make it possible to search for novel biochemical functions beyond the current capabilities of experimental biology. One of the grand challenges in bioenergy research is find better ways of degrading cellulose for its conversion to a biofuel ethanol. Novel cellulose-degrading capabilities can be found in genomic and metagenomic data collected from cellulose-degrading organisms and environments. However, there is a significant problem. Cellulases, enzymes that break down the cellulose polymer into metabolizable substrates, can be found within different families of carbohydrate-hydrolyzing enzymes that represent distinct protein folds. On the other hand, within any cellulose-containing specific fold/family there are closely related enzymes that have different substrate specificity (e.g. mannosidases, xylanases, etc.). This makes it extremely difficult to identify potential cellulases in genomic and metagenomic datasets. The goals of this work is to permit the accurate identification of cellulose degrading enzymes from amino acid sequence information and the prediction of organisms that are likely to efficiently utilize cellulose. Currently, hidden Markov models (HMM) are able to identify glycoside hydrolase domains based on conserved secondary structure, but the exact substrate specificity of these proteins cannot be determined. By combining bioinformatic and phylogenetic techniques with available biochemical information, we aim to improve the classification of potential cellulose degrading enzymes, enable sequence based prediction of substrate use, and identify residues critical to substrate specificity. Here we present a test case: the Glycoside Hydrolase 5 family.

The BioEnergy Science Center is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science.

78

Plant Biomass Deconstruction by Extremely Thermophilic Anaerobes of the Bacterial Genus *Caldicellulosiraptor*

Irina Kataeva^{1,4*} (kataeva@uga.edu), Sung-Jae Yang,^{1,4} Aleksandar Cvetkovic,¹ Farris L. Poole II,^{1,4} Janet Westpheling,^{1,4} Richard J. Giannone,^{2,4} Robert L. Hettich,^{2,4} Sara Blumer-Schuette,^{3,4} Derrick L. Lewis,^{3,4} Inci Ozdemir,^{3,4} Amy L. VanFossen,^{3,4} **Robert M. Kelly,**^{3,4} **Michael W.W. Adams,**^{1,4} and Martin Keller^{2,4}

¹University of Georgia, Athens; ²Oak Ridge National Laboratory, Oak Ridge, Tenn.; ³North Carolina State University, Raleigh; and ⁴BioEnergy Science Center, Oak Ridge, Tenn.

<http://bioenergycenter.org>

Project Goals (Abstracts 78-82): The BioEnergy Science Center (BESC) is focused on the fundamental understanding and elimination of biomass recalcitrance.

BESC's approach to improve accessibility to the sugars within biomass involves 1) designing plant cell walls for rapid deconstruction and 2) developing multitolerant microbes for converting plant biomass into biofuels in a single step (consolidated bioprocessing, or CBP).

BESC research in biomass deconstruction and conversion targets CBP by studying model organisms (e.g., *Clostridium thermocellum*) to understand novel strategies and enzyme complexes for biomass deconstruction. We are also searching natural diversity for enzymes and microbes—in particular, thermophilic anaerobes.

Microbial research targets how the CBP microbe interacts with the features of the pretreated or native plant cell wall and testing strategies for using pure or mixed cultures for degradation.

Very few cultivated microorganisms can degrade lignocellulosic biomass without chemical pretreatment. We have shown that some species of *Caldicellulosiraptor* efficiently utilizes various types of untreated plant biomass, as well as crystalline cellulose and xylan. In the case of *C. bescii* (previously termed *Anaerocellum thermophilum*), which grows up to 90°C, growth substrates included insoluble plant biomass obtained after washing (at 75°C for 18 h) hardwoods such as poplar and high lignin grasses such as switchgrass. The predominant end products from all growth substrates were hydrogen, acetate and lactate. *C. bescii* also grew well on first- and second-spent biomass, where spent biomass is defined as the insoluble growth substrate recovered after the organism had reached late stationary phase. Electron microscopy and growth studies indicate that *C. bescii* attaches dynamically to the plant biomass. The organism has been grown in 600-liter cultures on both crystalline cellulose and on switchgrass and the nature of its extracellular proteins are being investigated, using proteomic and transcriptional analyses. Genome sequences for several *Caldicellulosiraptor* species have been completed, facilitating functional genomics studies aimed at identifying specific ABC transporters for biomass-derived sugars and novel multi-domain glycoside hydrolases that deconstruct cellulose and hemicellulose. Since biomass deconstruction varies considerably across the genome-sequenced members of the genus *Caldicellulosiraptor*, efforts are underway to determine the differentiating features of this novel group of bacteria from this perspective.

The BioEnergy Science Center is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science.

79

Characterizing Cellulose Hydrolysis and Ethanol Production by the Extremely Thermophilic Cellulolytic Organism *Caldicellulosiruptor obsidiansis*

Zhiwu Wang* (wangz2@ornl.gov), Scott Hamilton-Brehm, Jennifer L. Morrell-Falvey, Jonathan Mielenz, Martin Keller, and **James G. Elkins**

Oak Ridge National Laboratory, Oak Ridge, Tenn. and BioEnergy Science Center, Oak Ridge Tenn.

<http://bioenergycenter.org>

Project Goals: See goals for abstract 78.

Biofuel production from renewable lignocellulosic biomass is dependent on the efficient enzymatic hydrolysis of cellulose in order to release fermentable sugars. Cellulolytic microorganisms inhabiting geothermally heated terrestrial hot springs are known to possess multidomain/multifunctional cellulases and hemicellulases that display increased heat-stability. *C. obsidiansis* is a newly characterized extremely thermophilic anaerobe capable of hydrolyzing cellulose, xylan, and pretreated lignocellulosic biomass (switchgrass and *Populus*) while fermenting the resulting sugars into acetate, lactate, CO₂, H₂, and ethanol. In this study, we used confocal laser scanning microscopy and 3-D image reconstruction to monitor the spatial and temporal dynamics of colonization and degradation of cellulose by *C. obsidiansis*. These data indicate that *C. obsidiansis* forms colonies that spread horizontally and vertically after attachment on the cellulose membrane, forming depressions that punctured the substrate within 72 hours. The distance between the cells and the substrate surface played a crucial role in the conversion rate of the cellulose, and thus overall fermentation efficiency. In addition to these modeling efforts, we show that end-product profiles including ethanol can be influenced by applying different growth conditions. By understanding the mechanism of microbial colonization of recalcitrant polymerized sugars, it may be possible to manipulate the fermentation conditions and target genetic modifications to improve the utilization efficiency of substrate carbon to produce desired end-products such as ethanol.

The BioEnergy Science Center is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science.

80

Advances in Organism Development for Consolidated Bioprocessing

David Hogsett^{1,4*} (dhogsett@mascoma.com), Vineet Rajgarhia,^{1,4} Kevin Wenger,^{1,4} Nicky Caiazza,^{1,4} John McBride,^{1,4} Allan Froehlich,^{1,4} William Kenealy,^{1,4} Chris Herring,^{1,4} Heidi Hau,^{1,4} Devin Currie,^{2,4} Adam Guss,^{2,4} Lee R. Lynd,^{1,2,4} and Martin Keller^{3,4}

¹Mascoma Corporation, Lebanon, N.H.; ²Thayer School of Engineering, Dartmouth College, Hanover, N.H.;

³Oak Ridge National Laboratory, Oak Ridge, Tenn.; and ⁴BioEnergy Science Center, Oak Ridge, Tenn.

<http://bioenergycenter.org>

Project Goals: See goals for abstract 78.

Biomass recalcitrance - that is, the difficulty of converting solid biomass to reactive intermediates such as sugars - is the primary obstacle to cost-effective production of ethanol and other fuels from cellulosic biomass. One-step microbial fermentation of pretreated cellulosic biomass without added enzymes, referred to as consolidated bioprocessing (CBP) is increasingly recognized as a potentially game-changing approach by which to overcome biomass recalcitrance. Organism development for CBP can proceed by one of two strategies: improve ethanol production by cellulose-utilizing microbes, and improve cellulose-utilization in microbes that produce biofuels well.

Results will be presented on the development of both eukaryotic and prokaryotic microorganisms capable of directly converting lignocellulosic biomass to ethanol with little or no addition of traditional, commercial cellulases. Data will be presented on the development of yeast strains which express multiple cellulases and are capable of CBP of pretreated materials to ethanol. Additional data will be presented on the development of thermophilic anaerobes for use in CBP, including development of cellulase expression in the pentose-fermenting *Thermoanaerobacterium saccharolyticum* and metabolic engineering of naturally cellulolytic *Clostridium thermocellum* to increase ethanol yields.

The BioEnergy Science Center is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science.

81

A Proteomic Approach to Quantifying the Mass Concentration of Cellulase Enzymes Produced by *Clostridium thermocellum*

Lois St. Brice^{1,4*} (Lois.A.St.Brice@dartmouth.edu), Andrew Dykstra,^{2,4} **Bob Hettich**,^{2,4} Miguel Rodriguez,^{2,4} Babu Raman,^{2,4} Joel Kreps,^{3,4} Lee R. Lynd,^{1,4} and Martin Keller^{2,4}

¹Thayer School of Engineering, Dartmouth College, Hanover, N.H.; ²Oak Ridge National Laboratory, Oak Ridge, Tenn.; ³Verenium, San Diego, Calif.; and ⁴BioEnergy Science Center, Oak Ridge, Tenn.

<http://bioenergycenter.org>

Project Goals: See goals for abstract 78.

To accurately assess rates of microbial cellulose utilization (MCU), it is essential to independently determine cell, cellulase and cellulose mass concentration. Such ability would enable us to directly address questions related to substrate utilization, allocation of cellular resources between cell and cellulase synthesis, cell- and cellulase-specific cellulose hydrolysis rates and bioenergetics. Current methodologies for cellulase determination involve work-intensive purification procedures. In this study we seek to develop and validate a method for cellulase determination which involves minimal manipulation of a fermentation sample. Using proteomic protein determination, we seek to reliably and robustly predict mass concentration of cellulase across varying growth conditions, substrates and cellulase types (cell free vs. cell associated). Our goal was to identify a core group of cellulosomal proteins from *Clostridium thermocellum* which can be assayed using proteomics to determine total cellulosomal protein. Ten proteins that comprise approximately 90% of total cellulosomal proteins have been identified in *Clostridium thermocellum* fermentations. From these ten proteins, 40 peptides have been selected for targeted analysis to determine cellulase mass concentration in cell digest samples. Good candidate peptide sequences were selected to use for quantification based on Mudpits, LTQ and triple quadrupole MS measurements. We analyzed the candidate proteins for variability in the fraction of total cellulosomal mass represented in samples from varying conditions. We optimized peptide selection for these protein components ensuring representative unique peptides with good signal quality for all proteins of interest were selected. We also examined the relationship between proteomic determination of total cellulase based on the peptides selected for analysis and protein determination using conventional protein measurement techniques like the Bradford assay. Once we established the core group, good peptides that could reproducibly be used to quantify them, and the correlation between the protein determination methods, a prediction capability curve was established to determine the total amount of cellulase in a sample.

The BioEnergy Science Center is a U.S. Department of Energy

Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science.

82

Deletion of Cel48S from *Clostridium thermocellum* and Its Affect on Cell Growth and Cellulosome Function

Daniel Olson^{1,2,4*} (dolson@mascoma.com), Shital Tripathi,² Richard Giannone,^{3,4} Nicky Caiazza,^{2,4} David Hogsett,^{2,4} **Lee Lynd**,^{1,4} and Martin Keller^{3,4}

¹Thayer School of Engineering, Dartmouth College, Hanover, N.H.; ²Mascoma Corporation, Lebanon, N.H.; ³Oak Ridge National Laboratory, Oak Ridge, Tenn.; and ⁴BioEnergy Science Center, Oak Ridge, Tenn.

<http://bioenergycenter.org>

Project Goals: See goals for abstract 78.

A method was developed for making targeted gene deletions in *Clostridium thermocellum* and used to delete the *cel48S* gene. This gene was chosen as our initial target because it has been widely studied, it makes up a large portion of the cellulosome and it is believed to play a key role in cellulose solubilization. The resulting *cel48S* mutant strain was analyzed for its ability to grow on crystalline cellulose and was found to digest it as completely as its parent strain, a surprising result. In fact, the only difference was a slight decrease in the rate of growth, and a reduction in biomass production when grown on Avicel. The deletion of *cel48S* is the first report of targeted deletion of a component of the *C. thermocellum* cellulosome.

The BioEnergy Science Center is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science.

83

BioEnergy Science Center Education and Outreach

S. Kral,^{1,4} J. Griffin,¹ H. Schulson,¹ P. Doney,^{2,4} S. Fowler,^{3,4} B. Davison,^{3,4} M. Keller,^{3,4} and **J. Westpheling**^{2,4*} (janwest@uga.edu)

¹Creative Discovery Museum, Chattanooga, Tenn.; ²University of Georgia, Athens; ³Oak Ridge National Laboratory, Oak Ridge, Tenn.; and ⁴BioEnergy Science Center

<http://bioenergycenter.org>

Project Goals: The BioEnergy Science Center (BESC) is focused on the fundamental understanding and elimination of biomass recalcitrance. BESC's approach

to improve accessibility to the sugars within biomass involves 1) designing plant cell walls for rapid deconstruction and 2) developing multitasking microbes for converting plant biomass into biofuels in a single step (consolidated bioprocessing).

BESC researchers in characterization, modeling, education and data management seek to provide integrating data and understanding across the center. We use advanced technologies to analyze chemical and structural changes within biomass. We have developed a coordinated characterization pipeline to provide composition, pretreatment, and enzyme digestibility data from native and modified plant samples and to store this data within a data management system. We are developing knowledge environments to allow access to improved data analysis and modeling tools.

Education and outreach to the general public is critical in the acceptance and deployment of bioenergy. In addition to leveraging successful education and training programs already in place at our partner institutions, BESC has developed educational lessons and activities that target elementary and middle school children.

In addition to our efforts to prepare a new generation of scientists for the emerging fields of bioenergy through the interdisciplinary training of graduate students and postdocs, our center has taken a novel approach in that our education efforts begin with fifth graders. We have developed lesson plans aimed at 4th, 5th and 6th grades to educate and inform students about the basics of energy production and utilization. They include basic concepts such as the carbon cycle, lignocellulosic biomass as substrate for the production of biofuels as well as technical and economic obstacles to a biobased fuel economy. The hands-on activities and guided questions are also designed to meet educational objectives for these grades. These lessons have been piloted in a hundred classrooms in North Georgia and Tennessee and will be made available to schools nationwide in the fall of 2009. We have also begun to pilot interactive “science night” programs offered to students and the general public through local schools, museums and community centers. We will present details of some of the lessons and science night activities.

The BioEnergy Science Center is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science.

Biofuels: Analytical and Imaging Technologies for Studying Lignocellulosic Material Degradation

84

Effects of Pretreatments onto Lignocellulosic Materials as Studied by Raman Microscopy and Mass Spectroscopy

Li-Qiang Chu^{1*} (lchu1@nd.edu), Zhen Li,² Rachel Masyuko,¹ Jonathan V. Sweedler,² and Paul W. Bohn¹

¹Dept. of Chemical and Biomolecular Engineering and Dept. of Chemistry and Biochemistry, University of Notre Dame, Notre Dame, Ind. and ²Dept. of Chemistry and Beckman Institute for Advanced Science and Technology, University of Illinois, Urbana-Champaign

Project Goals: The project is to develop correlated optical (Raman and SH-OCT) and mass spectrometric (SIMS and MALDI MS) imaging approaches for spatial and temporal characterization of lignocellulosic materials at specific processing stages.

The ability to efficiently use lignocellulosic materials (LCMs) to feed the biorefinery of the future depends on high-efficiency pre-enzymatic processing to render lignin separable from cellulose/hemicelluloses. The complex 3-D network structure and chemical characteristics of LCMs pose daunting challenges for imaging and molecular characterization: (1) they are opaque and highly scattering; (2) their chemical composition is a spatially variegated mixture of heteropolymers; (3) the nature of the matrix evolves in time during processing. Presently, there are few *in situ* characterization tools that can be applied to materials with these characteristics, especially during processing—yet acquiring this information is of paramount importance. Here we present a combination of Raman microscopy and mass spectrometric imaging (secondary ion mass spectrometry, SIMS, and laser desorption ionization mass spectrometry, LDI MS) to visualize the structural and chemical changes of LCMs upon various treatments, such as H₂SO₄, NaClO₂, NaOH, etc. *Miscanthus x giganteus* as a model LCM was sectioned into 50 μm thick, and was investigated in this work. Raman and SIMS imaging results indicated that lignin and cellulose are collocated in the cell wall of raw *miscanthus*. A globular structure, composed predominantly of hemicellulose and lignin, is associated with the interior cell wall. Pretreatment of *Miscanthus* using NaOH or NaClO₂ solutions results in the removal of lignin at long processing time. Interestingly, Raman experiments reveal that the H₂SO₄ treated *Miscanthus* exhibits a higher autofluorescence, which might be due to the formation of highly conjugated hydrocarbon species during H₂SO₄ treatment. We are currently correlating these Raman results with mass spectrometric imaging to identify the unknown

species, thereby realizing the full power of correlated optical-mass spectrometric imaging. Together these studies promise to provide a much more complete picture of the effects of various treatments on LCMs.

85

Dynamic Visualization of Lignocellulose: A Biofuels Scientific Focus Area

Barbara R. Evans^{1*} (evansb@ornl.gov), Jeremy C. Smith,^{1,3} Volker Urban,¹ Dean A. Myles,¹ William T. Heller,¹ Hugh M. O'Neill,¹ Art Ragauskas,² Ida Lee,^{1,3} and **Brian H. Davison**¹

¹Oak Ridge National Laboratory, Oak Ridge, Tenn.;

²Institute of Paper Science and Technology, Georgia Institute of Technology, Atlanta; and ³University of Tennessee, Knoxville

Project Goals: Lignocellulosic biomass is recalcitrant to deconstruction and saccharification due to its fundamental molecular architecture and multicomponent laminate composition. A fundamental understanding of the structural changes and associations that occur at the molecular level during biosynthesis, deconstruction, and hydrolysis of biomass is essential for improving processing and conversion methods for lignocellulose-based fuels production. The objective of this research is to develop and demonstrate a combined neutron scattering and computer simulation technology for multiple-length scale, real-time imaging of biomass during pretreatment and enzymatic hydrolysis. Integration and application of the combined capabilities of the Spallation Neutron Source (SNS), the High Flux Isotope Reactor (HFIR) and the National Center for Computational Science (NCCS) at ORNL will provide new information on lignocellulosic degradation at an unprecedented level of detail.

The Oak Ridge National Laboratory (ORNL) Scientific Focus Area (SFA) Biofuels program 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. This program integrates neutron scattering with computational simulation and molecular dynamics to understand the physicochemical processes taking place across multiple length scales during deconstruction of lignocellulosic biomass. These molecular-level methods are assisted and complemented by technical expertise in lignocellulose characterization at the Institute of Paper Science and Technology (Georgia Tech), and chemical force microscopy methods for surface characterization. A multipurpose neutron imaging chamber will be designed and used for in situ, dynamic observation of biomass processing. Deuteration of the biomass crop switch grass and other cellulose sources is being carried out to enable higher contrast between the components of lignocellulosic biomass for neutron scattering that will enable the examination of surface accessibility. These novel technological capabilities are being applied to

specific problems in the pretreatment and enzymatic hydrolysis of biomass to produce the fundamental understanding of plant cell architecture that is needed to develop the next generation of cost-effective cellulosic ethanol production. The interrelated research is organized as three principal tasks: (1) sample preparation and characterization; (2) neutron scattering and diffraction; and (3) computer simulation and modeling.

This overview poster will present the project goals as well as results in sample preparation and characterization. Two companion posters will specifically target the results from neutron scattering and from computer modeling.

The Biofuels Science Focus Area is supported by the Office of Biological and Environmental Research (BER) of the Office of Science (SC), U.S. Department of Energy (DOE), Genomic Science research program.

86

SANS Study of Dilute Acid Pretreatment of Switchgrass

Sai Venkatesh Pingali^{1*} (pingalis@ornl.gov), Volker Urban,¹ William T. Heller,¹ Hugh M. O'Neill,¹ Marcus Foston,² Dean A. Myles,¹ Art Ragauskas,² Barbara R. Evans,¹ and **Brian H. Davison**¹

¹Oak Ridge National Laboratory, Oak Ridge, Tenn. and

²Institute of Paper Science and Technology, Georgia Institute of Technology, Atlanta

Project Goals: Lignocellulosic biomass is recalcitrant to deconstruction and saccharification due to its fundamental molecular architecture and multicomponent laminate composition. A fundamental understanding of the structural changes and associations that occur at the molecular level during biosynthesis, deconstruction, and hydrolysis of biomass is essential for improving processing and conversion methods for lignocellulose-based fuels production. The objective of this research is to develop and demonstrate a combined neutron scattering and computer simulation technology for multiple-length scale, real-time imaging of biomass during pretreatment and enzymatic hydrolysis.

Small-angle neutron scattering (SANS) and wide-angle X-ray diffraction (WAXD) were used to obtain a better understanding of the morphology of the cellulose/lignin composite to aid in understanding and ultimately selecting biomass pretreatment methods that are required to prepare lignocellulosic biomass for conversion to ethanol.

The structural changes that occur during acid pretreatment of switchgrass were investigated by a series of SANS and WAXD experiments. Samples of switchgrass and its component biopolymers were prepared using two different chemical processes: (a) the dilute acid pretreatment method used to break down lignocellulosic biomass and (b) the

extraction treatment for removing one component at a time from the biomass without disrupting its overall structure. The pretreatment, extractions, and compositional analysis of the samples were carried out at the Institute of Paper Science and Technology (Georgia Tech). SANS experiments were carried out with the BIO-SANS instrument at the High Flux Isotope Reactor (ORNL). Dilute acid pretreatment (1) increases the small-scale structure which can be related either to the crystalline core cross-section or pores in the fibrils; (2) decreases in the interconnectivity of the fibrils and forms additional distinct structures at length scales of 100-150 Å that are due to formation of lignin aggregates; and (3) at length scales larger than 1000 Å, does not change the smooth domain boundaries. In contrast, the extraction treatment: (1) produced a smaller increase in the small-scale structure; and (2) did not create an additional structure assigned to the re-precipitation of lignin.

This Biofuels Science Focus Area is supported by the Office of Biological and Environmental Research (BER) of the Office of Science (SC), U.S. Department of Energy (DOE), Genomic Science research program.

87

Examination of Lignin Aggregation by Computer Simulation Integrated with Neutron Experiments: Molecular Dynamics Studies of Lignin

Loukas Petridis^{1*} (petridisl@ornl.gov), Jeremy C. Smith,^{1,3} Benjamin Lindner,³ Roland Shulz,³ Art Ragauskas,² Marcus Folston,² and **Brian H. Davison**¹

¹Oak Ridge National Laboratory, Oak Ridge, Tenn.;

²Institute of Paper Science and Technology, Georgia Institute of Technology, Atlanta; and ³University of Tennessee, Knoxville

Project Goals: Lignocellulosic biomass is recalcitrant to deconstruction and saccharification due to its fundamental molecular architecture and multicomponent laminate composition. A fundamental understanding of the structural changes and associations that occur at the molecular level during biosynthesis, deconstruction, and hydrolysis of biomass is essential for improving processing and conversion methods for lignocellulose-based fuels production. The objective of this research is to develop and demonstrate a combined neutron scattering and computer simulation technology for multiple-length scale, real-time imaging of biomass during pretreatment and enzymatic hydrolysis.

Computer simulation can provide the integration of structural information obtained from multiple imaging and analysis methods that is needed to visualize the molecular structure of lignocellulose and its changes during pretreatment and hydrolysis. The objective is the development and demonstration of computer simulation technology that will provide multi-length scale, real-time

imaging of biomass during pretreatment and enzymatic hydrolysis based on experimental data obtained through neutron science, surface force microscopy, and advanced NMR methods. The integration and application of these combined capabilities will provide new information on lignocellulosic degradation at an unprecedented level of detail.

Computer simulation can complement experimental techniques in gaining an atomic- and molecular-level understanding of the structure and dynamics of lignocellulose. The power of the high-performance computation facilities of the National Center for Computational Sciences (ORNL) enables the development of such dynamic, atomistic models of the very large, heterogeneous molecules that compose lignocellulosic biomass. Here we present how simulation is integrated with Small Angle Neutron Scattering (SANS) to examine the morphology of lignin aggregates in solution. A very brief description of the experimental input that was used to construct the simulation models as well as the derivation of a force field for lignin [1] will be followed by a presentation of the results of our Molecular Dynamics (MD) studies of lignin in solution. The models were built using information on composition, distribution, and location of covalent bonds in lignin from specific sources as determined by state-of-the-art techniques of chemical analysis and ¹³C- and ¹H/²H- NMR combined with chemical modification methods that were carried out at the Institute of Paper Science and Technology (Georgia Tech). The surface morphology and compactness of the lignin aggregates are examined and the results are discussed in the context of SANS experiments using the analogous samples that were carried out with the BIOSANS instrument at the High Flux Isotope reactor (ORNL). Simulation of more complex models that include cellulose and hemicellulose is incredibly computationally demanding because of the sheer size of these models. For this reason, a new strategy [2] is discussed that allows efficient simulation of such large systems on petaflop supercomputers, such as the JaguarXT5 at ORNL. These advances extend the length- and time-scales that can be probed using simulation and as a result microsecond time scale MD of multimillion-atom lignocellulose systems appear now within reach.

References

1. L. Petridis and J.C. Smith. 2009. "A Molecular Mechanics Force Field for Lignin," *J. Comput. Chem.*, 30, 457-467.
2. R. Shulz, B. Lindner, L. Petridis, and J.C. Smith. 2009. "Scaling of Multimillion-Atom Biological Molecular Dynamics Simulation on a Petascale Supercomputer," *J. Chem. Theory Comput.*, 5 (10), 2798-2808.

The Biofuels Science Focus Area is supported by the Office of Biological and Environmental Research (BER) of the Office of Science (SC), U.S. Department of Energy (DOE), Genomic Science research program.

88

Student Presentation

Understanding and Engineering Outer Membrane Protein Export in Gram-Negative Bacteria

Matthew P. DeLisa and Charles Haitjema* (chh65@cornell.edu)

Cornell University, Ithaca, New York

Project Goals: Extracellular secretion of cellulases by engineered bacteria

Escherichia coli K12 is a common host for preparative protein production, however it is limited in its ability to deliver proteins to the extracellular environment. The lack of robust outer membrane (OM) protein translocation machinery presents a bottleneck in the development of bioenergy as cellulosic biomass is not readily transported into the bacterial cell. A recent study identified a protein in *E. coli* called YebF that can efficiently carry recombinant proteins into the culture medium. Here, we identify a novel OM protein translocation mechanism mediated by the *yebEFG* gene cluster that is responsible for delivering YebF and its fusion partners across the OM. We also describe the development of multiple extracellular secretion assays that provide a rapid means to study and engineer the *yebEFG* translocon and type II secretion of gram-negative bacteria.

89

Systems Biology of Cellulose Fermentation to Ethanol: From Domestication of New Organisms to Understanding of Microbial Synergies

Nicolai S. Panikov^{1*} (n.panikov@neu.edu), Maria V. Sizova,¹ Slava S. Epstein,¹ Anthony V. Palumbo,² and Martin Keller²

¹Biology Dept., Northeastern University, Boston, Mass. and ²Oak Ridge National Laboratory, Oak Ridge, Tenn.

Project Goals: Present communication is based on preliminary results of research supported by DOE (Grant Number: ER64507, "To Bioethanol through Genomics of Microbial Synergies"). The strategic goal of this project is to facilitate development of cellulosic ethanol technology by achieving better understanding of lignocellulose degradation in nature: in anaerobic biocompost, anoxic soils and sediments. We believe that industrial consolidated bioprocessing of fuel from various feedstocks cannot be realized by a single 'superbug', even after profound metabolic engineering. More feasible solution would be to construct a set of specialized microbial consortia each one adapted to particular feedstock.

Emerging bioethanol technology uses narrow range of fermenting organisms from genera *Clostridium* and *Caldicellulosiruptor*, the *C. thermocellum* being the most carefully studied (Lynd et al., 2002). New organisms will likely come from the pool of as yet uncultivated species accounting for up to 99% of the natural microbial diversity. We have recently shown that such species could be recovered by using in situ incubation devices (Kaeberlein et al., 2002). We have demonstrated also that 'uncultivable' species can be grown in standard media in the presence of microbial helpers, deliberately added organisms which provide signaling metabolites, siderophores or other extracellular stimulatory factors to a 'difficult' microbe (Nichols et al., 2008). Here we report on further development of this approach combined with culture-independent techniques and dynamic mathematical simulations as applied to particular case of cellulose and lignocellulose digestion/fermentation.

Methods. Anaerobic cellulose degradation and fermentation were followed in situ as well as in laboratory incubation experiments with cellulose-amended soil and biocompost samples incubated under constant environmental conditions (temperature, moisture, gas flow). To improve recovery of natural cellulolytic organisms, we used cellulose-traps followed by standard serial dilution optimized for recovery of strict anaerobes. The isolated consortia and pure cultures were tested for their degrading activity by using high throughput screening system based on continuous off-gas analysis by IR- and mass-spectrometry. 16S rRNA survey of the isolates, consortia and natural communities (after cloning) provided taxonomic identification and assessment of degree of uncertainty: how many species (OTU's) involved in cellulose degradation remained uncultivable. Finally, the active strains were grown on suspended cellulose or pretreated wood under full fermentation control (pH, mixing intensity, temperature, red-ox conditions) and computer-aided instrumental monitoring of residual substrate, cell mass, fermentation products and base titration rate (Panikov and Lynd, 2010)

Results: Stable cellulose-degrading consortia were obtained in most tested soils and biocomposts. The typical fermentation products in enrichments and stabilized consortia were CO₂, H₂, CH₄, acetate, ethanol and lactate. In some cases, acetate was replaced with formate. At high sulfate concentration (some soils and sediments), cellulose degradation was suppressed, probably because of toxic by-products formed by sulfate-reducing bacteria. The methanogenic communities grew slow (specific growth rate 0.05–0.1 h⁻¹ at 55°C) with sustained oscillations of fermentation rate (the effect of reversible product inhibition).

More than 15 pure cultures of novel organisms related to *C. clariflavum*, *C. straminisolvens* and *C. thermocellum* have been isolated. *C. clariflavum* was able to ferment cellulose, xylan and their mixture as well as pretreated wood into ethanol, formate, CO₂ and H₂. Other isolates degraded only cellulose with acetate as end product. New strains varied in respect to growth and maintenance rates, enzymes localization (free and cell-bound), sensitivity to product inhibition as well as to starvation and O₂-stress (see illustration below).

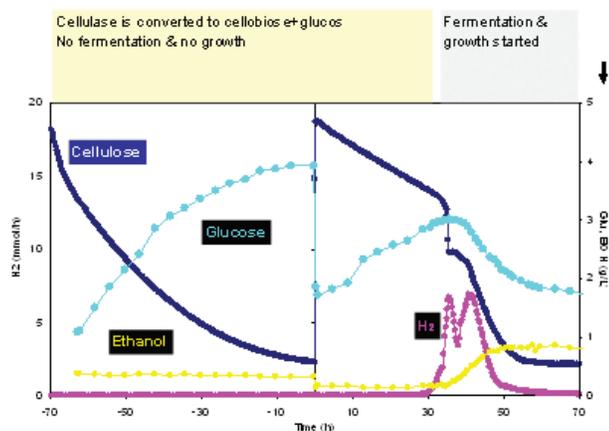


Fig 1. Example of starvation stress. Culture of cellulolytic bacteria *C. thermocellum* enters non-growing state after short-term starvation. First addition of cellulose-mineral medium (time -70 h) does not result in a normal cell growth: extracellular cellulases remaining from the past growing period decompose cellulose into glucose and cellobiose, but cells are unable to consume released sugars. At time zero (\downarrow), the second addition of fresh medium induces growth after another 30 h of latent phase. The growth and fermentation start abruptly and proceed until depletion of cellulose (but not glucose).

Mathematical simulation of community dynamics.

Growth of communities and consortia as well as pure cultures was simulated by structured dynamic model based on high-order set of non-linear ordinary differential equations (Panikov, 1995; 2008). Model takes into account differential gene expression in delayed response to concentration of limiting substrate orchestrated by transcription factors. The simulation of community was possible with aggregated model containing linear approximation of the vector of intracellular polymeric constituents. We are testing the validity of two basic ecological concepts on the nature of microbial cellulolytic community. The first concept identifies a community as a super-organism with firm internal interactions between individual populations stemming from the metabolic stoichiometry of decomposition network and regulatory effects of signaling metabolites. The second, continuum paradigm allows relative freedom for members to enter or leave community dependent upon their success in acquiring limiting nutrient resources. The competitive advantage of each population depends on inherited growth characteristics (growth rate, colonization potential, yield, affinity of transporters, maintenance, stress-tolerance).

Conclusion: Preventing starvation stress in industrial strains seems to be as important as the level of enzymatic activity. We discuss the ways to improve productivity and robustness of fermenting *Clostridia* by selection and metabolic engineering.

References

1. Kaerberlein, T., K. Lewis and S.S. Epstein (2002). Isolating "uncultivable" microorganisms in pure culture in a simulated natural environment. *Science* 296 (5570): 1127-1129.

2. Lynd, L.R., P.J. Weimer, W.H. van Zyl and I.S. Pretorius (2002). Microbial cellulose utilization: fundamentals and biotechnology. *Microbiol Mol Biol Rev* 66 (3): 506-577, table of contents.
3. Nichols, D., K. Lewis, J. Orjala, S. Mo, R. Ortenberg, P. O'Connor, C. Zhao, P. Vouros, T. Kaerberlein and S.S. Epstein (2008). Short Peptide Induces an "Uncultivable" Microorganism To Grow In Vitro. *Appl. Environ. Microbiol.* 74 (15): 4889-4897.
4. Panikov, N.S. (1995). *Microbial Growth Kinetics*. Chapman and Hall. 378 p.
5. Panikov, N.S. (2008). Kinetics, Microbial Growth. In: Michael C. Flickinger and Stephen W. Drew (Eds.) *Encyclopedia of Bioprocess Technology: Fermentation, Biocatalysts and Bioremediation*. New York: John Wiley & Sons, Inc.: 1513-1543.
6. Panikov, N.S. and L.R. Lynd (2010). Physiological and Methodological Aspects of Cellulolytic Microbial Cultures. In: *Manual of Industrial Microbiology and Biotechnology*, 3rd Ed. (Baltz, Davies, and Demain, eds.)

90

A New Solution-State NMR Approach to Elucidate Fungal and Enzyme/Mediator Delignification Pathways

Kenneth E. Hammel^{1*} (kehammel@wisc.edu), Daniel J. Yelle,¹ John Ralph,² Dongsheng Wei,¹ Alexander N. Kapich,¹ and Daniel Cullen¹

¹USDA Forest Products Laboratory, Madison, Wis. and ²Dept. of Biochemistry, University of Wisconsin, Madison
<http://www.fpl.fs.fed.us/>

Project Goals: New methods for solution-state NMR spectroscopy of lignocellulose are being used, in conjunction with enzymology and molecular biology approaches, to determine how filamentous fungi cleave lignin and what agents they employ to accomplish this chemistry.

Brown rot basidiomycetes remove cellulose from wood efficiently, even though this sugar polymer is initially shielded by a biochemically recalcitrant barrier of lignin. During this process, the lignin appears to remain in situ, which raises the question of how the polysaccharide-degrading systems of brown rot fungi circumvent the lignin to access their substrates. New results based on solution-state NMR analysis of ball-milled, dissolved, brown-rotted wood are now available to clarify this picture. ¹H-¹³C HMBC spectra of aspen degraded by the brown-rotter *Postia placenta* showed that the lignin sidechains had been cleaved between C_α and C_β, yielding new benzoic acid and benzaldehyde residues in the polymer. In addition, arylglycerol-β-aryl ether linkages had been cleaved in the lignin to generate new phenylglycerol residues, as shown by three-dimensional ¹H-¹³C HSQC-TOCSY spectra. The HSQC results, in conjunction with quantitative ¹³C NMR spectroscopy, indicated that roughly 6% of the monomeric units in the residual lignin were cleaved structures. Our results show that *P. placenta* is ligni-

nolytic, contrary to the prevailing view of brown rot. Since this fungus lacks ligninolytic peroxidases, it is also clear that some other mechanism is responsible for its ability to cleave lignin. Results to date suggest that reactive oxygen species generated via extracellular oxidation of a fungal metabolite may be the responsible oxidants: (a) The wood colonized by the fungus contained a laccase that is encoded in the *P. placenta* genome. (b) The biodegrading wood contained a fungal metabolite, 2,5-dimethoxyhydroquinone, and also Fe^{3+} as its oxalate complex. (c) Heterologously expressed *P. placenta* laccase oxidized 2,5-dimethoxyhydroquinone with concomitant production of perhydroxyl radicals, which are known initiators of hydroxyl radical production in the presence of Fe^{3+} complexes.

91

Real-Time Chemical Imaging of *Clostridium cellulolyticum* Actions on *Miscanthus*

Hoi-Ying N. Holman* (hyholman@lbl.gov), Zhao Hao, Ping Hu, and Andrew N. Tauscher

Lawrence Berkeley National Laboratory, Berkeley, Calif.

Project Goals: Our purpose is to perform technology research and development, and to apply the technology to develop a comprehensive picture of actions of cellulolytic microorganisms on natural plant biomass, which will help elucidating the important processes of different temporal and spatial scales underlying the microbial destruction of plant biomass for a cost-effective production of biofuel.

In nature, microorganisms are important agents in the cycling of elements. Those that can hydrolyze cellulose rapidly may come to play an important role in carbon cycling and breaking the barriers to cost-competitive production of cellulosic ethanol. Microbe-induced cellulose hydrolysis is generally a slow and incomplete process. However, many microorganisms among *Clostridia* species have been linked to elevated rates of cellulose hydrolysis in compost and landfills. Cellulolytic action by *Clostridia* sp. is facilitated primarily at the surface of cellulosic materials. *C. cellulolyticum* is a mesophilic anaerobic bacterium. The wealth of information on cellulosomes, genomics and carbon flux in *C. cellulolyticum* made it a prime model system for understanding microbial strategies in biofuels processing. We have developed synchrotron radiation-based Fourier transform infrared (SR-FTIR) spectromicroscopy which has enabled us to non-invasively make molecular measurement and images of *C. cellulolyticum* interactions with cellulose substrates, as well as plant materials. Our SR-FTIR results show that even in the simple cellulose system, the surface chemistry is quite variable spatially at scales that range from a fraction of a micron to hundreds of microns, with concentrated features locally. However, nearly all the kinetic studies within the last two decades were conducted in batch cultures, or in continuous cultures, or in chemostat cultures. These results point to the importance of physiochemical parameters at a microscopic level under relatively uniform and dilute conditions.

We are extending these observations with an improved SR-FTIR approach, to explore *C. cellulolyticum* actions on *Miscanthus*, a natural perennial plant that grows as tall as 13 feet with little to no fertilizer, and can be conveniently stored for an almost indefinitely time period. *Miscanthus* shows promises for more efficient biofuel production. Therefore, we use *Miscanthus* as the lignin-cellulose substrates in this study. The destruction of *Miscanthus* will be followed in real time by SR-FTIR chemical imaging. The controls will include known enzymes on various carbohydrate polymers. Our aim is to develop a comprehensive picture of actions of cellulolytic microorganisms on natural plant biomass, which will help elucidating the important processes of different temporal and spatial scales underlying the microbial destruction of plant biomass for a cost-effective production of biofuel.

92

Multi-Mode Real-Time Chemical Imaging as a Systems Biology Approach to Decipher Microbial Depolymerization of Lignocellulose

Jamie Cate^{1,2*}, Hoi-Ying N. Holman¹ (hyholman@lbl.gov) and Haw Yang³

¹Lawrence Berkeley National Laboratory, Berkeley, Calif.; ²School of Chemistry, University of California, Berkeley; and ³Dept. of Chemistry, Princeton University, N.J.

Project Goals: We are developing multi-mode real-time imaging methods to probe at the chemical level plant biomass depolymerization by living cellulolytic microbes. This integrated imaging effort will provide unique insights into the highly complex physical and chemical transformations that occur during the depolymerization problem. The integrated system should also prove useful for studying the fundamental chemical processes of other energy conversion technologies, such as next-generation solar energy conversion devices and fuel cells.

We are developing multi-mode real-time imaging methods to probe at the chemical level plant biomass depolymerization by living cellulolytic microbes. In nature, some microbes use suites of enzymes to break down the highly heterogeneous solid substrates present in plant biomass; others convert them to biofuels like ethanol. The mechanisms of these actions require surface chemistry, since the plant biomass substrates are solids. It is imperative, therefore, to study not only the enzymes, but also the properties of the substrates as they are degraded. Furthermore, the production of enzymes is affected by the physiological states of the microorganisms, which can be altered by the metabolites and end products including ethanol because of their inherent toxicity. To understand this dynamic system of biomass depolymerization for cellulosic ethanol production, we need to approach it at a systems biology level. Due to the highly complex nature of the substrates and enzyme mixtures, we think that a systems biology approach should be considered

in broader terms. It should include multi-mode chemical imaging methods.

To this end, a “grand challenge” is the acquisition of integrated knowledge on multiple time and length scales. We develop and use both single-molecule imaging of enzyme dynamics and Fourier transform infrared spectroscopy of solid substrates with living cellulolytic bacteria to probe plant cell wall depolymerization as a function of space and time. For example, using our newly developed real-time 3D single-particle tracking (RT-3DSPT) spectromicroscopies for single-molecule spectroscopy and imaging, we will examine cellulase and cellulosome processivity and cooperativity in the degradation of lignocellulose. These experiments will exploit the genetic tools that we are developing to incorporate fluorescent tags into the enzyme and enzyme complexes to enable tracking. We will also exploit the high temporal and spatial resolution of synchrotron radiation-based Fourier transform infrared (SR-FTIR) spectromicroscopy to follow changes in the chemical composition of the cell wall substrates as they are depolymerized (see our second poster, Holman et. al.). To obtain a truly quantitative understanding, we found that the heterogeneity inherent in biomass depolymerization ultimately cannot be tackled with separate measurements that occur at different times or on different, albeit similar, samples. This integrated imaging effort will provide unique insights into the highly complex physical and chemical transformations that occur during the depolymerization problem. The integrated system should also prove useful for studying the fundamental chemical processes of other energy conversion technologies, such as next-generation solar energy conversion devices and fuel cells.

93

In Vivo Mapping of ROS Gradients Produced by Wood Decay Fungi during Early Colonization

Christopher G. Hunt^{1*} (cghunt@fs.fed.us), **Kenneth E. Hammel**^{1,2} (kehammel@wisc.edu), and Don Carl Jones¹

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

http://www.fpl.fs.fed.us/research/units/performance_personel.shtml
<http://www.bact.wisc.edu/faculty/hammel/index.php>

Project Goals: In this project we have developed a modular system to image and quantify ROS (or other metabolite) from fungi in their natural substrate with minimal disturbance. We are using this knowledge of ROS concentrations to understand the mechanisms at work during incipient decay.

Background

Wood decay fungi are successful in removing carbohydrate from wood, despite the presence of lignin. We hope to improve prospects for woody biomass saccharification by understanding the mechanisms of fungal wood decay.

To make lignocellulosics permeable to enzymes, filamentous fungi use a variety of small, diffusible reactive oxygen species (ROS) such as hydroxyl radicals, peroxy radicals, and possibly phenoxy radicals. These radicals diffuse into the cell walls and initiate biodegradative radical reactions. When lignin is the target, radical attack results in various extents of oxidation and depolymerization.

These small diffusible oxidative species are important tools used by filamentous fungi to make the cell wall accessible to enzymes. Despite this, we have a poor knowledge of how these oxidants are spatially distributed in biodegrading lignocellulose relative to the fungal hyphae that produce them. The goal of this project is to remedy this deficit through fluorescence microscopy of newly designed sensors that will serve as in situ reporters of biodegradative radical production. We will use these sensors to produce oxidative maps that will help us to understand how fungi generate ROS and how they use these ROS to make cell walls more accessible to enzymes.

Method

We are placing fluorescent beads in wood at the start of fungal decay, and then imaging the beads after the fungus has colonized the wood a few days later. Our images can tell us the local concentration of oxidants as well as an overlay with the location of hyphae.

The strategy of covalently attaching fluorescent dyes to silica beads has many advantages. We design the bead to emit two fluorescent signals, so that the ratio of the two signal intensities provides quantitative information. Immobilized dyes are prevented from moving after reaction, so partitioning is impossible. In addition, they cannot be ingested, and the fluorescence from the dye is clearly distinguishable from background.

Our first bead has BODIPY 581/591[®] on a 3 μ m porous HPLC bead. This dye’s emission changes irreversibly from red to green upon oxidation by ROS. The ratio of red to green emission provides a quantitative measure of the cumulative oxidation at that point in space. Dyes with reactivity to specific ROS, pH, or other metabolites of interest are envisioned.

Results

We showed that our oxidant detection system is tied to wood decay by comparing bead oxidation in with the wood decay fungus *Phanerochaete chrysosporium* to the oxidation from a wood inhabiting fungus that does not degrade wood, *Ophiostoma piliferum*. The decay fungus cause much more oxidation, and also ate holes completely through the wood sections if left to incubate for a month.

By observing the oxidation of beads around a hyphal tip, we have concluded that oxidation occurs gradually over time. The extracellular enzymes to create ROS are typically excreted from the hyphal tip, but the enzymes do not immediately oxidize the wood. The enzymes make low molecular weight diffusible ROS which attack wood. These enzymes continue to operate over hours and days, continually creating ROS which gradually degrades cell walls.

While the current generation of beads is not sensitive enough to observe oxidation in the first few hours after the passing of a hyphal tip, three day old cultures show oxidation gradients around almost every hypha we investigated (see below). We expect these gradients to tell us about the relative rates of reactivity and diffusion for the oxidative species in the culture.

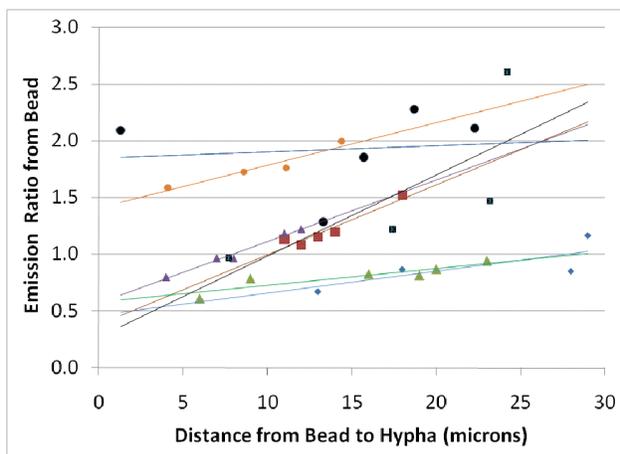


Figure 1: Fluorescent emission from individual beads as a function of distance from hypha. Lower Y axis values indicate more oxidation.

Calibration of the beads was done by measuring the oxidation of the beads after incubation with different concentrations of a free radical initiator in cultures. From this calibration, we can estimate the number of oxidant molecules produced per time for given conditions.

This work was supported by the U.S. Forest Service and the U.S. Department of Energy Office of Science Office of Biological and Environmental Research

94

Integrated Nondestructive Spatial and Chemical Analysis of Lignocellulosic Materials during Pretreatment and Bioconversion to Ethanol

Gary Peter* (gfpeter@ufl.edu), Choong Heon Lee, Kyle Lundsford, Steve Blackband, Lonnie Ingram, and Richard Yost

University of Florida, Gainesville, Florida

Project Goals: Our *long-term goal* is to develop a quantitative structural model for changes that occur in the organization and chemical composition of plant biomass during pretreatment, enzymatic degradation and bioconversion to ethanol or other products. The *objectives* of this proposed work are to 1) use advanced high resolution magnetic resonance microscopy (MRM), micro and nano x-ray computed tomography (x-ray CT), electron microscopy and imaging mass spectrometry (IMS) to quantify

changes in the architecture, porosity, permeability, surface area, pore size, interconnectivity, chemical organization and composition of bagasse and particularly *Populus* and pine wood chips during pretreatment and enzymatic degradation, and 2) integrate the results from these different quantitative imaging methods into a model for disassembly of the plant cell wall during pretreatment and bioconversion. We are developing methods for imaging biomass with MRM, x-ray CT and IMS.

IMS with a MALDI linear ion trap + MS: The full-scan MS of biomass shows intense ions at every mass-to-charge ratio (m/z), making the analysis very complex. The use of the linear ion trap (LIT) and MSⁿ fragmentation are required to interpret the complex spectra and map the distribution of cellulose, lignin and hemicelluloses within the biomass during pretreatment and hydrolysis. Because so many ions are present, we analyzed standard compounds that are normally present in wood. Full-scan and MSⁿ spectra were obtained for beta 1,4-glucan, 4-O-methylglucouronxylan, β -glucan, starch and microcrystalline cellulose. The complex carbohydrates typically present in lignocellulosic biomass yield oligomeric fragments of characteristic ionization patterns. Interestingly, starch fragmented and ionized very differently than microcrystalline cellulose. Figure 1 shows typical spectra from microcrystalline cellulose (top) and birch beta 1,4-glucan, 4-O-methylglucouronxylan (bottom).

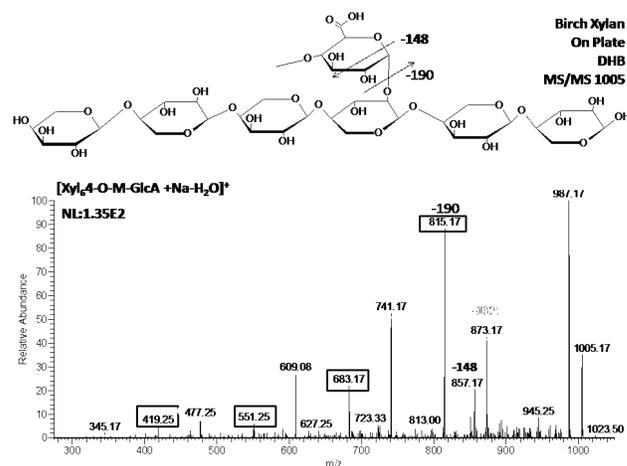
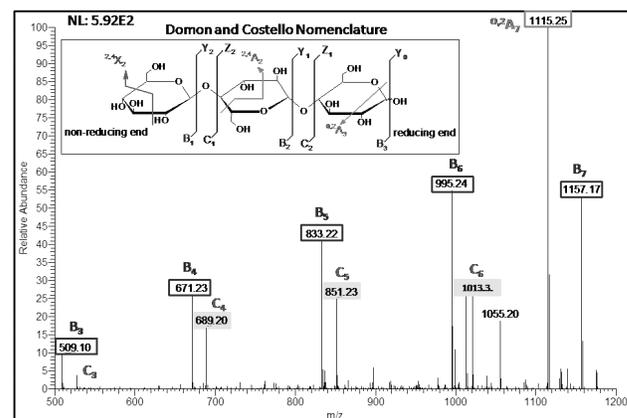


Figure 1

IMS images with radial sections of poplar wood pretreated with dilute acid at 145°C created by extracting the m/z 1175 derived from cellulose shows inconsistent patterns suggesting microheterogeneity in release of the ion.

Magnetic Resonance Microscopy: Excellent image quality is obtained from *Populus* wood and bagasse samples using T2 and diffusion weighted modes. In T2 images, vascular bundles appear dark consistent with the knowledge that lignified cells contain limited free water. A chemical shift has been found in some samples and the cause for this is under investigation. Image quality is quite comparable to optical microscopy. T2 weighted MRM images acquired at 39 μM resolution of *Populus* wood chips treated with mild acid and 145°C, conditions considered more typical of pretreatments considered commercially viable, were obtained. Small differences were observed comparing untreated with pretreated wood, even though chemical analyses show that the xylan was quantitatively removed and the wood clearly becomes more brittle and is substantially softer in the 2 and 3% acid treatments at 145°C as expected. We hypothesize that the resolution of the instrument using larger rf coils is too low to quantify the relatively subtle differences in structure induced by pretreatment. New images were collected with state of the art rf microcoils at 8 μM resolution. Figure 2 shows the improved signal to noise ratios provide much better images.

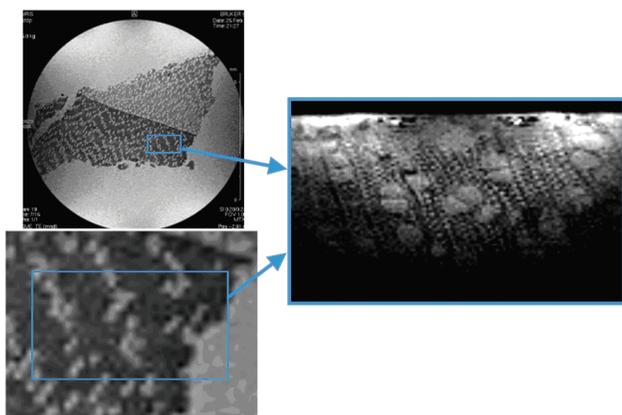
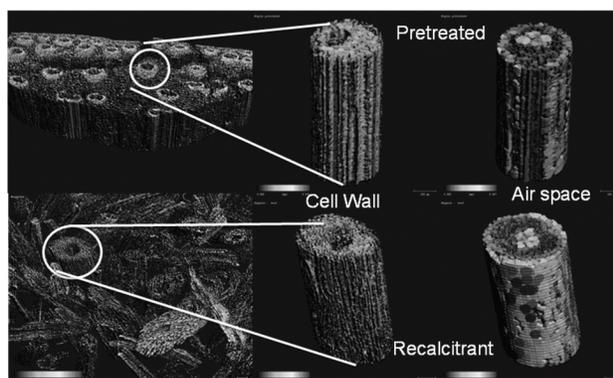


Figure 2. Illustration of the microcoil SNR improvement. On the left is a conventional 39 micron resolution image (top) with an expanded view (bottom). Using a microcoil an 8 micron image (right) with an equivalent field of view (indicated by the boxes) is shown – the clear improvement in the ability to see the wood microstructure is evident.

X-ray micro CT: Excellent images have been obtained at high resolution from *Populus*, pine, and bagasse samples. In addition to the basic density, images are readily segmented and the material and airspace sizes can be quantified. Small changes in surface area and surface area to volume ratios were observed after dilute acid pretreatment. Analysis of the recalcitrant material left after steam gun pretreatment and simultaneous saccharification and fermentation from our pilot facility shows that the recalcitrant material was mainly the lignified and dense vascular bundles as expected. Micro CT imaging of the recalcitrant material shows thinner cell

walls and some degree of degradation on the periphery compared with the internal regions of the bundles.



One constraint of CT imaging is that for quantification samples need to be dried and the wood shrinks by ~10% in the radial and tangential planes, thus dried wood measurements underestimate those in wet wood. To overcome this limitation we are exploring the use of nanoparticle contrast agents designed for CT imaging.

95 Cell Wall Assembly and Deconstruction Revealed through Multi-Platform Imaging in the *Zinnia elegans* Model System

Catherine I. Lacayo¹ (lacayo2@llnl.gov), Alexander J. Malkin,¹ Hoi-Ying N. Holman,² Shi-You Ding,^{3,4} Mona Hwang,¹ and Michael P. Thelen^{1*} (mthelen@llnl.gov)

¹Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory, Livermore, Calif.; ²Ecology Dept., Lawrence Berkeley National Laboratory, Berkeley, Calif.; ³National Renewable Energy Laboratory, Golden, Colorado; ⁴BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tenn.

Project Goals: To improve our knowledge of the native architecture of the plant cell wall and assess applications for efficient deconstruction, using a combination of imaging approaches, including atomic force microscopy (AFM), fluorescence microscopy, and synchrotron radiation Fourier-transform infrared spectroscopy (SR-FTIR).

With the ambition of manufacturing alternative fuels, the biotechnology industry has turned to the plant cell wall, a source of fermentable sugars, which can become the starting material for biofuel production. However, detailed changes in plant cell walls in response to chemical, enzymatic and microbial treatments have not been monitored at high resolution. Our project seeks to understand more about the structural organization of the cell wall and how it can be efficiently deconstructed. For this purpose, we are imaging single cells from *Zinnia elegans* that have been induced in culture to develop into tracheary elements (TEs), individual

components of xylem tissue. Mature TEs develop large secondary cell wall thickenings that are deposited underneath the primary cell wall and are rich in lignocellulose. We have imaged TEs using a variety of platforms, including atomic force microscopy (AFM), fluorescence microscopy and synchrotron radiation based Fourier-transform infrared spectroscopy (SR-FTIR). Our approach of imaging the ultrastructure of the cell wall at nanometer scale, coupled with the capability to reveal the corresponding chemical composition, can profoundly improve the fundamental understanding of the native architecture and mechanisms of deconstruction of the plant cell wall.

To probe the cell wall for the presence of specific polysaccharides, we used fluorescently-tagged carbohydrate binding modules (CBMs) from *Clostridium thermocellum*. After treating TEs with oxidative chemicals to remove lignin, we observed a dramatic increase in fluorescence using CBM3, a family 3 CBM that binds to crystalline cellulose. This increase in fluorescence suggested that that cellulose was more accessible or likely to bind to CBM3 following chemical treatments.

When we imaged the surface of *Zinnia* TEs by AFM, we observed that these were covered with pronounced granular structures, ranging in size from approximately 20 to 100 nm. After oxidative treatment, we found that this surface granular material was absent and that the underlying meshwork of cellulose fibrils (ranging in width from 10 to 20 nm) from the primary cell wall had become exposed. This result corroborated the increased physical accessibility of cellulose in the cell wall after oxidative treatment. When pre-treated TEs were examined by SR-FTIR, we found that their chemical composition changed significantly.

To examine secondary cell wall ultrastructure, we found that physical disruption of TEs using mild sonication was sufficient to produce cell fragments that were conducive to AFM imaging. We focused on discrete ring-like secondary wall structures, which revealed cellulose fibrils decorated with particles and arranged in parallel bundles. Chemical treatments generally removed particles from these cellulose bundles.

We are currently developing experimental techniques to structurally and chemically probe the dynamic response of *Zinnia* TEs to enzymatic and microbial degradation of lignocellulose. We anticipate that our imaging-based studies will help elucidate mechanisms of cell wall degradation and improve models of the organization and composition of the plant cell wall.

This work was funded by the U.S. Department of Energy GTL Program project number SCW0738 and performed under the auspices of the DOE by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. LLNL-ABS-421248.

96

Label-Free, Real Time Monitoring of Biomass Processing with Stimulated Raman Scattering Microscopy

Brian G. Saar^{1*} (saar@fas.harvard.edu), Yining Zeng,² Christian W. Freudiger,^{1,3} Yu-San Liu,² Michael E. Himmel,² Shi-You Ding,² (Shi.you.Ding@nrel.gov) and **X. Sunney Xie**¹ (xie@chemistry.harvard.edu)

¹Dept. of Chemistry and Chemical Biology, Harvard University, Cambridge, Mass.; ²Chemical and Biosciences Center, National Renewable Energy Laboratory, Golden, Colo.; and ³Dept. of Physics, Harvard University, Cambridge, Mass.

Project Goals: (a) To develop novel optical imaging technology based on coherent Raman scattering that is capable of real time, label-free chemical imaging and (b) to apply these techniques to image the process of biomass conversion to biofuels in real time in three dimensions in intact plant tissue. The new information available from these techniques will offer insight into this complex series of chemical reactions and help to better understand and optimize their efficiency.

The conversion of plant biomass into “cellulosic” ethanol is an alternative energy technology that has attracted significant research interest over the past decades and requires new tools to understand and optimize the conversion process. We demonstrate that stimulated Raman scattering (SRS) microscopy can be used to selectively map plant cell wall polymers, such as lignin and cellulose, simultaneously, at sub-micron spatial resolution and with linear concentration dependence and high speed. We then follow the acid chlorite delignification process, to further demonstrate the real-time imaging of lignin bleaching with a time resolution of a few seconds. SRS microscopy is a high sensitivity, label-free chemical imaging technique, and provides a new tool to improve our understanding of biomass conversion processes.

submitted post-press

New Imaging Tools for Biofuel Research: Correlated Soft X-ray Tomography and Visible Light Cryo-Microscopy

Mark. A. Le Gros,^{1,3*} Gerry McDermott,^{2,3} Maho Uchida,^{2,3} Christian G. Knoechel,^{2,3} Dilworth Y. Parkinson,^{2,3} Seema Sing,^{4,5} Blake Simmons,^{4,5} and **Carolyn A. Larabell**^{1,2,3} (larabell@lbl.gov)

¹Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, Calif. ²Dept. of Anatomy, University of California, San Francisco; ³National Center for X-ray Tomography, Advanced Light Source, Berkeley,

Calif.; ⁴Joint Bioenergy Institute, Lawrence Berkeley National Laboratory, Berkeley, Calif.; and ⁵Sandia National Laboratory, Livermore, Calif.

Project Goal: Develop correlated imaging technologies required to advanced bioenergy research.

In this poster we will present the recent results from our work developing instruments and methods for carrying out correlated soft x-ray tomography and high numerical aperture immersion light microscopy on cryogenic specimens^{1,2}. These new imaging modalities have enormous potential as precision structural phenotyping tools for bioenergy research. The novel use of a cryogenic immersion fluid in the cryolight microscope minimizes the refractive index mismatch between the specimen and lens, leading to a more efficient coupling of the light from the sample to the image forming system³. The instrument can be used for correlating detailed spectral imaging with a high fidelity x-ray tomographic map of any microorganism. We will show results of correlated imaging on yeast, and also show results of using soft x-ray tomography to phenotype algae for biofuel production.

For more information on the National Center for X-ray Tomography: <http://ncxt.lbl.gov>

References:

1. McDermott G, Le Gros MA, Knoechel CG, Uchida M, & Larabell CA (2009) Soft X-ray tomography and cryogenic light microscopy: the cool combination in cellular imaging. *Trends Cell Biol* 19(11):587-595.
2. Uchida M, *et al.* (2009) Soft X-ray tomography of phenotypic switching and the cellular response to antifungal peptoids in *Candida albicans*. *Proc Natl Acad Sci U S A* 106(46):19375-19380.
3. Le Gros MA, McDermott G, Uchida M, Knoechel CG, & Larabell CA (2009) High-aperture cryogenic light microscopy. *J Microsc-Oxford* 235(1):1-8.

Systems Biology and Metabolic Engineering Approaches for Biological Hydrogen Production

97

Metabolomics and Fluxomics of *Clostridium acetobutylicum* Part 1: Systems-Level Kinetic Flux Profiling Elucidates a Complete TCA Cycle

Daniel Amador-Noguez* (damador@princeton.edu), Xiao-Jiang Feng (xfeng@Princeton.EDU), Herschel Rabitz, and **Joshua D. Rabinowitz** (josh@princeton.edu)

Carl Icahn Laboratory, Dept. of Chemistry and Lewis-Sigler Institute for Integrative Genomics, Princeton University, N.J.

Project Goals (Abstracts 97-99): Microbial biofuel (i.e. hydrogen and butanol) production holds great promise as a source of renewable clean energy. A critical step towards more efficient biofuel production is improved understanding of the regulation of biofuel-related metabolism and the development of models that are sufficiently accurate to enable rational control of the network behavior. With the long term aim of enabling such control, we propose to develop integrated experimental-computational technologies for quantitative dissection of microbial biofuel-producing metabolism. These tools will be broadly applicable to many microbial biofuel producers. We plan to illustrate them with the organism *Clostridium acetobutylicum*.

Clostridium acetobutylicum is an organism with great potential for the commercial production of butanol and hydrogen. Much work has been done to elucidate the metabolic pathways by which solvents are produced in this anaerobic bacterium. However, there are still key pathways of primary metabolism that remain unresolved, including the TCA cycle and amino acid biosynthesis.

As is common for obligatory anaerobic organisms, *C. acetobutylicum* does not contain in its genome obvious homologues of many of the enzymes of the TCA cycle, including citrate synthase, fumarate reductase/succinate dehydrogenase, succinyl-CoA synthetase and α -ketoglutarate dehydrogenase. The apparent lack of these genes is inconsistent with the ability of *C. acetobutylicum* to synthesize α -ketoglutarate and the glutamate family of amino acids and to grow on minimal media. To address the inconsistency, prior metabolic modeling efforts proposed that an incomplete TCA cycle might function in the reductive (counterclockwise) direction to produce α -ketoglutarate. Alternatively, it was suggested that glutamate might be synthesized from ornithine by the arginine biosynthesis pathway running in reverse.

To elucidate the actual pathway that leads to α -ketoglutarate and glutamate production, and to investigate how the TCA cycle of *C. acetobutylicum* operates *in vivo*, we studied the dynamic incorporation of various isotope-labeled nutrients into metabolites in glycolysis, the TCA cycle, the pentose phosphate pathway and amino acid biosynthetic pathways using liquid chromatography-tandem mass spectrometry (LC-MS/MS). In contrast to the previously proposed hypotheses, our results demonstrate that this organism has a complete, albeit bifurcated, TCA cycle. Ketoglutarate is produced exclusively in the oxidative direction from oxaloacetate and acetyl-CoA via citrate. Succinate acts as a dead-end metabolite that can be produced in both the reductive direction from oxaloacetate via malate and fumarate and the oxidative direction via α -ketoglutarate. Our results therefore demonstrate the presence of the biochemical activity of all currently non-annotated enzymes of the TCA cycle including fumarate reductase, citrate synthase, α -ketoglutarate dehydrogenase and succinyl-CoA synthetase. The way in which the TCA cycle bifurcates in *C. acetobutylicum*, with its capacity to synthesize succinate both oxidatively and reductively, suggests that, in addition to its biosynthetic function, it may also play an important role in redox balance. This idea is supported by our observation that most of the succinate produced is excreted.

Our investigations also yielded important information about other unresolved primary metabolic pathways in *C. acetobutylicum*. We found that the Entner-Doudoroff pathway, an alternative pathway for glycolysis, is inactive. The oxidative pentose phosphate pathway is also inactive and this organism relies exclusively on the non-oxidative pentose pathway for the production of ribose-phosphate. Our investigation of the amino acid biosynthesis pathways revealed them to be complete and canonical with the exception of glycine. Glycine was formed from threonine instead of being synthesized by the canonical pathway via serine. Additionally, the one-carbon units required for the methionine, purine, and pyrimidine biosynthesis are not produced via the usual route from serine or glycine but are instead derived from the carboxyl group of pyruvate.

The observations obtained in this study are essential for the construction of an accurate genome-scale model of *C. acetobutylicum* metabolism and lay the groundwork for better understanding of integration of biosynthetic metabolism with solvent and hydrogen production.

98

Metabolomics and Fluxomics of *Clostridium acetobutylicum*, Part 2: Quantitative Flux Model Construction and Analysis

Xiao-Jiang Feng* (xfeng@Princeton.EDU), Daniel Amador-Noguez (damador@princeton.edu), Herschel Rabitz, and Joshua D. Rabinowitz (josh@princeton.edu)

Carl Icahn Laboratory, Dept. of Chemistry and Lewis-Sigler Institute for Integrative Genomics, Princeton University, N.J.

Project Goals: See goals for abstract 97.

In the first part of the research, we employed liquid chromatography-tandem mass spectrometry (LC-MS/MS) to probe the dynamic incorporation of various isotope-labeled nutrients into metabolites of *Clostridium acetobutylicum* in glycolysis, the TCA cycle, the pentose phosphate pathway and amino acid biosynthetic pathways. The labeling patterns of the metabolites identified the metabolic network structure, including a complete and bifurcated TCA cycle, which was unavailable from genome sequence analysis.

To obtain a quantitative understanding of the metabolic fluxes, we formulated an ordinary differential equation (ODE) model of the metabolic network. The model equations represent the quantitative dynamics of the labeled and unlabeled metabolites during exponential growth phase following introduction of isotope-labeled glucose. A nonlinear global inversion algorithm was employed to identify the unknown model parameters, including metabolic fluxes and some metabolite concentrations, that quantitatively reproduced the dynamic labeling data and several experimentally measured steady state constraints. Analysis of the identified model parameters indicates that the main proportion of the glycolytic flux is directed towards production of acids (butyric and acetic acid) through acetyl-CoA and amino acid biosynthesis through aspartate, while the fluxes through the two branches of the TCA cycle are relatively low.

Additionally, we performed model discrimination studies to distinguish multiple network models that can result in the same qualitative isotope labeling patterns. Traditional flux balance analysis suggests that malate and oxaloacetate are produced from fumarate in the TCA cycle. However, model identification results indicate that this structure will not be able to reproduce the observed quantitative data, and malate should be upstream of fumarate. Moreover, the model identification results also show that production of succinate from α -ketoglutarate cannot be achieved via coupling with methionine and lysine biosynthesis alone. The canonical TCA reaction of succinyl-CoA to succinate is required to describe the quantitative dynamics of the relevant metabolites.

In summary, the integrated laboratory and computational investigation generated a genome-scale quantitative flux model of *Clostridium acetobutylicum* metabolism. Model-

based analyses also provided a valuable means for unraveling certain ambiguities in the network structure. The flux model and the advanced techniques developed in the studies will serve as the basis for metabolic engineering of *Clostridium acetobutylicum* in order to achieve optimal biohydrogen production.

99

Student Presentation

Metabolomics and Fluxomics of *Clostridium acetobutylicum*, Part 3: Analysis of the Acidogenic–Solventogenic Transition

Daniel Amador-Noguez, Ian Brasg,* Xiao-Jiang Feng, Herschel Rabitz, and **Joshua D. Rabinowitz** (josh@princeton.edu)

Carl Icahn Laboratory, Dept. of Chemistry and Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, N.J.

Project Goals: See goals for abstract 97.

The solvent and hydrogen-producing bacterium *Clostridium acetobutylicum* has two major metabolic modes. During exponential phase growth it produces acids (butyric and acetic acid), and during stationary phase it takes up the acids previously produced and convert them into solvents (butanol, acetone and ethanol). Controlling this transition and stabilizing the solventogenic state are critical aspects for the commercial production of solvents using this anaerobic bacterium. To this end, it would be useful to have a comprehensive understanding of the intracellular metabolic changes that are associated with the transition between acidogenesis and solventogenesis states.

A previous attempt to tackle this question used microarrays to identify the global gene expression patterns associated with the solventogenic transition. In addition to gene expression changes in solvent producing genes, significant changes were found in a large number of primary metabolic genes in glycolysis and amino acid biosynthesis pathways. Changes in gene expression, however, do not necessarily reflect changes in enzyme activity. Moreover, since complex transcriptional alterations occurred even among genes within pathways (e.g., some increased and some decreased), the transcriptional data alone were insufficient to determine overall metabolic changes.

Kinetic flux profiling is a method for probing cellular metabolic fluxes that is based on the dynamics of cellular incorporation of isotope-labeled nutrient into downstream metabolites. We have previously used this approach to elucidate the metabolic network structure of various unresolved pathways in *C. acetobutylicum* during exponential growth phase. In this ongoing study, we are now applying this approach to investigate the metabolic differences (pathway flux changes and intracellular metabolite concentrations) between the acidogenic and solventogenic states of this organism.

We found that the flux through glycolysis does not change markedly during solventogenesis. Also, flux into the non-oxidative pentose phosphate pathway remains relatively unaffected. There was, however, a large decrease in the synthesis of most glycolysis-derived amino acids, with the notable exception of increased serine biosynthesis. Most of the fluxes coming out of pyruvate (the last metabolite in glycolysis), including into alanine, valine and oxaloacetate production were greatly decreased. This caused an increased flux into Acetyl-CoA, which cascades into increased flux through the acidogenic/solventogenic pathways.

In a related poster, we show that *C. acetobutylicum* has a complete TCA cycle in which succinate can be synthesized in either the oxidative or reductive direction. During solventogenesis, the reductive TCA cycle is completely shutdown. Interestingly, however, the oxidative TCA cycle remains active, producing succinate that is mostly excreted. This observation suggests that the right part of the TCA cycle may play a key role in solventogenesis by producing additional reducing power for solvent production.

Our results highlight the complex metabolic reorganization that takes place in solventogenic *C. acetobutylicum* and provide insight into some possible metabolic regulation points that could be exploited to enhance solvent production. These observations also lay the groundwork for the construction of a genome-scale dynamic quantitative model of the transition from acidogenic to solventogenic metabolism.

100

Photobiological H₂ Production in *Cyanobacterium* ATCC 51142

Matthew R. Melnicki^{1*} (matthew.melnicki@pnl.gov), Grigoriy E. Pinchuk,¹ Jennifer L. Reed,² Oleg V. Geydebrekht,¹ Eric A. Hill,¹ Jim K. Fredrickson,¹ Allan E. Konopka,¹ and **Alex Beliaev**¹

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

Project Goals (Abstracts 100-103): The PNNL Biofuels Scientific Focus Area (BSFA) will carry out fundamental research of microbial photoautotrophs with specific emphasis on photosynthetic energy conversion, reductant partitioning, and carbon metabolism in cyanobacteria focusing on: (i) functions of genes and proteins involved in photosynthetic metabolism; (ii) regulatory factors and networks governing the expression of photosynthetic machinery and the partitioning of reductant through central metabolic pathways; (iii) pathways related to photosynthetic growth and metabolism of cyanobacteria and subsystems (e.g. light-driven electron transfer, respiration, autotrophic carbon assimilation, macromolecule synthesis, nitrogen fixation) interactions; (iv) approaches to manipulate the metabolism of cyanobacteria to channel the reducing equivalents or photosynthetic intermediates to biofuels or biofuel precursors. Consistent with the goals

of the DOE BER Biological Systems Science Program, our long-term goal is to develop predictive systems-level understanding of photosynthetic metabolism through which one can identify and address key science issues that must be resolved to advance biofuel applications.

Biological H₂ production by bacteria and microalgae has been known for more than a century, and research directed at practical application of such microbial processes has been carried out for more than three decades. Although many biohydrogen production concepts have been described, fundamental technological challenges remain in making any such process a practical reality. Advances in microbial genome sequencing and functional genomics are greatly improving the ability to conduct system-level studies of microbial metabolism and to use the obtained knowledge to identify fundamental questions that must be resolved to advance biofuel applications. Genomics and metabolic engineering hold great promise for the rational design and manipulation of biological systems to make such systems efficient and economically attractive.

The research conducted as part of the PNNL Biofuels Scientific Focus Area (BSFA) focuses on elucidating the mechanisms of light-driven metabolism in a unicellular diazotrophic cyanobacterium *Cyanotheca* sp. strain ATCC51142. Conditions promoting H₂ production by *Cyanotheca* 51142 are being studied in order to develop a strategy for maximizing the output of H₂ using metabolic modeling approach. Initially, two-phase experiments have been employed to promote photosynthetically driven accumulation of glycogen that is subsequently converted to H₂. Specifically, during the first phase, strain 51142 was grown in continuous cultures under N-limitation in a photobioreactor sparged with CO₂-enriched Ar (0.3% v/v) and continuously illuminated at 150 μmol/m²·s. Upon reaching steady-state (biomass concentration 80 mg/l of ash-free dry weight), the cultures were incubated in N-free medium in the absence of CO₂ using 100% Ar as sparging gas while measuring the off-gas composition by in-line mass-spectrometry. Phase two was initiated by placing cultures either under dark or the light (200 μmol/m²·s) conditions. Appropriate controls consisting of light and dark cultures amended with ammonium or 3-(3,4-dichlorophenyl)-1,1-dimethylurea (DCMU), an inhibitor of photosystem II, were also included in the experiment. The data obtained revealed that, unlike non-nitrogen fixing cyanobacteria such as *Synechocystis* and *Synechococcus* spp., *Cyanotheca* 51142, is not capable of generating significant amounts of H₂ from stored glycogen under dark fermentative conditions, nor does it produce H₂ under light conditions in the presence of ammonia or N₂. However, illuminated cultures exposed to an Ar atmosphere and deprived of N₂ and CO₂ produced significant amounts of H₂. O₂ was also produced along with H₂ at 1:2 ratio, whereas DCMU significantly (4.5-fold) decreased H₂ generation. It should be noted that analysis of cell-free culture supernatants did not reveal any accumulation of organic acids. These results suggest that PSII and therefore water photolysis played a significant role in H₂ evolution by strain 51142. The inhibition of H₂ production by ammonia or N₂ strongly suggest that nitrogenase was the enzyme primarily

responsible for light-driven H₂ production, and whole cell assays revealed high nitrogenase activity in H₂-producing cells. Preliminary analysis of the material balance suggests the nitrogenase activity was supported by light-driven electron transfer. Within the scope of the proposed BSFA research, we will further elucidate and validate the pathways of light-driven two-step H₂ production by cyanobacteria and incorporate the experimental data into the metabolic model of *Cyanotheca* 51142 to identify the means for maximization of H₂ production by this organism.

101

Constraint-Based Modeling for Maximizing the Metabolic Potential of Photoautotrophic Microorganisms

Jennifer L. Reed^{1*} (reed@cae.wisc.edu), Trang Vu,¹ Jim K. Fredrickson,² Oleg V. Geydebrekht,² Eric A. Hill,² Allan E. Konopka,² Grigoriy E. Pinchuk,² Sergey M. Stolyar,² Jessica De Ingeniis,³ Andrei L. Osterman,³ Donald A. Bryant,⁴ and Alex S. Beliaev^{2*}

¹University of Wisconsin, Madison; ²Pacific Northwest National Laboratory, Richland, Washington; ³Burnham Institute for Medical Research, La Jolla, California; and ⁴Pennsylvania State University, University Park

Project Goals: See goals for abstract 100.

Photosynthetic microorganisms possess the unique ability to convert sunlight into chemical energy using water as the electron donor. Despite the wealth of information on the mechanistic aspects of bacterial photosynthesis and supramolecular complexes catalyzing the process of light conversion and CO₂ fixation, a system-level understanding of photosynthetic metabolism is yet to be achieved. Different phototrophic microorganisms display varying levels of light conversion efficiencies, which ultimately translate in different rates of electron transfer, ATP/NAD (P)H production, and growth. Understanding the origin of these properties will provide fundamental new insights that could be widely applied to the development of photosynthetic systems for biofuels development. Integral to that is the question, is the process of reductant partitioning in photoautotrophs which links energy-generating reactions with biosynthesis of biomass precursors and storage compounds. Within the scope of the PNNL Biofuels Scientific Focus Area (BSFA), we are exploring the mechanisms of energy conservation and carbon partitioning in cyanobacteria. One important outcome of the project will be development of a predictive tool, *i.e.* a genome-scale model, which provides a platform for integrating all knowledge and experimental data generated within the project. It will also have the ability to serve as an *in silico* tool for manipulating photosynthetic microorganisms to act as catalysts for solar energy conversion and will potentially allow development of a highly efficient biofuel production process.

As part of previous Genomics:GTL funding, we have built a genome-scale metabolic network for *Cyanotheca* sp. ATCC 51142, a unicellular diazotrophic cyanobacterium that can temporally separate the process of light-dependent autotrophic growth and glycogen accumulation from N_2 fixation. The resulting model currently includes 798 genes, 682 proteins, 630 metabolites, and 656 reactions accounting for common pathways such as central metabolism, nucleotide and amino acid biosynthesis, and those that are more unique to cyanobacteria such as photosynthesis, carbon fixation, and cyanophycin production. Photosynthesis was modeled as sequential reactions that occur in each photosystems, in order to study the effect of different light wavelengths, and separate photosystem activities on cellular growth and hydrogen production rate. Predicted results from the metabolic model, based on growth simulations of the constraint-based model under different carbon and nitrogen sources for photoautotrophic, heterotrophic and mixotrophic conditions qualitatively agree with experimental data. Using a custom-built photobioreactor, which allows for the control and monitoring of incident and transmitted light, we have also studied the physiological response of *Cyanotheca* sp. ATCC 51142 to nitrogen and light limitations imposed on photosystems I and II. Biomass composition and metabolite analyses were carried out to provide experimental validation for the model.

In addition, we have developed a draft metabolic network for *Synechococcus* sp. PCC 7002, a fast growing non-nitrogen-fixing cyanobacterium which exhibits the fastest growth rate of known cyanobacteria and is also remarkably tolerant to high light intensities. Understanding the origin of these properties could provide fundamental new insights that could be widely applied to the development of other biological systems for biofuels development. Initial comparisons between the reconstructed metabolic networks of *Cyanotheca* 51142 and *Synechococcus* 7002 suggested that both networks share a significant number of pathways. However *Synechococcus* 7002 also displays notable differences, specifically in pathways involved in amino acid and folate metabolism. Once the reconstruction of *Synechococcus* 7002 network is complete, we will apply metabolic engineering algorithms to identify strategies for modulating the efficiencies of light conversion, carbon fixation, and photosynthate production.

102

Genome-Enabled Studies of Photosynthetic Microorganisms for Bioenergy Applications

Alex S. Beliaev^{1*} (alex.beliaev@pnl.gov), **Allan E. Konopka**,¹ Harvey Bolton Jr.,¹ Mary S. Lipton,¹ Thomas O. Metz,¹ Grigoriy E. Pinchuk,¹ Aaron T. Wright,¹ Thomas C. Squier,¹ Sergey M. Stolyar,¹ Donald A. Bryant,² Matthew C. Posewitz,³ and Jennifer L. Reed⁴

¹Pacific Northwest National Laboratory, Richland, Wash.; ²Pennsylvania State University, University Park; ³Colorado School of Mines, Golden; ⁴University of Wisconsin, Madison

Project Goals: See goals for abstract 100.

With the increasing concerns over the sustainability of a crop-based biofuel economy, there is a renewed interest in photosynthetic microorganisms, which use solar energy, H_2O , and CO_2 , as effective alternatives for the production of biofuels and primary biomass. Cyanobacteria and microalgae have the potential to produce biofuels at a much higher productivity than vascular plants and they can be cultivated in freshwater and marine aquatic environments that do not compete for land resources with conventional agriculture. While structural and functional properties of protein complexes catalyzing the first steps of photosynthetic energy conversion reactions have been extensively explored, harnessing photosynthetic metabolism for biofuels production requires detailed knowledge of cellular subsystems and networks involved in electron transport, reductant partitioning, and energy storage pathways. The advances in microbial genome sequencing and functional genomics have greatly improved the ability to construct accurate systems-level models of microbial metabolism and to query the models for gene targets that enhance productivity by metabolic engineering.

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 focusing on: (i) functions of genes and proteins involved in photosynthetic metabolism; (ii) regulatory factors and networks governing the expression of photosynthetic machinery and the partitioning of reductant through central metabolic pathways; (iii) pathways related to photosynthetic growth and metabolism of cyanobacteria and subsystems (e.g., light-driven electron transfer, respiration, autotrophic carbon assimilation, macromolecule synthesis, nitrogen fixation) interactions; (iv) approaches to manipulate the metabolism of cyanobacteria to channel the reducing equivalents or photosynthetic intermediates to biofuels or biofuel precursors. Leveraging from the laboratory's cutting-edge technical capabilities, the research conducted under the PNNL BSFA will embody both scientific and technical tasks including development of continuous cultivation, biochemical, and genetic methods in conjunction with genomic, proteomic, metabolomic and modeling approaches for studying funda-

mental aspects of the phototrophic metabolism. Consistent with the goals of DOE BER Genomic Science Program, our long-term goal is to develop predictive systems-level understanding of photosynthetic metabolism through which one can identify and address key science issues that must be resolved to advance biofuel applications.

103

Phototroph-Heterotroph Co-Cultures for Studying Organism Interactions and Pathways of Solar Energy Conversion

Grigoriy E. Pinchuk^{1*} (grigoriy.pinchuk@pnl.gov), Oleg V. Geydebrekht,² Jennifer L. Reed,² Donald A. Bryant,³ Allan E. Konopka,¹ **Jim K. Fredrickson,^{1*}** and **Alex S. Beliaev^{1*}**

¹Pacific Northwest National Laboratory, Richland, Wash.;

²University of Wisconsin, Madison; and ³Pennsylvania

State University, University Park

Project Goals: See goals for abstract 100.

Much of terrestrial and aquatic photosynthesis occurs in complex microbial consortia, but little is presently known about interactions among microorganisms that contribute to their efficient solar energy capture and conversion. Under the PNNL Foundational Scientific Focus Area (FSFA), we are utilizing a combination of complementary laboratory-based research and exploration of naturally-occurring associations with combinations of defined photoautotroph-heterotroph cultures to understand the collective energy, carbon, and nutrient processing in microbial systems. The complex natural systems we have selected for study including the phototrophic mats of Yellowstone National Park (YNP) and surface waters of central Washington as of yet have not yielded cultivated members that can be used as model systems for detailed laboratory-based research. Due to the paucity of information on the heterotrophic population (s) in these natural systems, initial research and development of methodology for studying microorganism interactions is utilizing co-cultures of representative cyanobacteria and well-studied *Shewanella* species. While recognizing the opportunistic nature of this system, there is ample evidence that certain species of *Shewanella* live in association with autotrophic prokaryotes and examples of these associations are well documented. More importantly, we believe that a *Synechococcus-Shewanella* co-culture can be instrumental in gaining basic understanding of opportunistic interactions between photoautotrophic and heterotrophic bacteria. Our preliminary results using cyanobacteria-*Shewanella* co-cultures demonstrated that metabolic coupling and interactions between photoautotrophic and heterotrophic microorganisms may serve as a mechanism for controlling dissolved O₂ concentration, increasing Fe and Mn availability, and recycling nutrients in natural communities.

Complementary to the FSFA work, we are also exploring the potential of photoautotroph-heterotroph associations

for bioenergy applications as part of the PNNL Biofuels Scientific Focus Area. Although phototroph-heterotroph associations are abundant in nature, the co-culture approach has been seriously under-appreciated. To date, engineering of microbes for biofuel production is being carried out using single strains by enhancing or deleting specific steps of a pathway or modulating activities of specific enzymes. However, synthesis of biofuel precursor molecules requires precise coordination and interactions of many proteins within various pathways, where any adjustments or increases in expression and/or activity levels can lead to substantial metabolic burden and suboptimal yields. In that regard, engineering of photosynthetic organisms, which carry out simultaneous light- and dark-phase reactions, is inherently challenging. Photosynthetic production of biofuels often requires optimization of two or more metabolic functions which can be mutually exclusive in a single microbial cell and therefore require either spatial and/or temporal separation (e.g. O₂ evolution and H₂ production; O₂ evolution and N₂ fixation, sensitivity of RuBisCo to O₂). To that end, co-culturing of photosynthetic and heterotrophic microorganisms offers efficient ways to optimally engineer the photosynthetic production of biofuels. By engineering photosynthetic strains which excrete organic carbon compounds (organic acids or sugars) and co-culturing them with a heterotrophic organism capable of utilizing the excreted compounds, one can physically separate the processes of photosynthesis and photosynthate conversion while allowing for net CO₂ consumption. The co-cultivation of phototrophs and aerobic heterotrophs also eliminates technical problems associated with oxygen-sensitivity and substrate delivery by creating favorable microaerobic CO₂-enriched environments for the phototrophic microorganisms. Moreover, utilization of exogenously-added organic carbon by the heterotroph, decreases dissolved O₂ concentrations and induces the expression of O₂-sensitive enzymes (e.g. hydrogenase and nitrogenase) in the phototroph which in turn will generate reducing equivalents by light-driven water photolysis. Overall, we believe that the implementation of the co-culture approach will open new perspectives for designing efficient and cost-effective processes and will provide a novel platform for the development of consolidated bioprocessing methods leading to production of carbon-neutral energy at reduced economic and energetic costs.

104

Probing Metalloenzymes with Synchrotron Radiation – from Gamma Rays to Soft X-Rays

Stephen P. Cramer^{1,2*} (spjcramer@ucdavis.edu)¹Dept. of Applied Science, University of California, Davis and ²Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, Calif.**Project Goals: We are trying to develop a resource for biological and environmental x-ray spectroscopy at ALS beamline 4.0.2.**

Nitrogenase is the enzyme responsible for the ‘fixation’ of nearly inert atmospheric dinitrogen to ammonia. It is ultimately responsible for half of the world’s protein, while the other half depends on industrial fertilizer produced with hydrogen derived from fossil fuels. Nitrogenase uses a complex MoFe₇S₉X-homocitrate ‘FeMo-cofactor’ to convert N₂ to NH₃, but the detailed mechanism remains poorly understood.¹⁻³ Another type of enzyme, hydrogenase, catalyzes the interconversion of dihydrogen with protons and electrons. These enzymes use unusual forms of Fe-S clusters or Fe carbonyls, and their catalytic mechanisms are not understood.

One way to study Fe in biological systems is Nuclear Resonance Vibrational Spectroscopy (NRVS). In this synchrotron radiation technique, a sample is excited with a ~1 meV bandwidth beam near a Mössbauer resonance, and the delayed fluorescence is recorded as a function of excitation energy. When applied to Fe samples, NRVS is only sensitive to vibrations involving motion of ⁵⁷Fe. We will present results on model compounds, small Fe-S proteins, nitrogenase, and hydrogenase, and the needs and prospects for future improvements will be discussed.⁴

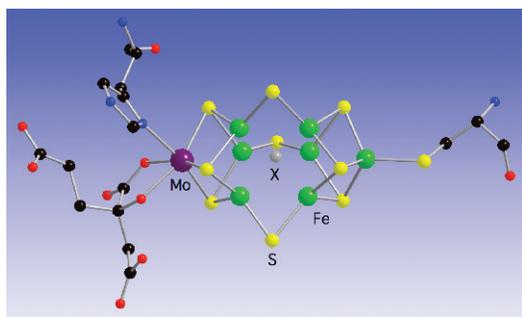


Figure. Structure of the nitrogenase FeMo-cofactor.

Although NRVS is a powerful tool for studying Fe, it is not applicable to the other metal centers of interest: Mo in nitrogenase or Ni in hydrogenase. We have thus resorted to soft x-ray spectroscopy as a probe for these sites. Some Ni L-edge and Mo M-edge spectra will be presented, and the issue of radiation damage with soft x-rays will be discussed.⁵⁻⁶

Finally, our efforts to develop a soft x-ray resource, ABEX, for biological and environmental science at the ALS will be summarized.

References

1. Peters, J.W.; Szilagyi, R.K. *Curr. Opin. Chem. Biol.*, 2006, 10, 101-108.
2. Barney, B.M.; Lee, H.-I.; Santos, P.C.D.; Hoffman, B.M.; Dean, D. R.; Seefeldt, L.C. *Dalton Trans.*, 2006, 2277-2284.
3. Dance, I. *Chem. Asian J.*, 2007, 2, 936-946.
4. Xiao, Y.; Fischer, K.; Smith, M.C.; Newton, W.; Case, D.A.; George, S.J.; Wang, H.; Sturhahn, W.; Alp, E.E.; Zhao, J.; Yoda, Y.; Cramer, S.P. *J. Am. Chem. Soc.*, 2006, 128, 7608-7612.
5. George, S.J.; Fu, J.; Guo, Y.; Drury, O.; Friedrich, S.; Rauchfuss, T.; Volkers, P.I.; Peters, J.C.; Scott, V.; Brown, S.D.; Thomas, C.M.; Cramer, S.P. *Inorg. Chim. Acta*, 2008, 361, 1157-1165.
6. George, S.J.; Drury, O.B.; Fu, J.; Friedrich, S.; Doonan, C.J.; George, G.N.; White, J.M.; Young, C.G.; Cramer, S.P. *J. Inorg. Biochem.*, 2009, 103, 157-167.

105

Student Presentation

Stopped-Flow IR Spectroscopy of Hydrogenases

Celestine Grady,^{1,2*} Simon George,^{1,2} Stephen P. Cramer^{1,2} (spjcramer@ucdavis.edu), Eric Shepard,³ Shawn McGlynn,³ Joan Broderick,³ John Peters,³ Jon M. Kuchenreuther,⁴ James A. Stapleton,⁴ and James R. Swartz⁴

¹Dept. of Applied Science, University of California, Davis; ²Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, Calif.; ³Dept. of Chemistry and Biochemistry, Montana State University, Bozeman; and ⁴Dept. of Chemical Engineering, Stanford University, Stanford, Calif.

Project Goals: Studying hydrogenases with IR spectroscopy can provide information on the creation of hydrogenase, the molecular and atomic makeup of the enzyme active site, the surrounding amino acid environment, and the mechanisms by which the enzyme reacts with hydrogen or inhibitors like oxygen. The understanding of these properties could lead to the development of improved mutant enzymes that can provide a substantial source of hydrogen.

As we move toward a future that depends less on hydrocarbons and more on a variety of sustainable energy sources, we recognize that hydrogen could be an important part of a new clean energy infrastructure. A possible effective and feasible solution for the mass production of hydrogen is through the manipulation of hydrogenases, enzymes that reversibly catalyze the evolution of molecular hydrogen from protons and electrons. Hydrogenases are found in organisms

such as algae that could potentially be harvested for hydrogen production.

Studying hydrogenase with IR spectroscopy can provide information on the creation of the enzyme, the molecular makeup of the enzyme active site, the surrounding amino acid environment, and the mechanisms by which the enzyme reacts with hydrogen or inhibitors such as CO and oxygen. An IR spectrum is obtained by shining IR light through a sample and measuring the amount of light that is transmitted. There will be visible absorption lines for the frequencies of light that have excited some vibration within the molecule. Each vibration is unique to a functional group, and the IR spectrum can be interpreted to determine the composition of the sample. Here we present IR spectra of as-isolated hydrogenase samples, and hydrogenases combined with sodium dithionite, CO and oxygen. In addition, we have used IR techniques such as Stopped Flow-FTIR and photolysis to obtain time-dependent observations of these hydrogenase reactions.

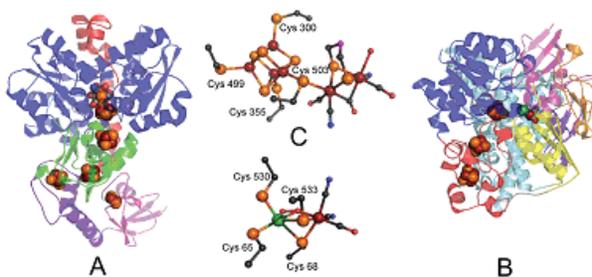


Figure. Structures of (A) Cpl H₂ase, (B) [NiFe] H₂ase, (C) H-cluster, and (D) NiFe active site.

106 Novel Hydrogen Production Systems Operative at Thermodynamic Extremes

Michael J. McInerney,¹ Ralph S. Tanner,¹ Gregory Hurst,² Lars Rohlin,³ and **Robert P. Gunsalus**^{3*} (rob@microbio.ucla.edu)

¹University of Oklahoma, Norman; ²Oak Ridge National Laboratory, Oak Ridge, Tenn.; ³University of California, Los Angeles

<http://www.mimg.ucla.edu/faculty/gunsalus/>

Project Goals: The goals of this collaborative project are to develop new research strategies to address the Genomics:GTL program needs in the area of bio-hydrogen production. This includes the delineation of the molecular machinery involved in hydrogen production from thermodynamically difficult substrates, as well as the characterization of new microbial model systems that generate high H₂ concentrations approaching 17% of the gas phase. We are performing systems-based studies of bio-hydrogen production in model anaerobic consortia as well

as with pure culture model strains to identify key regulated steps. The results of these studies will greatly expand our ability to predict and model systems for H₂ production in novel anaerobes that are currently very poorly understood.

The goals of this collaborative project are to develop new research strategies to address the Genomics:GTL program needs in the area of bio-hydrogen production. This includes the delineation of the molecular machinery involved in hydrogen production from thermodynamically difficult substrates, as well as the characterization of new microbial model systems that generate high H₂ concentrations approaching 17% of the gas phase.

To identify the strategies used for hydrogen production in one model butyrate-degrading organism, we are performing genomic, proteomic, and transcript analysis on *Syntrophomonas wolfei*. This microbe is representative of an important but poorly understood class of hydrogen-producing organisms that are capable of syntrophic fatty and aromatic acid metabolism when co-cultured with suitable microbial partner(s). Their ability to produce H₂ requires reverse electron transport with energy input. Analysis of the *S. wolfei* genome reveals many genes with potential to accomplish this task. It possesses genes for three cytoplasmic and two externally located formate dehydrogenases plus two cytoplasmic and one externally located hydrogenase. By implication, either hydrogen or formate could be produced by *S. wolfei* during syntrophic interactions. Interestingly, the three cytoplasmic formate dehydrogenases plus one of the soluble-type hydrogenases appear to be NADH-linked since the respective gene clusters contain genes for NADH:quinone oxidoreductases chains E and F. This suggests that *S. wolfei*, like several other sequenced syntrophic metabolizers and anaerobes known to produce high molar ratios of hydrogen from glucose, may produce H₂ and/or formate from NADH by an electron bifurcation mechanism. To determine which of the above enzymes are utilized in pure culture, proteomic studies were performed initially using crotonate-grown cells. This was accomplished by analyzing whole cell-derived peptide mixtures with two-dimensional liquid chromatography/tandem mass spectrometry (2D LC-MS-MS) via the MudPIT approach. Two highly expressed hydrogenase enzymes were detected plus a novel, electron transfer flavoprotein-linked FeS-type reductase complex that is probably used to process electrons generated by the oxidation acyl-CoA intermediates. *S. wolfei* metabolizes fatty acids by the β -oxidation pathway and surprisingly, the genome reveals multiple homologues for many of the enzymatic steps even though it has an extremely restricted range of fatty acid substrates. Provisional protein assignments were also made by LC-MS-MS for each of the eight reactions leading to acetate formation. Energy is harvested by substrate level phosphorylation via acetate kinase to yield one ATP. Additional proteomic and transcript studies are in progress to further characterize the expression of genes/proteins involved in the reversed electron transfer process for hydrogen production. These studies will yield improved understanding of how *S. wolfei* and other syntrophic metabolizers thrive at low thermodynamic driving forces not possible for many other anaerobes.

In a companion project we are characterizing the genetic, biochemical, and physiological properties of a newly isolated anaerobic bacterium called *Anaerobaculum hydrogeniformans* strain OS1 that can generate H₂ at concentrations up to 17%. In one approach, genomic sequencing is being performed on this representative member of the Synergistetes group. The genome of approximately 2.4 MB in size has a GC content of 46.5%. The current assembly consists of 403 contigs with about 1.9 MB contained in the top six contigs. Machine annotation and manual curation is currently in progress to support a metabolic reconstruction of the cellular metabolism leading to hydrogen formation in strain OS1 when grown on hexoses and pentoses. Since strain OS1 can also grow syntrophically in the presence of a H₂-consuming methanogen, it suggests an ability for a more complex alternative lifestyle.

107

Hydrogen Production Comes at the Expense of Calvin Cycle CO₂ Fixation During Photoheterotrophic Growth by *Rhodospseudomonas palustris*

James B. McKinlay* (mckinla1@uw.edu), Yasuhiro Oda, Colin R. Lappala, and **Caroline S. Harwood**

Dept. of Microbiology, University of Washington, Seattle

Project Goals: The goals of this project are (i) to use ¹³C-metabolic flux analysis and other approaches to understand the central metabolic changes involved in H₂ production by *R. palustris* and (ii) to use the resulting information to guide the metabolic engineering for improving H₂ production characteristics.

There is currently a pressing need for renewable fuels to negate the adverse social, economic, and environmental impacts of burning fossil fuels. H₂ is a promising biofuel, having about three-times the energy content of gasoline. Although most manufactured H₂ comes from fossil fuels H₂ can also be produced biologically. *Rhodospseudomonas palustris* uses energy from sunlight and electrons from organic waste to produce H₂ via nitrogenase. In order to understand and improve this process we used ¹³C-acetate to track and compare central metabolic fluxes in non-H₂ producing wild-type *R. palustris* and an H₂-producing mutant. Wild-type cells metabolized 22% of the acetate to CO₂ and then fixed 68% of this CO₂ into cell material using the Calvin cycle. This Calvin cycle flux enabled *R. palustris* to re-oxidize nearly half of the reduced cofactors generated during acetate oxidation. The H₂-producing mutant produced a similar amount of CO₂ but the Calvin cycle flux was much lower, re-assimilating only 12% of the CO₂. In this mutant, H₂ production assumed much of the redox balance burden as about 90% of the electrons for H₂ production were diverted away from the Calvin cycle. Microarray and Q-PCR analyses showed that the shift of electrons from the Calvin cycle to H₂ production involved transcriptional control of Calvin cycle operons. However, this transcriptional control did not require the

redox-sensing two-component regulatory system RegSR. When Calvin cycle flux was disrupted completely by deleting the genes encoding ribulose 1,5-bisphosphate carboxylase, *R. palustris* was forced to use H₂ production alone to maintain redox balance. This mutant exhibited a 1.5-fold increase in H₂ yield but at a cost to the growth rate. These results demonstrate how systems level approaches, such as ¹³C-metabolic flux analysis, can lead to effective strategies to improve H₂ yield. Furthermore, our results underscore that the Calvin cycle and nitrogenase have important electron-accepting roles separate from their better known roles in ammonia production and biomass generation.

108

Optimization of NSR-seq for Transcriptome Analysis in *Rhodospseudomonas palustris*

Christopher D. Armour,^{1,2} Yasuhiro Oda^{3*} (yasuhiro@u.washington.edu), John C. Castle,^{1,4} Somsak Phattarasukol,³ Colin R. Lappala,³ Matthew C. Biery,^{1,2} **Caroline S. Harwood**,³ and Christopher K. Raymond^{1,2}

¹Rosetta Inpharmatics LLC, Merck and Co., Inc., Seattle, Wash.; ²NuGEN Technologies, Inc., San Carlos, Calif.; ⁴Institute for Translational Oncology and Immunology, Mainz, Germany; and ³Dept. of Microbiology, University of Washington, Seattle

Project Goals: The goal of this project is to exploit next generation sequencing technology to characterize the transcriptional networks underlying hydrogen production in the photosynthetic bacterium *Rhodospseudomonas palustris*.

The emergence of next generation sequencing (NGS) technology has opened up new opportunities for the optimization of biological systems for alternative energy production. The open query data format of NGS digital readouts is particularly useful for characterizing the transcriptional networks underlying biofuel production in genetically diverse bacterial strains. The success of sequence-based expression profiling, however, depends on the availability of efficient methods for the construction of high complexity cDNA libraries that are compatible with NGS platforms. Conventional random-priming techniques produce libraries that are largely composed of ribosomal RNA (rRNA) transcripts, so affinity purification schemes are commonly applied to reduce rRNA content prior to reverse transcription. While this approach has been moderately effective at removing rRNA in some systems, it has generally been ineffective for organisms with genomes of high G + C content. Moreover, affinity-based rRNA depletion requires high RNA inputs to obtain ample purified material for cDNA synthesis. To overcome these limitations, we have developed an alternative strategy, called Not-So-Random (NSR) priming, which utilizes computationally designed hexamers to synthesize cDNA selectively from non-rRNA template molecules (Armour et al., 2009). In addition to reducing rRNA load, NSR library construction preserves transcript strand polarity

and requires only 1 µg of total RNA input. We have adapted NSR-seq methodology, which was originally developed in mammalian systems (Armour et al., 2009), to the photosynthetic bacterium *Rhodospseudomonas palustris* to dissect the biochemical pathways involved in hydrogen production using the Illumina Sequencing-By Synthesis platform.

With the ultimate goal of profiling diverse strain backgrounds, we designed NSR hexamers against rRNA transcripts obtained from six *R. palustris* strains for which complete genome sequences were available. Alignment of all possible hexamer sequences (4,096) to the 5S, 16S, and 23S rRNA sequences from each strain resulted in the identification of 1,203 NSR primers that had no perfect match complementarity to any of the rRNA filter transcripts. Oligonucleotides containing each NSR hexamer were synthesized individually with a 10 nt universal tail sequence at the 5' terminus and pooled prior to library construction. An NSR hexamer pool synthesized in the forward strand orientation was used for second strand synthesis (sense), whereas a pool containing the reverse complements was used in the antisense cDNA reaction. Distinct 5' tail sequences were used for the first and second strand synthesis primer pools, so that transcript strand orientation would be maintained through the library construction process.

Sequence analysis of a test library generated with total RNA isolated from *R. palustris* strain CGA009 indicated that NSR-priming did result in mRNA enrichment compared to conventional random-priming, but the effect was modest relative to a random-primed library built with RNA that had been pre-treated with the MICROBExpress™ Bacterial mRNA Enrichment Kit (Ambion, Inc.). Only 3% of the reads mapping to the genome aligned unambiguously in the random primed control library, compared to 7% and 11% for the NSR and MICROBExpress™ treated libraries, respectively. Moreover, we found that NSR-priming had a differential effect on rRNA species; 23S abundance was reduced by 50%, whereas 16S levels increased slightly relative to the control. A closer inspection of the distribution of NSR reads across rRNA loci revealed that only a few template sites accounted for the majority of rRNA priming events. This allowed us to remove problem hexamer sequences from the original NSR primer pool without significantly diminishing sequence complexity. Constructing libraries with a refined set of NSR primers that included 925 of the original 1,203 hexamers, the so-called 'cut300' primer set, increased the number of unambiguous alignments to 22% of all mapped reads. We also observed that specific bases in the universal primer sequence upstream of the NSR hexamer site contributed to rRNA priming, thus offering another opportunity to enhance primer selectivity. Re-engineering the 3' end of the tail sequence (CGA>TTA) further increased mRNA enrichment. With these improvements, 42% of NSR reads that map to the genome align to unique sites. Further testing indicated that the resulting NSR-seq mRNA expression profiles were highly reproducible and strand-specific.

Reference

1. Armour, C. D., J.C. Castle, R. Chen, T. Babak, P. Loerch, S. Jackson, J.K. Shah, J. Dey, C.A. Rohl, J.M. Johnson and

C.K. Raymond (2009) Digital transcriptome profiling using selective hexamer priming for cDNA synthesis. *Nat Methods* 6:647-649.

109

Strand-Specific NSR RNA-seq Analysis of *Rhodospseudomonas palustris* Reveals Additional Features of its Transcriptome that May Influence Hydrogen Production

Yasuhiro Oda^{1*} (yasuhiro@u.washington.edu), Colin R. Lappala,¹ Somsak Phattarasukol,¹ Christopher D. Armour,^{2,3} John C. Castle,^{2,4} Matthew C. Biery,^{2,3} Christopher K. Raymond,^{2,3} and **Caroline S. Harwood¹**

¹Dept. of Microbiology, University of Washington, Seattle; ²Rosetta Inpharmatics LLC, Merck and Co., Inc., Seattle, Wash.; ³NuGEN Technologies, Inc., San Carlos, Calif.; and ⁴Institute for Translational Oncology and Immunology, Mainz, Germany

Project Goals: The overall goal of this project is to use a systems level approach to dissect metabolic and regulatory networks necessary for nitrogenase-catalyzed hydrogen production by a phototrophic bacterium *Rhodospseudomonas palustris*.

Hydrogen gas has good potential for as a transportation fuel because it is clean burning and has a high energy content. Bacteria can produce hydrogen by several different enzymatic routes. The photosynthetic bacterium *Rhodospseudomonas palustris* produces copious amounts of hydrogen via the enzyme nitrogenase, which generates both ammonia and hydrogen as products of dinitrogen gas reduction. When nitrogen gas is not available, nitrogenase uses only protons and electrons as substrates and produces pure hydrogen gas. This reaction requires large amounts of ATP and electrons, which *R. palustris* can obtain from sunlight and biomass, respectively, by fairly complex metabolic routes. Thus hydrogen production involves the appropriate integration of dozens of metabolic reactions. Our long-term goal is to integrate transcriptomic and phenotypic data from up to 100 *R. palustris* strains using Bayesian network analysis to identify all genes involved in hydrogen production. This would include genes; such as central carbon metabolism genes that may change very little in expression and are therefore not easily recognized when just a few strains are analyzed by conventional transcriptomic techniques.

As a start, we have been working to develop improved techniques of transcriptome analysis by deep cDNA sequencing. This is necessary because strain-to-strain variation precludes the use of traditional microarrays. We modified for use in bacteria a strand-specific cDNA sequencing method called Not-So-Random (NSR) RNA-seq (Armour et al., 2009). This method uses a collection of computationally selected oligonucleotides to selectively enrich non-rRNA cDNAs. Also the cDNA libraries are prepared in such a way as to preserve strand specificity and therefore reveal the overarch-

ing themes of sense and antisense strand transcription across the genome (Armour et al., 2009). The cDNA libraries are sequenced by Illumina sequencing technology (25 bp sequence reads).

When we applied this method to *R. palustris* strains CGA009 and TIE-1, on the order of 60% of the total sequencing reads were non-rRNA reads starting from 1 μg of total RNA. Of the remaining reads (on the order of 2 million) approximately 70% mapped to genes with no base pair mismatches. We tested three growth conditions: nitrogen-fixing (hydrogen-producing) – high light, nitrogen-fixing (hydrogen-producing) – low light, and ammonia – high light. In each condition over 90% of the genes in each genome were expressed. The most highly expressed genes in the genomes were light harvesting 2 (LH2) and light harvesting 4 (LH4) genes. Interestingly the LH4 operon was expressed at higher levels under nitrogen-fixing as compared to ammonia-grown conditions at high light, perhaps reflecting the increased need for cellular ATP to supply to the nitrogenase enzyme. The LH4 operon was expressed at its highest levels under low light conditions.

We developed software to visually map cDNA reads onto the chromosomal map. Color-coding allows us to quickly visualize open-reading frame transcripts (the largest class of reads), anti-sense reads within genes, and the 5' untranslated regions of transcripts, allowing estimation of transcription start sites. Transcripts that map to intergenic regions in the opposite orientation from flanking genes are candidates for small trans-acting RNAs. Such a candidate sRNA was found in the nitrogenase gene cluster. Anti-sense reads in the 5' untranslated regions of genes may represent antisense cis-acting sRNAs. A candidate for this type of regulatory RNA is found in the region 5' of the LH4 operon.

Our results show that the NSR approach is an effective way to circumvent the problem of excessive rRNA reads, which – because rRNA is present in such overwhelming amounts in cells relative to other RNAs – will dominate sequence-based transcriptional analysis if permitted to do so. In addition the strand-specific sequencing information obtained using the NSR protocol has allowed us to identify potential cis- and trans-acting sRNAs that could constitute a previously unrecognized layer of regulatory control over hydrogen gas production.

Reference

1. Armour, C. D., J.C. Castle, R. Chen, T. Babak, P. Loerch, S. Jackson, J.K. Shah, J. Dey, C.A. Rohl, J.M. Johnson and C.K. Raymond (2009) Digital transcriptome profiling using selective hexamer priming for cDNA synthesis. *Nat Methods* 6:647-649.

110

Genetic Manipulation of the Hyperthermophilic Hydrogen Producer, *Thermotoga maritima*

D. White,^{1*} R. Singh,¹ K. Noll,² R. Kelly,³ and P. Blum¹

¹Beadle Center for Genetics, University of Nebraska, Lincoln; ²Dept. of Molecular and Cell Biology, University of Connecticut, Storrs; and ³Dept. of Chemistry and Biomolecular Engineering, North Carolina State University, Raleigh

Project Goals: The anaerobic hyperthermophile, *Thermotoga maritima* (Tma), ferments carbohydrates to form molecular hydrogen (H₂) as one of its by-products. A high overall H₂ yield makes *T. maritima* a preferable model to evaluate hydrogen production. Our focus is to develop and use genetic methods to manipulate the *Tma* genome to investigate the metabolic flux of carbon and hydrogen synthesis.

The anaerobic hyperthermophile, *Thermotoga maritima* (*Tma*), ferments carbohydrates to form molecular hydrogen (H₂) as one of its by-products. A high overall H₂ yield makes *T. maritima* a preferable model to evaluate hydrogen production. Our focus is to develop and use genetic methods to manipulate the *Tma* genome to investigate the metabolic flux of carbon and hydrogen synthesis. A new *Tma* genetic marker has been developed by screening for uracil auxotrophs among spontaneous mutants resistant to the pyrimidine analog, 5-fluoroorotic acid (5-FOA). The *pyrE-64* mutant (strain PBL3001) arose by a two nt deletion (-TG) at chromosomal positions 351,539 (-T) and 351,538 (-G), 155 nt from the end of *pyrE*. This mutation results in a premature stop codon (TGA) 64 nt before the natural stop and therefore reduces protein length by 21 AA (from an original 187 AA). Auxotrophy was confirmed by demonstrating growth in a defined medium was dependent upon uracil supplementation. Stability of *pyrE-64* was evaluated by enrichment and characterization of gain-of-function prototrophic suppressors. The *pyrE-100* mutant (strain PBL3021) restored the *pyrE* reading frame by deletion of an additional one nt flanking the primary lesion. Additional genetic markers may also arise from studies on spontaneous novobiocin resistant isolates that target gyrase. Current efforts are underway to repair *pyrE-64* by directed recombination using a suicide vector and, by complementation using a *groESp::pyrE* promoter fusion fragment carried on a replicating shuttle vector based on a synthetic copy of pRQ7 fused to pUC19. Recombination strategies will target carbon catabolic pathways and components of the multiple hydrogenases while complementation strategies will be used to import new traits into *Tma*. Both genetic approaches will support collaborative efforts on transcriptomics, studies of the *Tma* toga and metabolic modeling.

111

Carbohydrate Fermentation to Hydrogen by Hyperthermophilic *Thermotoga* Communities

Student Presentation

A.D. Frock,^{1*} S.R. Gray,¹ J.S. Notey,¹ K.M. Noll,² P.H. Blum,³ and R.M. Kelly¹ (rmkelly@ncsu.edu)

¹Dept. of Chemistry and Biomolecular Engineering, North Carolina State University, Raleigh; ²Dept. of Molecular and Cell Biology, University of Connecticut, Storrs; and ³Beadle Center for Genetics, University of Nebraska, Lincoln

Project Goals (Abstracts 111-112): Objective 1—Examine the regulation of substrate catabolic proteins and pathways as this relates to carbon partitioning, disposition of reducing power, and H₂ generation in *Thermotoga maritima*. Objective 2—Dissect catabolic and regulatory pathways using genetic approaches based on past success with other hyperthermophiles. Objective 3—Thermotogales biodiversity arises from adaptive specialization that expands on a conserved minimal genome; physiological characterization of selected novel traits will be done to expand understanding of bihydrogenesis.

Members of the hyperthermophilic bacterial genus *Thermotoga* are of special interest for biological hydrogen production due to high yields and the ability to use a broad range of complex carbohydrates. These high hydrogen yields are related to a more narrow range of fermentation products characteristic of hyperthermophiles compared to mesophiles. High temperature bioconversion also benefits from reduced risk of contamination and less recalcitrant biomass.

Comparison of several hyperthermophilic *Thermotoga* species (*T. maritima*, *T. neapolitana*, *T. petrophila*, and *T. sp. RQ2*) revealed that these species share a “core genome” of about 1500 genes, which generally includes most genes and pathways involved in central metabolism. Among genes involved in converting carbohydrates to hydrogen, those involved in carbohydrate degradation and transport appear to be more divergent among these species. In order to study species-specific characteristics and inter-species interactions in mixed culture, we have created a multi-species genus level cDNA microarray. This array is an expansion of a pre-existing *T. maritima* whole genome array, to which unique genes from *T. neapolitana*, *T. petrophila*, and *T. sp. RQ2* were added.

Hydrogen production rate was measured for these four species in pure culture, as well as four-species mixed culture, during growth on a simple substrate (glucose) and a complex substrate (mix of seven different polysaccharides). Under these conditions there was no significant variation in growth rate, final cell density, or hydrogen production, although it should be noted that substrate consumption and hydrogen yield were not measured in these experiments.

The multi-species microarray also presents the possibility of following the evolution of the community structure for

the mixed culture. Differentiation of these closely-related species by visual inspection was not possible nor were 16S rRNA approaches appropriate, given the 99%+ identity among these species. However, the multi-species microarray could be used to “count” each member species in the mixed culture. The first step for this method is identification of cDNA probes which are unique for one of the four species. Unique probes were predicted by BLAST against the four genomes, and the probes were experimentally tested by isolating genomic DNA from each of the species and hybridizing to the multi-species array. Initial results indicate that approximate enumeration of these species is possible provided that “unique” probes have been confirmed by microarray hybridization and that cDNA hybridized to the array is not present in spot-saturating quantities. This array is now being used to follow the evolution of the mixed community in batch and chemostat culture.

112

Functional Genomic Analysis of the Microbial Ecology of Hyperthermophilic *Thermotoga* Species

J.S. Notey,¹ S.R. Gray,¹ A.D. Frock,¹ K.M. Noll,² P.H. Blum,³ and R.M. Kelly¹ (rmkelly@eos.ncsu.edu)

¹Dept. of Chemistry and Biomolecular Engineering, North Carolina State University, Raleigh; ²Dept. of Molecular and Cell Biology, University of Connecticut, Storrs; and ³Beadle Center for Genetics, University of Nebraska, Lincoln

Project Goals: See goals for abstract 111.

The completed genome sequences of several hyperthermophilic *Thermotoga* species have given rise to functional genomic-based studies of the microbial ecology of these bacteria in pure and mixed cultures. Transcriptional response analysis of *T. maritima* growing syntrophically with the hyperthermophilic archaeon *Methanocaldococcus jannaschii* triggered quorum sensing-behavior and led to identification of a putative signaling peptide responsible for exopolysaccharide production (Johnson et al., 2006; Montero et al., 2006). Co-culture of *T. maritima* with another hyperthermophilic archaeon *Pyrococcus furiosus* led to identification to a genome locus containing putative bacteriocins and toxin-antitoxin loci (Montero, 2005; Gray et al., in preparation). Recent work focusing on multispecies cultures of hyperthermophilic *Thermotoga* species induced transcription of ORFs in this same locus, suggesting that this segment of the genome encoded genes important for ecological interactions. The genomes of *T. maritima*, *T. petrophila*, *T. sp. RQ2*, and *T. neapolitana*, as well as their transcriptomes in pure and mixed cultures, were analyzed to examine the similarities and differences among these bacteria with respect to microbial ecology. Also, experiments underway describe our current efforts to understand interspecies interactions in high temperature biotopes.

References

1. Johnson, M.R., S.B. Connors, C.I. Montero, C.J. Chou, K.R. Shockley, and R.M. Kelly. 2006. "The *Thermotoga maritima* Phenotype Is Impacted by Syntrophic Interaction with *Methanococcus jannaschii* in Hyperthermophilic Coculture. *Appl Environ Microbiol* 72 (1), 811–18.
2. Montero, C.I., D.L. Lewis, M.R. Johnson, S.B. Connors, E.A. Nance, J.D. Nichols, and R.M. Kelly. 2006. "Colocalization of Genes Encoding a tRNA-mRNA Hybrid and a Putative Signaling Peptide on Complementary Strands in the Genome of the Hyperthermophilic Bacterium *Thermotoga maritima*. *J Bacteriol* 188 (19), 6802–07.

113

Systems-Level Understanding of *Thermotoga maritima* – The Transcriptional Architecture

DaeHee Lee^{1*} (dlee@bioeng.ucsd.edu), Dmitry Rodionov,² Vasily Portnoy,¹ **Andrei Osterman**,² **Bernhard O. Palsson**,¹ and **Karsten Zengler**¹

¹Dept. of Bioengineering, University of California, San Diego; and ²Burnham Institute for Medical Research, La Jolla, Calif.

Project Goals: This project is focused on a systems-level understanding of biological hydrogen production using *Thermotoga maritima* as a model organism. The project will address the basic science required to improve our understanding of hydrogen production from various carbon sources including glucose, cellulose, starch and xylan by this thermophilic microorganism. The overall goal is 1) to reconstruct the regulatory and metabolic network in *T. maritima* using various sets of "omics" data, 2) to integrate regulatory and metabolic networks into one "integrated" genome-scale model, 3) to confirm and validate the ability of the integrated model to predict processing of various environmental signals.

This project is focused on a systems-level understanding of biological hydrogen production using *Thermotoga maritima* as a model organism. We therefore have integrated a metabolic reconstruction of *T. maritima* that contains 479 metabolic genes, 565 metabolites (non-unique) and 646 internal and external metabolic reactions with 478 protein structures to generate the first three-dimensional reconstruction of the central metabolic network of a bacterium. To understand how the flow of information from the genome to the different states of the metabolic network is archived we studied the transcriptome architecture of *T. maritima*. A protocol for chromatin immunoprecipitation (ChIP) has been adapted to work for *T. maritima*. Furthermore, we developed a method that allows for the genome-wide determination of transcription start sites (TSSs) with a single base-pair resolution. Genome-wide transcription profiles using high-density tiled arrays were integrated with binding regions of RNA polymerase and TSS information to generate an experimentally verified map of the transcriptional landscape, laying the foundation of the experimental elucidation of the operon structure in *T. maritima*. In addition we used an integrated

approach to systematic annotation and reconstruction of transcriptional regulons in the available genomes of the Thermotogales. Two major components of this analysis are (i) annotation and propagation of previously known regulons from model organisms to others (e.g., arabinose regulon AraR, arginine regulon ArgR), and (ii) *ab initio* prediction of novel regulons (e.g., inositol regulon InoR, mannose regulon ManQ). In addition to playing a key role in regulon reconstruction, regulons provide an additional layer of genome context, helping to significantly improve the accuracy of functional annotations and metabolic reconstruction. Comparative analysis of the InoR regulon led to a discovery of a new pathway of inositol catabolism, which is currently under experimental investigation by in vitro enzymatic assays.

114

A Report on the Investigation of the Cell Envelope Proteins of *Thermotoga maritima*

Amanda K. Petrus^{1*} (amanda.petrus@uconn.edu), Paul H. Blum,² Robert M. Kelly,³ and **Kenneth M. Noll**¹

¹Dept. of Molecular and Cell Biology, University of Connecticut, Storrs; ²Beadle Center for Genetics, University of Nebraska, Lincoln; and ³Dept. of Chemical and Biomolecular Engineering, North Carolina State University, Raleigh

Project Goals: We have undertaken an effort to purify the toga fraction of *T. maritima* cells and to identify the proteins that constitute the toga when cells grow on different substrates.

The outer envelope of a bacterium creates the interface that controls that organism's interactions with its environment and its utilization of carbon and energy sources. This cell envelope is a very important cell component when those sources are large and insoluble polymers, such as plant-derived polysaccharides. These bulky food sources require enzymatic degradation prior to being taken up into the cell. Thermophilic bacteria of the order Thermotogales have outer envelopes formed by a structure called the "toga" (1). The toga balloons over the cytoplasmic aspect of these cells forming a pronounced periplasmic space. The goal of this study is to characterize the proteins that compose this toga and to determine their roles in allowing cells to utilize complex, insoluble polysaccharides. These compounds can serve as a renewable energy source for the biohydrogenesis carried out by these cells.

Thermotoga maritima is the most extensively studied species of the Thermotogales. It has an optimal growth temperature of 77° C and can grow on a variety of simple and complex sugars, which lead to the production of carbon dioxide, hydrogen and acetic acid as the major products of fermentation. A complete genome sequence is available for *T. maritima* and several of its sugar hydrolases have been isolated and characterized (2). By contrast, there are only two major toga structural proteins currently identified, OmpA and

Omp β , and little is known about their functions (3). No evidence has been reported of lipids in the outer envelope.

Omp β is a porin protein that constitutes a large fraction of the toga (4). Omp α is a rod-shaped spacer protein that connects the outer envelope to the cell. Its carboxy terminus is hydrophobic and most likely anchored into the Omp β layer. It also remains associated with that layer in the parts of the toga that dissociate from the cytoplasmic membrane (3). It is not clear if Omp α is attached by its amino terminus and, if so, whether it is attached to the cytoplasmic membrane or the peptidoglycan layer. In addition to these structural proteins, at least two sugar hydrolases, a xylanase (1) and an amylase (5), have been identified in the toga fraction.

We have undertaken an effort to purify the toga fraction of *T. maritima* cells and to identify the proteins that constitute the toga when cells grow on different substrates. We have successfully utilized a freeze-thaw/homogenization mechanical shearing method to selectively release toga proteins without major contamination by cytoplasmic proteins. Our efforts have resolved the fraction using 2-D electrophoresis and the resulting protein spots will be subjected to analysis by mass spectrometry, with the goal of identifying those proteins by referencing the annotated genome sequence. Though the genome sequence is available, the gene encoding Omp β has yet to be identified. Omp α is encoded by ORF TM1729. The toga fraction has been resolved to fewer than twenty proteins on 2-D gels and a major spot migrating to a position consistent with a putative Omp β ORF has been found. The toga fraction obtained in our study will be analyzed by mass spectrometry to identify the sequence of the Omp β protein as well as other proteins associated with the toga.

References

1. Liebl, W., Winterhalter, C., Baumeister, W., Armbrrecht, M., Valdez, M. 2008. Xylanase attachment to the cell wall of the hyperthermophilic bacterium *Thermotoga maritima*. *J. Bacteriol.* 190:1350-1358.
2. Nelson, K. E., Clayton, R. A., Gill, S. R., Gwinn, M. L., Dodson, R. L., Haft, D. H., Hickey, E. K., Peterson, J. D., Nelson, W. C., Ketchum, K. A., McDonald, L., Utterback, T. R., Malek, J. A., Linher, K. D., Garrett, M. M., Stewart, A. M., Cotton, M. D., Pratt, M. S., Phillips, C. A., Richardson, D., Heidelberg, J., Sutton, G. G., Fleischmann, R. D., Eisen, J. A., White, O., Salzberg, S. L., Smith, H. O., Venter, J. C., Fraser, C.M. 1999. Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of *Thermotoga maritima*. *Nature.* 399:323-329.
3. Engel, A. M., Brunen, M., Baumeister, W. 1993. The functional properties of Omp β , the regularly arrayed porrin of the hyperthermophilic bacterium *Thermotoga maritima*. *FEMS Microbiol. Lett.* 109:231-236.
4. Rachel, R., Engel, A. M., Huber, R., Stetter, K. O., Baumeister, W. 1990. A porin-type protein is the main constituent of the cell envelope of the ancestral bacterium *Thermotoga maritima*. *FEBS Lett.* 262:64-68.
5. Schumann, J., Wirba, A., Jaenicke, R., Stetter, K.O. 1991. Topographical and enzymatic characterization of amylases from the extremely thermophilic eubacterium *Thermotoga maritima*. *FEBS Lett.* 282:122-126.

115

Structure-Assisted Modeling of a Metabolic Interactome in *Thermotoga maritima*

Ying Zhang,¹ Marat Kazanov,¹ Xiaoqing Li,¹ Dmitry Rodionov,¹ **Karsten Zengler,**² **Bernhard O. Palsson,**² Kenneth Noll,⁴ Adam Godzik,¹ and **Andrei Osterman**^{1,3*} (osterman@burnham.org)

¹Burnham Institute for Medical Research, La Jolla, Calif.; ²Dept. of Bioengineering, University of California, San Diego, La Jolla; ³Fellowship for Interpretation of Genomes, Burr Ridge, Ill.; and ⁴Dept. of Molecular and Cell Biology, University of Connecticut, Storrs

Project Goals: This project is focused on a systems-level understanding of biological hydrogen production using *Thermotoga maritima* as a model organism. The project will address the basic science required to improve our understanding of hydrogen production from various carbon sources including glucose, cellulose, starch and xylan by this thermophilic microorganism. The overall goal is 1) to reconstruct the regulatory and metabolic network in *T. maritima* using various sets of “omics” data, 2) to integrate regulatory and metabolic networks into one “integrated” genome-scale model, 3) to confirm and validate the ability of the integrated model to predict processing of various environmental signals.

Genomics-based metabolic reconstruction technology strongly impacts fundamental understanding of cellular organisms and drives multiple applications in bioengineering. However, a shortcoming of a traditional metabolic reconstruction is its “structure-blindness” with respect to atomic-level interactions between metabolic enzymes and metabolites. A recent progress in high-throughput structure determination led to a nearly complete coverage of a relatively limited fold space of protein families that comprise metabolic networks of model bacteria, opening an opportunity to bridge a gap in structural understanding of *metabolic interactome*. This study extends our previously reported integration of genomics-based metabolic reconstruction of *T. maritima* with experimentally determined and computationally modeled 3D structures towards mapping of molecular interactions between metabolic proteins (enzymes, transporters, transcriptional regulators) and their cognate ligands (substrates, products, cofactors, effectors). Metabolite cross-docking appears to be one of the promising approaches to modeling such interactions, identification of ligand-binding sites and even prediction of specific ligands for proteins of unknown function. We have assessed publically available docking tools for their ability of accurate enzyme-substrate recognition. Cross-docking against a comprehensive set of small-molecule metabolites and comparative analysis of score distribution was performed for a panel of ~ 50 enzymes conserved and essential in the metabolic network of *T. maritima* (and most other bacteria). While showing some encouraging trends, this analysis revealed many limitations of a brute-force global cross-docking approach. Some of these limitations may be partially resolved by narrowing

down a set of compared enzymes and ligands as illustrated by the analysis of six groups of FGGY sugar kinase family with distinct substrate specificities. Representatives of these groups were identified in *T. maritima* and experimentally characterized in our previous study (reported at DOE-GTL, 2009). A comparative analysis of substrate (sugars) and product (sugar-phosphates) ranking allowed us to identify conformations that improved or impaired “dockability” in a panel of experimental 3D structures and homology-based 3D models. Combining limited structural data with massive comparative sequence analysis allows us to accurately map Specificity Determining Residues (SDR) in large protein families with variations in substrate specificity. This is also illustrated by the example of FGGY sugar kinase family where >800 proteins from hundreds microbial genomes were analyzed by a modified mutual information-based method of Gelfand and Mirny. This analysis revealed a combination of divergent and convergent scenarios in the evolution of substrate specificity within this large and functionally versatile family. A similar approach was successfully applied for the analysis of specificity evolution in a family of transcriptional regulators from ROK family represented by six proteins from *T. maritima* that were identified and characterized in this project. Structure-assisted identification of functional sites in enzymes and other metabolic proteins is expected to improve accuracy of assignment, cross-genome projection and prediction of previously unknown gene functions. To capture and provide access to this and other types of information about genes and proteins supporting evolutionary and systems-level analysis of *T. maritima*, we initiated a development of the WIKITOGA web site. This web site will integrate several types of automated annotations (structural, functional, regulatory) and modeling tools with Wikipedia-style community contribution. Its content will be gradually extended from the initial focus on metabolic and regulatory networks of *T. maritima* towards whole-genome analysis of all Thermotogales species.

116

Using Hydrogen Isotopes to Assess Proton Flux during Biological Hydrogen Production

1. Determining Fractionation Factors and the Proton Transfer Pathway in Hydrogenases

Helen W. Kreuzer^{1*} (Helen.Kreuzer@pnl.gov), Hui Yang,² Liang Shi,¹ Nathaniel E. Ostrom,³ Hasand Gandhi,³ and **Eric L. Hegg**² (EricHegg@msu.edu)

¹Pacific Northwest National Laboratory, Richland, Wash.; ²Dept. of Biochemistry and Molecular Biology and ³Dept. of Zoology, Michigan State University, East Lansing

Project Goals: To improve our understanding of biological H₂ production using H/D isotope ratios.

Biological H₂ production by hydrogenase enzymes has enormous potential as an environmentally sustainable source of energy. Hydrogenases, found throughout nature in many diverse organisms, are among the most efficient

H₂-producing catalysts known. Although considerable progress has been made in elucidating the metabolic pathways involved in H₂ metabolism, many uncertainties remain. One major impediment to improving our understanding of H₂ metabolism is our inability to adequately define the regulation of and the flux through key pathways involved in H₂ production. Thus far, few attempts have been made to utilize hydrogen isotopes to improve our knowledge of the H₂ metabolic pathways, perhaps because the source of protons for hydrogenase enzymes is intracellular water. Until recently, intracellular water was generally assumed to be isotopically equivalent to extracellular water, and therefore it was perhaps thought that hydrogen isotopes would not be informative. We are exploiting our recent discoveries that intracellular water can be isotopically distinct from extracellular water, and that the contribution of protons from metabolic substrates to intracellular water can be quantified, to develop the use of hydrogen isotopes for studying intracellular proton trafficking.

We predicted that the isotope ratio of H₂ produced by various hydrogenases would differ because of slight differences in the active sites and proton transfer pathways. We further predicted that we can measure this difference via isotope-ratio mass spectrometry, and that the H/D isotope ratios would allow us to address fundamental questions concerning biological H₂ production including the source of the H₂. To test this predictions, we purified five different hydrogenases (three [FeFe]-H₂ases and two [NiFe]-H₂ases) and established conditions that allowed us to quantify the specific activity of the purified H₂ases and the amount of H₂. In addition, we built a custom chromatographic system for the analysis of H₂ and interfaced this system with an isotope ratio mass spectrometer (IRMS). We obtained a reproducibility of better than 3‰ for δD, and this precision is maintained down to a lower sensitivity limit of 0.2 μmol H₂ in 1 mL of headspace volume. In addition, because H₂ diffuses so readily through most materials (resulting in isotopic fractionation), it was necessary to develop and validate a robust protocol for capturing biologically-produced H₂. Subsequently, reaction conditions were developed that allow for the reproducible formation and capture of the optimal concentration of H₂.

Using the enzymes and optimized protocols established above, we determined the isotope ratio of the H₂ produced by three different [FeFe]-H₂ases (*Clostridium pasteurianum*, *Shewanella oneidensis*, and *Chlamydomonas reinhardtii*) and two [NiFe]-H₂ases (*Shewanella oneidensis* and *Desulfovibrio fructosovorans*). Significantly, the data indicate that all 5 hydrogenases produce H₂ with a unique isotopic signature. This proves our initial hypothesis, that different H₂-producing enzymes have different fractionation factors, and that these differences are reflected in the isotope ratio of the H₂.

Building on these results, we are using this data to help elucidate the proton transfer pathway in [FeFe]-H₂ases. To accomplish this task, we mutated a number of residues proposed to be critical for proton transport in the *C. pasteurianum* enzyme. Mutations that affect the proton pathway will change the fractionation factor by changing the ener-

genetics of proton vs. deuteron migration, while mutations that affect H_2 ase activity by other mechanisms will not change the fractionation factor. We have generated nine variants and all have been shown to alter H_2 ase activity. Future studies will ascertain if these mutations also alter the H/D isotope ratio of the H_2 .

In the second phase of the project, we are also utilizing the experimentally-determined H_2 ase fractionation factors for in vivo studies to identify the major H_2 -forming pathway in *S. oneidensis* under a variety of growth conditions (see accompanying poster).

117

Using Hydrogen Isotopes to Assess Proton Flux during Biological Hydrogen Production 2. In Vivo Studies with *Shewanella oneidensis*

Eric A. Hill,¹ James J. Moran^{1*} (James.Moran@pnl.gov),
Eric L. Hegg,² and Helen W. Kreuzer¹

¹Pacific Northwest National Laboratory, Richland, Wash. and ²Dept. of Biochemistry, Michigan State University, East Lansing

Project Goals: To improve our understanding of biological H_2 production using H/D isotope ratios.

Biological hydrogen production by hydrogenase enzymes represents a potentially sustainable, non-polluting source of energy. In order to fully exploit the capabilities of biological systems for hydrogen production, a deeper understanding of regulation of and fluxes through key pathways involved in hydrogen production is needed. We have developed methods using stable isotope measurements that allow us to trace protons from extracellular water and metabolic substrates into intracellular water and cellular metabolites, including hydrogen gas. In this part of our project, we are applying those methods to elucidate hydrogen production pathways in vivo, using *S. oneidensis* as a model organism.

Shewanella oneidensis MR-1 is a facultative anaerobe capable of transferring electrons to a variety of terminal acceptors including iron, manganese, and other metals. *S. oneidensis* encodes two hydrogenase enzymes, the [FeFe]-hydrogenase HydA and the [NiFe]-hydrogenase HyaB. Hydrogenases catalyze the reversible reaction of protons plus electrons to form hydrogen gas. If electron acceptors in its growth environment are limited, *S. oneidensis* reduces protons, producing hydrogen gas.

The objectives of this portion of our project are to determine:

- **What is the contribution of each hydrogenase enzyme to hydrogen production?** When the organism is perturbed, does the manipulation differentially affect the flux through each hydrogenase?
- **Is there channeling of protons between organic substrates and hydrogenases?** Although it is well estab-

lished that the addition of organic substrates increases H_2 production under certain conditions (e.g. acetate to the growth media of green algae or glucose to the growth media of cyanobacteria), the precise mechanism by which this occurs is not entirely clear. Some have suggested that protons are directly channeled from specific substrates into hydrogen production. We will test this hypothesis.

We have characterized hydrogen production by wild-type and electron transfer-deficient *S. oneidensis* strains. When the wild-type organism is cultured in sealed headspace vials with limited electron acceptors, the headspace hydrogen concentration initially increases, then decreases, then steadily increases. In strains deficient in metal reduction, we do not observe the decrease in headspace H_2 concentration, demonstrating that the electron transport deficiency impacts hydrogenase activity.

We hypothesize that we will be able to use stable isotopes to dissect proton fluxes through the two hydrogenase enzymes under these and other culture conditions. We have determined that HydA and HyaB evolve isotopically distinct hydrogen from the same substrate water (presented in a companion poster). In this, the second phase of our project, we are measuring the stable isotope content of the intracellular water that presumably is the substrate for hydrogenase activity, as well as that of the hydrogen gas produced in vivo, in the wild-type and mutant strains. Similar measurements will enable us to dissect proton trafficking in the presence of different organic substrates.

118

Systems Biology of Hydrogen Regulation in *Methanococcus maripaludis*

Qiangwei Xia,¹ Erik L. Hendrickson,¹ Sung Ho Yoon,² Thomas J. Lie,³ William B. Whitman,⁴ Murray Hackett,¹ Nitin Baliga,² and John A. Leigh^{3*} (leighj@u.washington.edu)

¹Dept. of Chemical Engineering, University of Washington, Seattle; ²Institute for Systems Biology, Seattle, Wash.; ³Dept. of Microbiology, University of Washington, Seattle; and ⁴Dept. of Microbiology, University of Georgia, Athens

<http://maggie.systemsbiology.net>
<http://faculty.washington.edu/leighj/>

Project Goals: 1. Use transcriptomics, proteomics, and metabolomics to study the systems biology of H_2 metabolism, formate metabolism, nitrogen fixation, and carbon assimilation in *Methanococcus maripaludis*. 2. Determine the mechanism of H_2 sensing and transcriptional regulation by H_2 .

Background

We are engaged in a long-term effort to understand regulatory networks in hydrogenotrophic methanogens, members

of the Archaea whose energy metabolism specializes in the use of H_2 to reduce CO_2 to methane. (Many hydrogenotrophic methanogens can use formate as an alternative to H_2 and CO_2). Our studies focus on *Methanococcus maripaludis*, a model species with good laboratory growth characteristics, facile genetic tools, and a tractable genome of 1722 annotated ORFs. Much of our work to date has focused on the response that occurs when supplies of essential nutrients are decreased to growth-limiting levels. Thus, we have studied the responses to H_2 limitation, nitrogen limitation, phosphate limitation, and leucine limitation (using a leucine auxotroph) (1-3). A key aspect of our approach is the use of continuous culture for maintaining defined nutrient conditions (4).

Proteomics

We used high-coverage quantitative proteomics to determine the response of *M. maripaludis* to growth-limiting levels of H_2 , nitrogen, and phosphate (1). Six to ten percent of the proteome changed significantly with each nutrient limitation. H_2 limitation increased the abundance of a wide variety of proteins involved in methanogenesis. However, one protein involved in methanogenesis decreased: a low-affinity [Fe] hydrogenase, which may dominate over a higher-affinity mechanism when H_2 is abundant. Nitrogen limitation increased known nitrogen assimilation proteins. In addition, the increased abundance of molybdate transport proteins suggested they function for nitrogen fixation. An apparent regulon governed by the euryarchaeal nitrogen regulator NrpR was identified. Phosphate limitation increased the abundance of three different sets of proteins, suggesting that all three function in phosphate transport. The global proteomic response of *M. maripaludis* to each nutrient limitation suggests a wider response than previously appreciated. The results give new insight into the function of several proteins, as well as providing information that should contribute to the formulation of a regulatory network model.

Five different approaches were compared for measuring protein abundance ratios (5). The results suggest that at the limit of deep sampling, frequency based measurements are competitive with metabolic stable isotope labeling in terms of power to detect abundance change. In addition, false discovery rates and local false discovery rates were compared as complementary approaches to multiple hypothesis testing for quantitative significance. These findings will be discussed in detail in a poster by M. Hackett.

Transcriptomics and metabolomics

A tiling array was designed and used to measure gene expression changes along a growth curve. An initial transcriptome map has been constructed and is currently being hand-annotated. The results will be discussed in detail in a poster by S.H. Yoon and N. Baliga. In addition, methods are being worked out for the measurement of key metabolites.

Hydrogen regulation and metabolism

mRNA levels for key enzymes of methanogenesis are regulated by H_2 availability (3). Results to be presented suggest that this regulation relies on sensing of some intracellular redox indicator, rather than on the external H_2 concentration.

We are also investigating whether H_2 is a necessary intermediate during growth on formate (6). This is of interest in the context of the potential for H_2 production by nitrogenase during growth on formate. If H_2 is not a necessary intermediate, then we should be able to eliminate enzymes that would deplete H_2 that is produced by nitrogenase. Experiments are underway to test the essentiality of genes whose products could produce H_2 during growth on formate.

This research was supported by the Office of Science (BER), U.S. Department of Energy, Award No. DE-FG02-08ER64685.

References

1. Xia Q, *et al.* (2009) Quantitative proteomics of nutrient limitation in the hydrogenotrophic methanogen *Methanococcus maripaludis*. *BMC Microbiol* 9:149.
2. Hendrickson EL, *et al.* (2008) Global responses of *Methanococcus maripaludis* to specific nutrient limitations and growth rate. *J Bacteriol* 190:2198-2205.
3. Hendrickson EL, Haydock AK, Moore BC, Whitman WB, and Leigh JA (2007) Functionally distinct genes regulated by hydrogen limitation and growth rate in methanogenic Archaea. *Proc Natl Acad Sci U S A* 104:8930-8934.
4. Haydock AK, Porat I, Whitman WB, and Leigh JA (2004) Continuous culture of *Methanococcus maripaludis* under defined nutrient conditions. *FEMS Microbiol Lett* 238:85-91.
5. Xia Q, Wang T, Beck DAC, Taub F, Leigh JA, and Hackett M (2010) Quantitative local false discovery rates, deep sampling and detecting protein abundance change for the model organism *Methanococcus maripaludis*. *Proteomics* (in review).
6. Lupa B, Hendrickson EL, Leigh JA, and Whitman WB (2008) Formate-dependent H_2 production by the mesophilic methanogen *Methanococcus maripaludis*. *Appl Environ Microbiol* 74:6584-6590.

119

Quantitative Local False Discovery Rates, Deep Sampling and Protein Abundance Change for *Methanococcus maripaludis*

Qiangwei Xia,^{1,2,4} Tiansong Wang,² David A.C. Beck,³ Fred Taub,² **John A. Leigh,² and Murray Hackett^{1*}** (mhackett@u.washington.edu)

¹Dept. of Chemical Engineering, ²Dept. of Microbiology, and ³eScience Institute, University of Washington, Seattle; and ⁴Dept. of Chemistry, University of Wisconsin, Madison

Project Goals: 1. To find the best balance between statistical power to detect protein abundance change and the need to decrease the time required for a complete proteome analysis for *Methanococcus maripaludis*. 2. To establish just how much sampling is required for spectral counting to become as efficient or perhaps even more efficient for generating protein abundance ratios relative to traditional metabolic stable isotope labeling. The answer to this question is heavily dependent on the specific mass

spectrometry instrumentation and methods used in the investigation, both of which have experienced several upgrades and improvements since we last addressed this question in the peer reviewed literature in 2006. 3. To establish the most efficient data reduction and transformation procedures for the use of local false discovery rates (lfdr) with spectral counting data. Lfdr is one approach among several to multiple hypothesis testing, based on the more general concept of false discovery rate. Although proposed initially in the context of microarray-based transcriptome data, lfdr is equally applicable to large-scale quantitative proteomic experiments and RNA-Seq.

Protein abundance ratios were measured using five different approaches for the Archaeon *Methanococcus maripaludis*, a model organism of interest for analytical studies because of the tractable nature of its proteome in terms of size, proteome extraction efficiency and other positive features. Multidimensional capillary HPLC coupled with tandem mass spectrometry was used for analysis of heavy (^{15}N) and natural abundance (^{14}N) tryptic digests of *M. maripaludis* grown in chemostats. Here we report our comparison of abundance ratios based on heavy and light proteomes mixed prior to mass spectrometry; spectral counting of heavy and light proteomes mixed; spectral counting of heavy and light proteomes analyzed separately; summed signal intensities for mixed heavy and light proteomes; and summed signal intensities for heavy and light proteomes analyzed separately. Protein identifications were saturated and proteome penetration maximized at $\sim 91\%$ of the predicted protein-encoding open reading frames. False discovery rates and local false discovery rates were compared as complementary approaches to multiple hypothesis testing for quantitative significance. Power calculations, dynamic range and other observations reported suggest that at the limit of deep sampling frequency based measurements are competitive with metabolic stable isotope labeling in terms of power to detect abundance change.

This research was supported by the Office of Science (BER), U.S. Department of Energy, Award No. DE-FG02-08ER64685.

120 Development of Metabolic Network Models of *Rhodobacter sphaeroides* for the Prediction of Quantitative Contributors to H_2 Production

L. Safak Yilmaz^{1*} (yilmaz@cae.wisc.edu), Timothy J. Donohue² (tdonohue@bact.wisc.edu), and Daniel R. Noguera¹ (noguera@engr.wisc.edu)

¹Dept. of Civil and Environmental Engineering and
²Dept. of Bacteriology, University of Wisconsin, Madison

Project Goals: 1) Develop a metabolic network model of *Rhodobacter sphaeroides*. 2) Use this model to (a) understand the bioenergetics of growth conditions that produce hydrogen from different carbon sources, so that optimal conditions of hydrogen production can be predicted

and (b) determine reactions and pathways that detract electrons from hydrogen production, so that hydrogen production can be potentially maximized by genetic manipulations.

Introduction and Objective

The metabolically versatile organism *Rhodobacter sphaeroides* produces H_2 while using light as an energy source and organic substrates as an electron donor. The overall goal of this project is to elucidate gene networks that contribute to and detract from H_2 production and genetically modify *R. sphaeroides* to divert as high a fraction of substrate electrons as possible to H_2 generation. As a tool to achieve this goal, we are developing metabolic network models of *R. sphaeroides* which can integrate genomic knowledge and experimental data, thereby allowing the prediction of optimization strategies.

Methodology

Stoichiometric reaction network models (Palsson, 2006) were reconstructed using the KEGG database (<http://www.genome.jp/kegg/pathway.html>), the genome of *R. sphaeroides* strain 2.4.1 (<http://www.rhodobacter.org/>), and other metabolic network models available in the literature (http://gcrp.ucsd.edu/In_Silico_Organisms). Model development is divided into three stages based on the scale of metabolic pathways represented, as explained below.

Stage 1 activities built a core network model that represented central carbon metabolism. Carbon pathways including glycolysis, citric acid cycle, and the pentose-phosphate pathway are connected to an electron transport chain previously proposed for purple non-sulfur bacteria (Klamt et al., 2008). The energetic contribution of light is part of the electron transport chain, H_2 production is added as a nitrogenase reaction in the absence of N_2 based on available data (see poster by Kontur et al.). The synthesis of the known electron sink polyhydroxybutyrate (PHB) is part of the network as a pathway starting from acetyl coenzyme A. To simulate cell synthesis, precursor metabolites are assembled in a biomass reaction taken from *Escherichia coli* core network models (Palsson, 2006).

Ongoing activities in Stage 2 expands the Stage 1 model to include amino acid and CO_2 fixation pathways. The existing biomass equation (Stage 1) will also be modified based on the utilization of precursor metabolites in the new pathways and on the average protein composition of purple non-sulfur bacteria (Kobayashi and Kobayashi, 1995) to include amino acids as part of biomass.

A planned Stage 3 model will add other major metabolic pathways (lipid metabolism, nucleic acid metabolism, glycan biosynthesis, and metabolism of cofactors) to the Stage 2 model. Biomass assembly will be based on building blocks using average cell composition of purple non-sulfur bacteria (Kobayashi and Kobayashi, 1995).

Once reaction networks are constructed, the models are established in MATLAB (MathWorks, Natick, MA) and GAMS (GAMS Development Corporation, Washington, DC) to carry out in silico flux balance analysis (FBA) with

linear programming. The models are trained and tested with experimental results from a systematic analysis of electron flow in *R. sphaeroides* growing photosynthetically on a variety of carbon sources (Yilmaz et al., in review). The data include H₂ generation, PHB synthesis, biomass formation, and production of soluble microbial products (SMP) during exponential and stationary phases of batch cultures, and similar data is being generated from chemostats in ongoing work. All cultures are fed with glutamate as the sole nitrogen source to maximize H₂ production, while single organic acids (succinate, lactate, pyruvate, and fumarate) or sugars (mostly glucose) are used as the carbon source.

Results

The Stage 1 model qualitatively captures aerobic growth without H₂ production and anoxygenic photosynthetic growth with H₂ production. The model predicts that H₂ production is a necessary electron accepting pathway in photosynthetic growth, without which excess electrons cannot be recycled. Another surprising prediction is that, SMP production, which was a significant electron sink in experimental cultures (Yilmaz et al., in review), occurs during dimmer light conditions (i.e. with a limited flux of the light reaction), possibly representing shadowing in dense cultures. The FBA results always predict a larger CO₂/H₂ ratio (≥ 1.0 in partial pressure) than what is found in the headspace of batch cultures (~ 0.2). This difference is attributed to the lack of CO₂ fixation pathways in our Stage 1 model. Although unconstrained simulations do not quantitatively match experimental data, constrained FBA fits to experimental results for the production of key electron sinks.

The Stage 1 model lacks pathways to simulate glutamate consumption as a carbon, nitrogen and electron source, but the Stage 2 can account for this due to addition of amino acid pathways. This addition and the introduction of CO₂ fixation pathways to the Stage 2 model are expected to provide quantitative and qualitative predictions for the mechanisms of growth and H₂ production as a function of carbon sources. The databases used were checked for the absence of missing reactions in amino acid and carbon fixation pathways before we began reconstruction of the Stage 2 model. A major improvement in the Stage 3 model will be the ability to include a global analysis of metabolism with FBA and existing microarray data. To this end, we developed a prototype visualization tool that overlays flux and gene expression profiles to KEGG pathway maps. The poster will present results from the Stage 1 model, the current Stage 2 model, and illustrations from the visualization tool for integrated microarray and flux data.

References

1. Klamt, S, Grammel, H, Straube, R, Ghosh, R, and Gilles, E D (2008) Modeling the electron transport chain of purple non-sulfur bacteria. *Molecular Systems Biology*. 4:156.
2. Kobayashi, M, and Kobayashi, M (1995) Waste Remediation and Treatment Using Anoxygenic Phototrophic Bacteria. In: Blankenship, R E, Madigan, M T, and Bauer, C E (ed), *Anoxygenic Photosynthetic Bacteria*. Kluwer Academic Publishers Dordrecht, Netherlands, pp 1269-1282

3. Palsson, B O. 2006. *Systems Biology*. Cambridge University Press, New York, NY.
4. Yilmaz, L S, Kontur, W, Sanders, A P, Sohmen, U, and Donohue, T J. Electron Partitioning During Light- and Nutrient-Powered Hydrogen Production by *Rhodobacter sphaeroides*. *Bioenergy Research*. In review.

121

Cellular Redox Balance and the Integrative Control of Carbon Assimilation and Hydrogen Production in *Rhodobacter sphaeroides* and *Rhodospseudomonas palustris*

Rick A. Laguna* (laguna.2@osu.edu) and F. Robert Tabita

Dept. of Microbiology, Ohio State University, Columbus

Project Goals: Elucidate molecular mechanisms by which evolved strains of NSP bacteria derepress the synthesis of nitrogenase, allowing this enzyme system to catalyze the production of high levels of hydrogen gas. Perform computational and molecular modeling studies to identify further control circuits that might be manipulated to improve hydrogen production. Determine how these different regulatory networks are integrated in the cell. Determine how various regulatory protein complexes we have implicated contribute towards regulating hydrogen production. Examine the intracellular organization of these complexes and determine how small effector metabolites influence their function. Integrate these studies to engineer the most efficient strain for maximal hydrogen production.

Nonsulfur purple (NSP) photosynthetic bacteria are characterized by their metabolic versatility. These organisms are capable of growth during photosynthetic and non-photosynthetic conditions, in the absence or presence of oxygen, respectively, and they can synthesize cell mass via the assimilation of either organic or inorganic carbon sources, with needed energy obtained via photochemical or dark chemical processes. During aerobic chemolithoautotrophic and anaerobic photolithoautotrophic growth conditions, CO₂ serves as the sole carbon source and is reduced into cellular carbon by the Calvin-Benson-Bassham (CBB) reductive pentose phosphate pathway. Under photoheterotrophic growth conditions, organic carbon compounds, usually organic acids, are oxidized into cell mass. In NSP bacteria such as *Rhodobacter sphaeroides* and *Rhodospseudomonas palustris*, when malate is used as the carbon source during photoheterotrophic growth, the reduction of metabolically produced CO₂ through the CBB cycle is the preferred means by which these organisms consume excess reductant produced from the assimilation of the organic electron donor (1-3). However, when the CBB cycle is inactivated, these organisms face the problem of either not growing or adapting by using other redox balancing mechanisms. In one interesting adaptive scenario, excess reductant is consumed by the derepression of the nitrogenase complex, resulting in the production of hydrogen gas in both *Rb. sphaeroides*

and *Rps. palustris*, as well as *Rhodospirillum rubrum* and *Rb. capsulatus* (4,5). This is of interest considering that malate is more oxidized than cell mass. Therefore the need to consume excess reductant through the CBB cycle should not be required. However upon examining the assimilation of malate in *Rb. sphaeroides* and *Rps. palustris* (both assimilate malate through the tricarboxylic acid pathway) and by using various CBB mutant strains, it has been shown that excess reductant is produced and must be consumed by the CBB cycle in order to maintain cellular redox balance. The reason for this is that as malate is assimilated into cell mass in both organisms, excess reductant is generated. Acetate can also be used as a carbon source for photoheterotrophic growth in *Rb. sphaeroides* and *Rps. palustris*. Unlike malate, acetate is at the same oxidation state as cell mass. Therefore excess reductant should not be generated by the assimilation of acetate and the CBB cycle should not be required. In *Rb. sphaeroides* this is the case; however in *Rps. palustris* there is a need for the CBB cycle. The reason for this is that the two organisms use different metabolic pathways to assimilate acetate. *Rb. sphaeroides* uses the ethylmalonyl-CoA pathway to assimilate acetate (6), which does not generate excess reductant from this process. Therefore CBB mutant strains of *Rb. sphaeroides* are capable of photoheterotrophic growth using acetate as the carbon source. However *Rps. palustris* uses the glyoxylate pathway to assimilate acetate (7), which does produce excess reductant. Therefore CBB mutant strains of *Rps. palustris* cannot grow photoheterotrophically with acetate as the carbon source. We are currently examining how the assimilation of different carbon sources affects redox balance and hydrogen production in *Rb. sphaeroides* and *Rps. palustris*, as well as *R. rubrum* and *Rb. capsulatus*, during photoheterotrophic growth conditions. Indications are that there are differences that might be eventually exploited to maximize hydrogen production.

References

1. Falcone, D. L. and F. R. Tabita. 1991. Expression of endogenous and foreign ribulose 1,5-bisphosphate carboxylase-oxygenase (RubisCO) genes in a RubisCO deletion mutant of *Rhodobacter sphaeroides*. *J. Bacteriol.* 173:2099-2108.
2. Romagnoli, S. and F. R. Tabita. 2006. A novel three-protein two-component system provides a regulatory twist on an established circuit to modulate expression of the *cbbI* region of *Rhodopseudomonas palustris* CGA010. *J. Bacteriol.* 188:2780-2791.
3. Joshi, G. S., Romagnoli, S., VerBerkmoes, N. C., Hettich, R. L., Pelletier, D., and Tabita, F. R. 2009. Differential accumulation of form I RubisCO in *Rhodopseudomonas palustris* CGA010 under photoheterotrophic growth conditions with reduced carbon sources. *J. Bacteriol.* 191:4243-4250.
4. Joshi, H. M. and F. R. Tabita. 1996. A global two component signal transduction system that integrates the control of photosynthesis, carbon dioxide assimilation, and nitrogen fixation. *Proc. Natl. Acad. Sci. U. S. A.* 93:14515-14520.
5. Tichi, M. A., and Tabita, F.R. 2000. Maintenance and control of redox poise in *Rhodobacter capsulatus* strains deficient in the Calvin-Benson-Bassham pathway. *Arch. Microbiol.* 174:322-333.

6. Alber, B.E., Spanheimer, R., Ebenau-Jehle, C., and Fuchs, G. 2006. Study of an alternate glyoxylate cycle for acetate assimilation by *Rhodobacter sphaeroides*. *Mol. Microbiol.* 61:297-309.
7. Larimer, F.W., Chain, P., Hauser, L., Lamerdin, J., Malfatti, S., Do, L., Land, M.L., Pelletier, D.A., Beatty, J.T., Lang, A.S., Tabita, F.R., Gibson, J.L., Hanson, T.E., Bobst, C., Torres, J.L., Peres, C., Harrison, F.H., Gibson, J., Harwood, C.S. 2004. Complete genome sequence of the metabolically versatile photosynthetic bacterium *Rhodopseudomonas palustris*. *Nature Biotechnol.* 22:55-61.

122

Development of *Cyanotheca* as a New Model Organism for Photobiological Hydrogen Production

Anindita Banerjee,¹ Jana Stöckel,¹ Lawrence Page,¹ Xueyang Feng,² Bing Wu,² Yinjie Tang,² Hongtao Min,³ Sujata Mishra,³ Xiaohui Zhang,³ Louis A. Sherman^{3*} (lsherman@purdue.edu), Stephen J. Callister,⁴ Richard D. Smith,⁴ and Himadri B. Pakrasi^{1,2}

¹Depts. of Biology and ²Energy, Environmental and Chemical Engineering, Washington University, St. Louis, Mo.; ³Dept. of Biological Sciences, Purdue University, West Lafayette, Ind.; and ⁴Biological Sciences Division, Pacific Northwest National Laboratory, Richland, Wash.
sysbio.wustl.edu/Pakrasi/projects/hydrogen.php

Project Goals: The objective of this proposal is to develop the cyanobacterium *Cyanotheca* as a model organism for photobiological hydrogen production. Members of the genus *Cyanotheca* are unicellular oxygenic prokaryotes with the ability to fix atmospheric nitrogen. Our long-term goal is to develop a deep understanding of the metabolism of these microbes as it pertains to H₂ evolution. Specifically, we are using genome sequencing, transcriptomics, proteomics, metabolomics, mutagenesis, biochemical analysis and physiological approaches, all of which are encased in a systems biology framework.

Hydrogen production by *Cyanotheca*: *Cyanotheca* are unicellular, nitrogen-fixing cyanobacteria. They possess both the nitrogenase and hydrogenase enzyme systems that can catalyze biological hydrogen production. Hydrogen production by different *Cyanotheca* strains was assessed by employing appropriate physiological perturbations, based on knowledge from prior transcriptomic and proteomic studies. *Cyanotheca* sp. ATCC 51142 exhibited exceptionally high rates of hydrogen production under aerobic conditions. This is in striking contrast to other known unicellular photosynthetic hydrogen producing strains, which produce hydrogen under largely anaerobic conditions. The rates of hydrogen production were significantly enhanced by growing *Cyanotheca* cells in the presence of glycerol or high levels of CO₂. Both of these carbon sources enhanced the level of glycogen, an intracellular energy reserve in the cells. Our studies revealed that the high rate of hydrogen production by this strain is

largely mediated by an efficient nitrogenase enzyme system. Programmable photobioreactors equipped with sensors to monitor several critical parameters are now being used to further characterize these strains to achieve higher rates of hydrogen production.

Comparative genomics: Complete genome sequences of five *Cyanotheca* strains (ATCC 51142, PCC 7424, PCC 7425, PCC 8801, PCC 8802) are currently available and two more (PCC 7822, ATCC 51472) are in the process of completion at the DOE Joint Genome Institute. The sequences reveal significant metabolic diversity within this group of cyanobacteria. The genome sequence information is being used to generate a *Cyanotheca* pan-genome (in collaboration with JGI), comprising of the “core genome” (containing all of the genes common to each genus member) and the “dispensable genome” (containing unique genes or genes shared between two or more strains). The unique genes are likely to confer strain-specific attributes and will be analyzed for their role in hydrogen production.

Genetics: *Cyanotheca* 7822 was successfully transformed by electroporation using a modification of the asymmetric PCR technique originally developed for eukaryotic algae. The first target was *nifK*; the gene encoding one of the two subunits of the nitrogenase MoFe protein and a spectinomycin-resistance cassette was inserted in the middle of the *nifK* gene. This strain has been stable for more than 9 months and is resistant to spectinomycin and is incapable of growing on plates or in liquid medium that is lacking combined nitrogen (*i.e.*, it cannot fix atmospheric N₂). The strain is also completely unable to produce hydrogen under any growth condition, although it can still reduce acetylene at rates that are particularly high when the cultures are incubated under argon prior to measurement. This indicates that the strain can still assemble a MoFe complex, although it cannot produce the main products of N₂ fixation, H₂ and NH₃.

Metabolomic Studies: Over the past year, we have developed new metabolite extraction methods and mass spectrometry techniques. Using such new approaches, we characterized the central metabolic pathways in *Cyanotheca* 51142, and discovered a novel isoleucine biosynthesis pathway that involves citramalate synthase. We have studied impacts of carbon and nitrogen sources on the central metabolism and hydrogen production by *Cyanotheca* 51142, using both biochemical methods and ¹³C isotopomer approaches. We quantitatively determined CO₂ fixation and carbon substrate utilization under mixotrophic growth conditions in *Cyanotheca* 51142. We are currently developing ¹³C-assisted dynamic flux models to study autotrophic metabolism in cyanobacteria.

Proteomic Studies: In preparation for conducting relative quantitative proteomics analyses, using the AMT tag proteomics approach, construction of reference peptide databases for 6 of 7 *Cyanotheca* strains has been completed. Over 460 LC-MS/MS datasets have been generated and analyzed using the high-throughput proteomics capabilities at PNNL. Complete databases correspond to strains *Cyanotheca* sp. PCC 8801, PCC 8802, PCC 7424, PCC 7425, PCC 7822, and ATCC 51142. LC-MS/MS datasets for

Cyanotheca sp. ATCC 51472 have also been generated with identification of peptide sequences waiting upon the draft completion of the genome sequence (at the Joint Genome Institute) of this strain. Percent observed coverage of predicted proteins from unique peptides (10% false discovery rate) ranges from roughly 40% to 70%, which inversely correlates to the size range of genome sequences; *Cyanotheca* 7424 has the largest genome sequence (~6.5 Mb) and smallest percent observed coverage among the strains.

Project number - DE-FC02-07ER64694

123 Genome-Wide Network Analysis of Metabolism in *Chlamydomonas reinhardtii*

Roger L. Chang^{1*} (rlchang@ucsd.edu), Ani Manichaikul,² Balaji Santhanam,³ Lila Ghamsari,³ Erik F.Y. Hom,⁴ Kourosh Salehi-Ashtiani,³ and Jason A. Papin²

¹Dept. of Bioengineering, University of California, San Diego; ²Dept. of Biomedical Engineering, University of Virginia, Charlottesville; ³Center for Cancer Systems Biology and Dept. of Cancer Biology, Dana-Farber Cancer Institute, and Dept. of Genetics, Harvard Medical School, Boston, Mass.; and ⁴Dept. of Molecular and Cellular Biology, Harvard University, Cambridge, Mass.

<http://bme.virginia.edu/csbl>

<http://ccsb.dfci.harvard.edu>

Project Goals: Our goal in this study is to carry out genome-wide network reconstruction of *C. reinhardtii* metabolism integrated with structural and functional annotation, to provide a framework to integrate these metabolic pathways for investigation of global properties of metabolism.

Algae have garnered significant interest in recent years, both for their potential as a source of biofuel, as well as for their use in the production of nutritional supplements. Among eukaryotic microalgae, *Chlamydomonas reinhardtii* has established its position as an ideal model organism, popular for its relatively fast doubling time and its ability to grow under standardized conditions in the laboratory. Characterization of metabolic functions in *C. reinhardtii* provides a framework for developing engineering strategies towards generation of strains with improved production of commercial targets, as well as studying diverse cellular processes such as photosynthesis and cell motility. Extensive literature on *C. reinhardtii* metabolic function and many mutants with metabolic phenotypes provide a solid foundation toward detailed characterization of individual metabolic pathways in this organism.

Our goal in this study is to carry out genome-wide network reconstruction of *C. reinhardtii* metabolism integrated with structural and functional annotation, to provide a framework to integrate these metabolic pathways for investigation of global properties of metabolism. The availability of complete

genome sequence data (JGI v4.0), recently released for *C. reinhardtii*, allowed us to perform in-house annotation of the metabolic genes encoded within this genome (*please see our accompanying abstract by Ghamsari et al. for details*). Briefly, functional annotations enable the identification of the presence of enzymes encoded within the genome of the organism to define the scaffolding of the reconstruction. Comprehensive literature searches are used as the primary form of evidence to establish the structure of all metabolic pathways of interest. Finally, this information is supplemented with more general knowledge of metabolic pathways, as provided in classical biochemistry textbooks and also available in online databases.

We report the first genome-scale reconstruction of *C. reinhardtii*, accounting for all pathways and metabolic functions indicated by the latest release of the genome (JGI v4.0) combined with our in-house generated functional annotation. The reconstruction accounts for 978 genes, associated with 1671 reactions, and includes 1029 unique metabolites. As the most comprehensive metabolic network reconstruction of *C. reinhardtii* to date, ours is the first to account for multiple wavelengths of light involved in metabolism and includes considerable expansion of fatty acid metabolism over previous reconstructions, with pathway details accounting for metabolism of individual R-groups. Further, the metabolic network reconstruction presented here provides a greater level of compartmentalization than existing reconstructions of *C. reinhardtii*, with the inclusion of the lumen as a distinct component of the chloroplast for photosynthetic functionality, and the eyespot used to guide the flagella in phototaxis.

We present simulations under a variety of growth conditions (e.g. acetate/no acetate, light/no light, aerobic/anaerobic), and physiological validation of *in silico* gene knockout against known mutant data for a variety of phenotypes (e.g. increased use of acetate; light; CO₂; nitrogen; and other media components, amino acid requiring, altered color). We further present detailed simulations demonstrating how photon absorption and different wavelengths of light affect downstream metabolic processes, elucidating the benefits of sunlight versus artificial light conditions. Our well-validated and comprehensive genome-scale reconstruction of *C. reinhardtii* metabolism provides a valuable quantitative and predictive resource for metabolic engineering toward improved production of biofuels and other commercial targets.

124

Identifying the Metabolic Potential of *Chlamydomonas reinhardtii* by Large-Scale Annotation of its Encoded Open Reading Frames

Lila Ghamsari^{1*} (Lila_Ghamsari@dfci.harvard.edu), Balaji Santhanam,¹ Roger L. Chang,² Ani Manichaikul,³ Erik F.Y. Hom,⁴ Marc Vidal,¹ Jason A. Papin,³ and Kourosh Salehi-Ashtiani¹

¹Center for Cancer Systems Biology and Dept. of Cancer Biology, Dana-Farber Cancer Institute, and Dept. of Genetics, Harvard Medical School, Boston, Mass.; ²Dept. of Bioengineering, University of California, San Diego; ³Dept. of Biomedical Engineering, University of Virginia, Charlottesville; and ⁴Dept. of Molecular and Cellular Biology, Harvard University, Cambridge, Mass.

<http://ccsb.dfc.harvard.edu>

<http://bme.virginia.edu/csbl>

Project Goals: The release of the complete genome sequence of *C. reinhardtii* has made this unicellular alga an ideal model for metabolic engineering; however, the annotation of the relevant genes has not been verified yet and the much-needed metabolic network model is currently unavailable. Using the integrated annotation and metabolic network modeling that we recently established (Manichaikul et al., *Nature Methods* 2009), we are engaged in efforts to: 1) assign enzymatic functions to the annotated proteome of *C. reinhardtii*, 2) experimentally verify or refine the structure of the annotated open reading frames (ORFs), and 3) build a genome-wide metabolic network model for the organism based on the assigned metabolic functions (*please also see our accompanying abstract by R.L. Chang et al.*).

Results: We used the new JGI “filtered transcript models” (Chlre4_best_transcripts and Chlre4_best_proteins), and the *Augustus 5* models released through the JGI portal (<http://genome.jgi-psf.org/Chlre4/Chlre4.home.html>) for both functional assignments and structural annotation verifications. Enzymatic functional assignments were made by associating Enzyme Classification (EC) numbers through reciprocal blast searches against Uniprot enzyme database (with over 100,000 protein entries). The best match for each translated ORF was identified (with an e-value threshold of 10⁻³) and the EC number from the Uniprot best match was transferred on to the ORF. We extended the EC assignments to the respective paralogs of the ORFs by clustering ORFs using BLASTCLUST (sequence identity cut-off of 35% and sequence length cut-off of 70%) within each annotation group (i.e., *Augustus 5* and JGI filtered models). Altogether, we were able to assign 970 EC numbers to 1,448 JGI and to 1,877 *Augustus* models. Over 93% of the EC terms were assigned to both JGI and *Augustus* models (Fig. 1A). We then carried out all possible pairwise alignments between the JGI and *Augustus* transcripts that had been assigned the same EC numbers by the above-mentioned

procedure. In contrast to the high overlap between the two models in terms of EC assignments, less than half of each set were found to be 100% identical in sequence (Fig. 1B), indicating that the structural annotation of many of the two sets differ from one another.

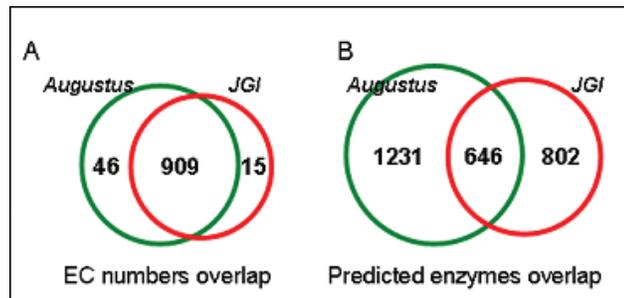


Figure 1. Annotation and overlap of *JGI* and *Augustus* ORF models.

To experimentally verify the structure of both *JGI* and *Augustus* ORF models, we carried out open reading frame (ORF) verification by RT-PCR on all ORFs that we had assigned EC numbers to (as well as a set of positive control ORFs). Following optimization of the RT-PCR procedure for high GC content of the *C. reinhardtii* transcriptome, we were able to observe positive RT-PCR products for approximately 70% of the transcripts. Following cloning, we carried out 454FLX sequencing of the ORFs and aligned the 454FLX reads to the ORF reference sequences (Fig. 2). We obtained 95-100% coverage of the ORF length for 940 of the ORF models; 215 ORFs with 50 to 95% coverage, 207 with 10 to 50%, and less than 10% coverage for the remainder. These results indicate that at least half of the transcript models are accurately annotated. Cloning and parallel sequencing of the remaining *JGI* metabolic ORFs are in progress.

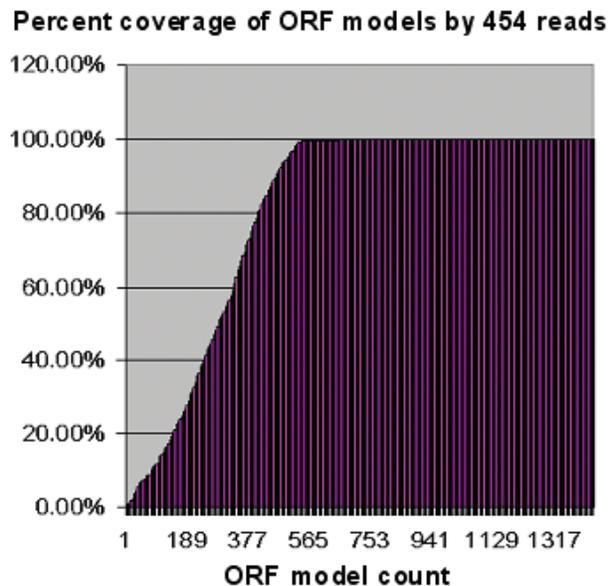


Figure 2. Sequence verification of metabolic ORF models. 1,877 ORF models were tested by RT-PCR, recombinational cloning and 454 FLX sequencing.

The verified metabolic ORF clone resource that we have generated will be made available without restrictions to the research community.

125 Pathway of Fermentative Hydrogen Production by Sulfate-Reducing Bacteria

Kimberly L. Keller^{1,2*} (kellerkl@missouri.edu), Steve D. Brown,^{3,2} Iris Porat,³ and **Judy D. Wall**^{1,2} (WallJ@missouri.edu)

¹University of Missouri, Columbia; ²VIMSS (Virtual Institute of Microbial Stress and Survival) <http://vimss.lbl.gov/>, Berkeley, Calif.; and ³Oak Ridge National Laboratory, Oak Ridge, Tenn.

Project Goals: The production of hydrogen by fermentative pathways of the anaerobic sulfate-reducing bacteria (SRB) of the genus *Desulfovibrio* is the focus of the project. The limitations to hydrogen production identified in these model organisms may be informative for those microbes chosen for industrial hydrogen generation. We propose to determine the contribution of substrate-level phosphorylation to respiratory growth on sulfate and the contribution of respiration to fermentation of pyruvate. The enzymes for pyruvate oxidation will be established in two strains of *Desulfovibrio* to identify the reduced product available for hydrogen generation. Electron sinks potentially competing with protons will be eliminated individually and together to determine the plasticity of electron flow to hydrogen.

We are exploring the production of hydrogen by fermentative pathways of the anaerobic sulfate-reducing bacteria (SRB) of the genus *Desulfovibrio*. The SRB have not been considered model organisms for hydrogen production, yet they have a multifaceted hydrogen metabolism. Strains of *Desulfovibrio* can ferment organic acids in the absence of terminal electron acceptors and produce rather large amounts of hydrogen. We believe a study of the limitations to hydrogen production in these model organisms (in particular, *Desulfovibrio* G20 and *Desulfovibrio vulgaris* Hildenborough) may be informative to decipher the flow of electrons in those organisms chosen for industrial application for hydrogen production. We have proposed to:

- Determine the contribution of fermentation to the respiratory energy budget as well as the contribution of respiration processes to the fermentation of pyruvate. Ultimately, we seek to separate these pathways and deduce the dependence of the bacterium on these combined processes.
- Identify the enzyme (s) responsible for the oxidation of pyruvate to elucidate the reduced substrate initially available for hydrogen generation during fermentation.
- Eliminate alternative electron sinks that are likely to compete for protons, reducing the overall yield of hydrogen.

An analysis of metabolic end products of *Desulfovibrio* G20 cultures grown on pyruvate as the sole electron donor and acceptor has indicated that respiration and fermentation occur simultaneously. Conversely, cells respiring sulfate with lactate also apparently gain energy by substrate level phosphorylation. We are attempting to delete the gene encoding acetate kinase to confirm that this enzyme is essential for substrate-level phosphorylation during pyruvate fermentation and determine the effect of this deletion on the efficiency of pyruvate and lactate respiration.

To determine what pathways are present for energy generation during respiration or fermentation, we have initiated an analysis of the proteomics in each condition. Cultures from wild-type *Desulfovibrio* G20 grown by pyruvate fermentation and by sulfate respiration with lactate are being used to identify enzymes which appear differentially expressed between growth modes, as well as those are found in high abundance in both growth modes. These results provide candidates for further exploration

The construction of a Tn5 transposon mutant library in *Desulfovibrio* G20 is allowing us to establish the importance of the candidate enzymes identified by the proteomics. The mutants are grown on different substrates and their growth and metabolites are compared to those of wild-type *Desulfovibrio* G20. We have already begun an analysis of hydrogen production and metabolite changes of a number of G20 Tn5 mutants, including those lacking a fumarate reductase, molybdopterin oxidoreductase, formate dehydrogenases, and malic enzyme. No growth differences were found between wild-type and mutants lacking formate dehydrogenases or malic enzyme mutants, probably because there are apparently multiple isozymes encoding these enzymes that could

compensate for the single enzyme loss. As expected, there was no succinate measured for the mutant with the Tn5 inserted in the fumarate reductase. In addition, the fumarate reductase mutant was unable to grow when fumarate only was provided as the sole electron donor and acceptor. This lack of fumarate dismutation has also been observed with a *Desulfovibrio* G20 plasmid insertion mutation in the *cycA* gene which encodes the type-1 tetraheme cytochrome c3.

Advances continue to be made in the genetic manipulation *D. vulgaris*, including a markerless deletion system allowing the sequential deletion of multiple genes. The system, which uses the *upp*-encoded uracil phosphoribosyltransferase as an element for counterselection, is being refined to increase efficiency of mutant selection. Preliminary experiments show the wild-type G20 strain is quite sensitive to 5-fluorouracil and attempts to make *Desulfovibrio* G20 more amenable to gene specific mutations continues. We are currently in the process of generating a barcoded transposon library in *Desulfovibrio vulgaris* Hildenborough.

126 Transcriptomic Analyses of the Sulfate-Reducing Bacterium *Desulfovibrio* G20 and a Type-1 Tetraheme Cytochrome c3 Mutant during Their Transitions into Stationary Phase

Iris Porat^{1*} (porati@ornl.gov), Kimberly L. Keller,^{2,3} Barbara J. Giles,² Timothy J. Tschaplinski,¹ Steve D. Brown,¹ and **Judy D. Wall**^{2,3} (WallJ@missouri.edu)

¹Oak Ridge National Laboratory, Oak Ridge, Tenn.; ²University of Missouri, Columbia; and ³VIMSS (Virtual Institute of Microbial Stress and Survival) <http://vimss.lbl.gov/>, Berkeley, Calif.

Project Goals: The production of hydrogen by fermentative pathways of the anaerobic sulfate-reducing bacteria (SRB) of the genus *Desulfovibrio*. is the focus of the project. The limitations to hydrogen production identified in these model organisms may be informative for those microbes chosen for industrial hydrogen generation. We propose to determine the contribution of substrate-level phosphorylation to respiratory growth on sulfate and the contribution of respiration to fermentation of pyruvate. The enzymes for pyruvate oxidation will be established in two strains of *Desulfovibrio* to identify the reduced product available for hydrogen generation. Electron sinks potentially competing with protons will be eliminated individually and together to determine the plasticity of electron flow to hydrogen.

Fermentative hydrogen generation provides a mechanism for anaerobic microbes to release electrons in a neutral fashion during oxidation of vast quantities of organic matter. We are examining the capacity of a soil anaerobe, the sulfate-reducing bacterium *Desulfovibrio* G20, to generate hydrogen from organic acids. These apparently simple pathways have yet to be clearly established. Previously, an insertion muta-

tion in *cycA*, encoding the type-1 tetraheme cytochrome c_3 , was constructed in *Desulfovibrio* G20 (Rapp-Giles et al. 2000). The growth rate of the *CycA*⁻ mutant was similar to that of the wild type during lactate-supported sulfate respiration or pyruvate fermentation. However, *CycA*⁻ appeared to be unable to respire sulfate with pyruvate as the electron donor and it was unable to dismutate fumarate. Interestingly, *CycA*⁻ growing with lactate/sulfate generated more hydrogen gas than G20, suggesting that electron flow was rerouted.

To determine any changes in gene expression that might account for these differences, both *Desulfovibrio* G20 and the *CycA* mutant were grown in pH-controlled fermentors in define medium by lactate-supported sulfate respiration or by pyruvate fermentation. RNA was prepared and microarray analysis performed to determine differential gene expression between exponential and stationary phase G20 cells grown by respiration or by fermentation and then determine any differences in gene expression in the *CycA* mutant grown under comparable conditions. The transcriptomic results demonstrated gene expression profile changes of G20 in response to carbon and energy depletion upon entering stationary phase as well as those occurring in response to pyruvate fermentation versus lactate respiration with sulfate. Hierarchical cluster analysis of transcriptomic profiles for G20 and the *CycA* mutant revealed 12 distinct clusters. Also, the arrays were clustered by growth phases. As would be predicted, in both strains genes involved in translation and transcription were expressed at high levels during the log phase. In addition, some genes for biosynthesis of amino acids, co-enzymes, lipid and carbohydrates were up-regulated during logarithmic growth. During the stationary phase, large numbers of genes encoding proteins for energy generation were up-regulated, including hydrogenases, dehydrogenases and ATP synthase. Also, once the electron donors were exhausted, genes involved in the flagella biosynthesis and in stress responses were increased in expression.

Interestingly, two well defined clusters with 14 ORFs each showed significant transcript differences between the two strains and the two media. In the *CycA* mutant, one of the clusters that included a large operon encoding the enzymes for conversion of pyruvate to succinate was down regulated. The second cluster showed decreased expression in pyruvate medium compared to lactate sulfate medium in both the log and stationary phases for G20 as well as for the *CycA* mutant. That cluster encodes two operons, one containing the type II c_3 transmembrane complex genes and a second with other energy related protein genes.

Metabolic profiles were obtained for G20 and *CycA* cultures in respiring and fermenting cultures. Comparing the metabolites from the *CycA* mutant to those from G20 in lactate/sulfate medium, higher concentrations of glycerol-1-phosphate, stearate, citramalate, aspartate, glycine and nucleotides were detected regardless of growth phase. Lower concentrations of lactate and succinate were found in the mutant. In contrast, in pyruvate fermenting cultures, the concentrations of glutamate, lactate, aspartate, N-acetylaspartate, glycine, and 5-oxo-proline were higher in the *CycA*

mutant than in G20 in log phase and opposite in stationary phase. The reverse was found for concentrations of succinate and trehalose in the mutant, lower than those from G20 in log phase and higher in stationary. These intracellular changes may also reflect overall redirection of substrates and electron flow in the two strains.

Reference

1. Rapp-Giles, B.J., L. Casalot, R.S. English, J.A. Ringbauer, Jr., A. Dolla, and J.D. Wall. 2000. Cytochrome c_3 mutants of *Desulfovibrio desulfuricans*. Appl. Environ. Microbiol. 66:671-677.

Student Presentation

127

Genome Resequencing of Thermoacidophilic Archaeal Carbon Flux Mutants

Y. Maezato* and P. Blum (pblum1@unl.edu)

Beadle Center for Genetics, University of Nebraska, Lincoln

Project Goals: The project (DE-FG36-08G088055) supports research on the ability of extremely thermoacidophilic microbes and their proteins to accelerate lignocellulose processing from switchgrass and to maximize sugar release. Delineation of carbon metabolic pathways in *Sulfolobus solfataricus* (*Sso*) is an essential step towards engineering traits associated with lignocellulosic bioprocessing. In *Sso*, carbon catabolism varies between two phenotypic states. In *car* mutants, growth on hexoses is blocked while pentose catabolism remains unaffected. In light of this characteristic, identification of *car* was undertaken.

The project (DE-FG36-08G088055) supports research on the ability of extremely thermoacidophilic microbes and their proteins to accelerate lignocellulose processing from switchgrass and to maximize sugar release. Delineation of carbon metabolic pathways in *Sulfolobus solfataricus* (*Sso*) is an essential step towards engineering traits associated with lignocellulosic bioprocessing. In *Sso*, carbon catabolism varies between two phenotypic states. In *car* mutants, growth on hexoses is blocked while pentose catabolism remains unaffected. In light of this characteristic, identification of *car* was undertaken. To genetically map *car* and its regulated targets, a genomic large insert library (BACs) was sorted by selection for clones that restored hexose metabolism in the *car* mutant. Shotgun subclone libraries led to the identification of a glycolytic gene, *kdgK*, and a chromatin modification gene, *hdac-1*, as components of the *car* mutant glycolytic defect. Genome resequencing was conducted to further characterize the genetic basis for *car*. Nearly 50 differences were identified distinguishing genomes of the *Sso* wild type and *car* mutant derivative. These included the CRISPR associated RAMP module Cas gene *cmr2*. Additional three-way whole genome comparisons further narrowed strain differences. Detailed genotyping combined with genomic reconstruction is underway to further assess these whole genome sequencing results.

128

Metabolic Fluxes: Quantifying Competition Between Nitrate and Proton Reduction during Fermentative NAD (P)H Formation in Real-Time

Gennady Ananyev,* Nicholas J. Skizim, and G. Charles Dismukes (dismukes@rci.rutgers.edu)

Waksman Institute of Microbiology and Dept. of Chemistry and Chemical Biology, Rutgers University, Piscataway, N.J.

Project Goals (Abstracts 128-129): The goal of this project is the design and utilization of tools which will allow a quantitative understanding of the mechanism and kinetics of hydrogen production in cyanobacteria.

The ability to continuously monitor metabolite concentrations *in vivo* as they undergo transient changes in response to environmental perturbations is among the most sought goals of metabolomics. Here we provide direct measures of the quantitative flux of intracellular reductant and its temporal dynamics between competing pathways. We have developed a fluorometric method for monitoring the concentration of intracellular reduced pyridine nucleotides, NAD (P)H, and combined it with a real-time electrochemical assay of dissolved H₂ concentration for simultaneous measurements on whole cells. Here we describe an application to quantitative kinetics in the cyanobacterium *Arthrospira maxima*, the most prolific fermentative H₂ producer in this group. *A. maxima* produces H₂ via a bidirectional NiFe-hydrogenase, the sole H₂-metabolizing enzyme in this organism. We demonstrate that the two temporal phases of H₂ production induced following the onset of anaerobiosis arise from distinct metabolic processes responsible for (a) anabolic production of NADPH (Phase-1) and (b) catabolic production of NADH (Phase-2). Phase-1 starts within minutes of anaerobiosis and may extend up to 2-3 h. Phase-1 H₂ is shown to correlate with the residual pool of photosynthetically produced NADPH; it decreases with increasing aerobic dark time prior to anaerobiosis. The second phase starts within 3-20 h and follows the rise of intracellular NAD (P)H with a short lag time ($\Delta t = 24$ min) indicative of the time to achieve a redox poise ($[\text{NAD (P)H}]/[\text{NAD (P)}^+] > 100$) sufficient for H₂ formation. The major Phase-2 H₂ is produced by autofermentation of carbohydrate reserves. The yield of Phase-2 H₂ is shown to be inversely related to the concentration of NO₃⁻ in the medium. A positive linear relationship is observed between the NO₃⁻ concentration and the delay time for onset of Phase-2 H₂ production over more than two decades change in concentration. Experiments carried out at various extracellular concentrations of inhibitors of nitrate reductase (N₃⁻, CN⁻) provide solid evidence for a direct competition for consumption of cellular reductant by hydrogenase and nitrate reductase (*nar*). This evidence provides a firm basis for proposed metabolic engineering of pathways.

Supported by DOE-GTL DE-FG02-07ER64488 and AFOSR-MURI FA9550-05-1-0365.

129

Real-Time Co-Detection of Dissolved H₂ and Intracellular NAD (P)H Concentrations in Microbes

Gennady Ananyev, Nicholas J. Skizim,* and G. Charles Dismukes (dismukes@rci.rutgers.edu)

Waksman Institute of Microbiology and Dept. of Chemistry and Chemical Biology, Rutgers University, Piscataway, N.J.

Project Goals: See goals for abstract 128.

In an effort to probe the kinetics and quantify the concentration of hydrogen produced by living microorganisms, here we describe a home-built combination cell for simultaneous real-time detection of dissolved H₂ and reduced pyridine nucleotide levels, [NAD (P)H]. Electrochemical H₂ detection (LOD: 2 nC = 1 nM H₂) from live cells with sample volumes as small as 5 μ l is accomplished with a custom made Clark-type cell (or reverse fuel cell) comprised of a membrane covered Pt/Ir electrode poised at a bias of +220 mV. Total reduced pyridine nucleotide = [NADPH] + [NADH] is assayed by selectively exciting NAD⁺ and NADP⁺ through an ultraviolet light emitting diode (365 nm) operating in pulse mode, and subsequently measuring fluorescence emission at 470 \pm 20 nm with a photodiode. Calibration curves with standards reveal linear responses for both H₂ and NAD (P)H detection using inactivated microbes as background, from which quantitative concentrations could be assessed. Using these tools we have been able to observe for the first time multiple temporal phases of H₂ production in anaerobically poised cyanobacteria which possess a [NiFe]-hydrogenase, as well as those containing a nitrogenase. We show that two main phases of hydrogen production correlate with the availability of residual NADPH produced via photosynthesis prior to the onset of anaerobiosis (phase 1), and NADH produced by anaerobic autofermentation of glycogen reserves (phase 2) in strains containing a [NiFe]-hydrogenase as their only hydrogen metabolizing enzyme. Application of these tools to N₂ fixing cyanobacteria has similarly revealed at least three kinetic phases of dark hydrogen production coupled to fermentative metabolism. Experiments are planned on two diazotrophic *Synechococcus* thermophiles which lack all types of hydrogenase. Together these data shall provide a basis for discriminating between and quantifying the H₂ evolution pathways involving hydrogenase and nitrogenase.

Supported by DOE-GTL DE-FG02-07ER64488 and AFOSR-MURI FA9550-05-1-0365

130

Systems Level Approaches to Understanding and Manipulating Heterocyst Differentiation in *Nostoc punctiforme*: Sites of Hydrogenase and Nitrogenase Synthesis and Activity

Harry Christman,¹ Elsie Campbell,¹ Douglas Rissler,¹ Brett Phinney,² Wan-Ling Chiu,³ and **John C. Meeks**^{1*} (jcmeeks@ucdavis.edu)

¹Dept. of Microbiology and ²Genome Center, University of California, Davis; and ³Dept. of Biology, Virginia Commonwealth University, Richmond

Project Goals: Heterocyst-forming cyanobacteria, such as *Nostoc punctiforme*, are applicable for cost-effective photo-biohydrogen production, providing the frequency of heterocysts in the filaments can be increased by genetic manipulation and the metabolic end product, H₂, is uncoupled from growth.

We hypothesize that if the frequency of heterocysts, sites of nitrogen fixation and hydrogen evolution, can be increased approximately 3-fold above the less than 10% normally found in filaments, then heterocyst-forming cyanobacteria would be applicable for cost effective photo-biohydrogen production. In *Nostoc punctiforme*, the heterocyst frequency of 8% in the free-living state is increased to 30-35% when in symbiotic association with terrestrial plants, such as the hornwort *Anthoceros* spp. and the angiosperm *Gunnera* spp. We are applying transcriptomic and proteomic analyses with wild-type and mutant strains to identify the regulatory circuits of free-living heterocyst differentiation and how those circuits have been co-opted during symbiotic growth.

Transcriptomics. HetR and HetF are two positive regulatory elements in the differentiation of heterocysts. Mutations in either structural gene in *N. punctiforme* lead to the inability to differentiate heterocysts in free-living cultures or to support N₂-dependent growth of the hornwort plant partner, although the mutants do infect the plant. Conversely, over expression of the genes from a multicopy plasmid in trans, results in the differentiation of multiple heterocysts that are localized in distinct clusters with a nonrandom spacing pattern in the filaments. The heterocyst frequency can reach 40% of the cells, but the heterocysts do not fix nitrogen in support of vegetative cell growth. A question is whether the gene products operate at the same pathway in the regulatory cascade of heterocyst differentiation. Time course (0.5 to 24 h) DNA microarray experiments, using a Nimblegen platform, with the data analyzed in the R statistical environment and subsequently clustered using the Genesis program, yielded two interesting results. A total of 124 genes were up-regulated in both mutants in response to combined nitrogen limitation. Only one known up-stream gene for heterocyst differentiation gene (*nrrA*) was up-regulated. The up-regulated genes include those for assimilation of the alternative inorganic nitrogen sources nitrate and urea. Only one amino acid transport gene was

up-regulated, implying *N. punctiforme* does not search for organic nitrogen sources during nitrogen starvation. An additional common cluster of 330 genes were down-regulated in the two mutants and these included many encoding proteins of central metabolism, protein synthesis and photosynthetic energy metabolism. Approximately 53% of the up- and down-regulated transcripts encode proteins of unassigned function. The second result of interest is that 130 and 775 genes were uniquely up-regulated, and 29 and 1063 were uniquely down-regulated in the *hetF* and *hetR* mutants, respectively. These results imply that HetF and HetR have physiological roles in addition to heterocyst differentiation, which is consistent with the presence of the genes in filamentous cyanobacteria that do not differentiate heterocysts. We are now examining the transcriptional profiles of wild type strains over expressing HetR and HetF in order to identify genetic targets and any differences in the targets in the two constructs for comparisons to the loss of function mutant data. We have also initiated analysis of a unique pattern mutant in which a high heterocyst frequency (~ 35%) is manifest as multiple singular heterocysts with a 3-4 vegetative cell spacing between heterocysts. This is the pattern we observe in the symbiotic growth state. However, this mutant also does not grow with N₂ as the sole nitrogen source.

Proteomics. We have completed a single Mudpit-based MS/MS run of the total proteome of a N₂-grown culture of *N. punctiforme*. The cell extract was processed into three fractions; an initial 14,000 x g pellet of membrane proteins, a 150,000 x g supernatant of soluble proteins and a 150,000 x g pellet of primarily carboxysomes and phycobilisomes. The fractions were sub-fractionated by PAGE, sliced sections digested with trypsin and the eluted products further fractionated by 2 dimensional LC followed by electrospray injection. This 600 member proteome is comparable to one we previously defined of the supernatant proteins of an ammonium-grown culture, except for an enrichment in membrane proteins, plus proteins associated with heterocysts and nitrogenase function. Notable is the constitutive presence of DNA photolyases and enzymes for metabolism of reactive oxygen species, stress factors to which cyanobacteria are constantly exposed. The value of this proteome will be enhanced by comparison to a N₂ plus fructose grown mixotrophic culture now in progress. In the symbiotic growth state *N. punctiforme* grows as a photomixotroph in *Anthoceros* and most likely as a heterotroph in *Gunnera*. Thus, these proteomes are essential for comparisons to symbiotic growth with the accompanying high heterocyst frequency and high rates of nitrogenase activity.

131

Student Presentation

Contributors to Light- and Feedstock-Powered Hydrogen Production in *Rhodobacter sphaeroides*

Melanie A. Spero^{1*} (spero@wisc.edu), Ugur Sohmen^{2*} (sohmen@wisc.edu), Wayne S. Kontur,¹ L. Safak Yilmaz,² Daniel R. Noguera² (noguera@engr.wisc.edu), and Timothy J. Donohue¹ (tdonohue@bact.wisc.edu)

¹Dept. of Bacteriology; and ²Dept. of Civil and Environmental Engineering, University of Wisconsin, Madison

Project Goals: Our research seeks to generate a quantitative understanding of light- and feedstock-powered hydrogen (H₂) production in the photosynthetic bacterium *Rhodobacter sphaeroides*. We will determine the source (s) of H₂ production in *R. sphaeroides* and determine the effect on fuel production of eliminating or minimizing processes predicted to siphon reducing power from H₂ production. This will be accomplished by deleting genes in pathways predicted to impact H₂ production and assaying for the amount of reductant that is partitioned to H₂, biomass, as well as other electron sinks. Mutations tested to date include those in genes encoding enzymes predicted to generate H₂ (nitrogenase), detract from H₂ production (hydrogenase), and in pathways predicted to compete for reducing power that could be utilized for H₂ production (synthesis of polyhydroxybutyrate, carbon dioxide fixation). This work also compares H₂ production in log-phase batch cultures to late stationary phase cultures, where the amount of reducing power shuttled into biomass production is minimized. Each mutant will be assayed for H₂ production under active or non-growing conditions on a variety of carbon sources to determine the impact of these cellular processes on H₂ production.

These studies are performed in the α -proteobacterium *R. sphaeroides*, the most-studied photosynthetic bacterium, since it is known to produce relatively large amounts of H₂ under photoheterotrophic growth conditions. A working model of the processes predicted to impact H₂ production was developed based on prior knowledge of metabolic pathways, the genome sequence of this bacterium, and global gene expression data (see posters by Yilmaz et al., and Kontur et al.).

H₂ production in *R. sphaeroides* is proposed to be mainly or completely associated with the nitrogenase enzyme. To determine the contribution of nitrogenase to H₂ production, a mutant strain of *R. sphaeroides* was generated that lacks the structural genes of nitrogenase (*nifHDK*). The growth properties of the Δ Nif mutant support the hypothesis that nitrogenase-mediated H₂ production is an important electron sink with some, but not all, carbon feedstocks. Experiments are ongoing to use the Δ Nif mutant to determine if other *R. sphaeroides* enzymes can contribute to H₂ production under defined conditions.

One enzyme predicted to detract from H₂ production is uptake hydrogenase (Hup), which can oxidize H₂ into protons and electrons. A mutant strain of *R. sphaeroides* lacking the Hup structural genes (*hupSL*) is being used to determine the impact of Hup activity on H₂ production, especially since global gene expression data predict a wide range of Hup activities present when cells grow in media containing different carbon sources as feedstocks.

We are also testing the effects of other electron sinks on H₂ production, by blocking pathways known or predicted to compete for reducing power. For example, mutant strains with gene deletions in polyhydroxybutyrate synthesis or carbon dioxide sequestration (via the Calvin cycle) are being analyzed to determine effects on H₂ production. Each of these mutants is being assayed for H₂ production on various carbon sources in order to evaluate how altering the flow of reducing power into other electron sinks can alter the distribution of reducing power into various products.

Because biomass is as a significant electron sink in exponentially growing cells, the distribution of electrons into different products is also being analyzed in stationary phase cultures, when electron flow towards biomass production is predicted to be minimized. Preliminary results indicate that partitioning of reducing power to H₂ production varies between growing and stationary phase cultures on some carbon feedstocks. We will report on experiments with wild type and mutant strains that seek to monitor electron partitioning during stationary phase to better understand how feedstock supply impacts the flow of electrons to H₂ and other electron sinks.

132

SurR Regulates Hydrogen Production in *Pyrococcus furiosus* by a Sulfur-Dependent Redox Switch

Gina L. Lipscomb,^{1*} Hua Yang,^{1,4} Annette Keese,² Gerrit J. Schut,¹ B.C. Wang,¹ Michael Thomm,² Janet Westpheling,³ Michael W.W. Adams,¹ and Robert A. Scott¹ (rscott@uga.edu)

¹Dept. of Biochemistry and Molecular Biology, University of Georgia, Athens; ²Dept. of Microbiology, University of Regensburg, Germany; ³Dept. of Genetics, University of Georgia, Athens; and ⁴Centers for Disease Control and Prevention, Atlanta, Georgia

Project Goals: The goal of SAPHyRe (Systems Approach to Probing Hydrogen Regulation) is to develop a detailed systems-level description of the regulatory and metabolic networks controlling hydrogen production in the hyperthermophilic archaeon *Pyrococcus furiosus* (Pf). Pf will be used as the model organism to investigate its response to various environmental conditions relevant to all hydrogen-producing microorganisms, such as carbon and nitrogen sources, metal availability, and oxidative and reductive stresses. The outcome of this project will

serve two purposes: 1) it will bring us one step closer to utilizing *P. furiosus* in development of H₂ as an alternative energy source and 2) it will serve as a model methodology for investigating the regulatory pathways of hydrogen production in other organisms.

We present structural and biochemical evidence for a redox switch in the archaeal transcriptional regulator SurR of *Pyrococcus furiosus*, a hyperthermophilic anaerobe. *P. furiosus* produces H₂ during fermentation, but undergoes a metabolic shift to produce H₂S when elemental sulfur (S⁰) becomes available. Changes in gene expression occur within minutes of S⁰ addition, and the majority of these S⁰-responsive genes are regulatory targets of SurR, a key regulator involved in primary S⁰ response. SurR was shown *in vitro* to have dual functionality, activating transcription of some of these genes, notably the hydrogenase operons, and repressing others, including a gene encoding sulfur reductase. This work demonstrates that the activity of SurR is modulated by cysteine residues in a CxxC motif that constitute a disulfide switch. Oxidation of the switch with S⁰ inhibits sequence-specific DNA binding by SurR, leading to *deactivation* of genes related to H₂ production and *derepression* of genes involved in S⁰ metabolism.

133

Pathways and Regulatory Network of Hydrogen Production from Cellulose by *Clostridium thermocellum*

Jonathan Millen^{1*} (jonathan_millen@hotmail.com), Ji Hye Jo,² Shiv Thammannagowda,² Grant Pennington,² Matthew Heckman,¹ Pin-Ching Maness,² Mary Lipton,³ **J.H. David Wu**¹ (davidwu@che.rochester.edu)

¹University of Rochester, N.Y.; ²National Renewable Energy Laboratory, Golden, Colo.; and ³Pacific Northwest National Laboratory, Richland, Wash.

Project Goals: The overall objective of this research is to understand H₂ metabolic pathways in *Clostridium thermocellum* and the underlying regulatory network at the molecular and systems levels. We hypothesize that: multiple hydrogenases work concertedly to contribute to hydrogen evolution; transcription factors control hydrogenase expression at the transcription level; and hydrogen metabolism is coupled to cellulolysis and fermentative pathways in this bacterium. We are determining the hydrogenase expression and metabolic network nodes on cells subjected to different culture conditions and metabolic pathway inhibitors, to probe differential expression of the various hydrogenases and their interrelationship with other cellular metabolic pathways. We are also mapping connections in the transcription factor network controlling linked metabolic pathways. Finally, we will purify FeFe-hydrogenases from its native producer and heterologously expressed *E. coli* to determine their subunit compositions, endogenous redox partners, and the direc-

tion of reaction (hydrogen production vs. uptake) to shed light on their roles in hydrogen metabolism.

Clostridium thermocellum, a thermophilic, ethanogenic, and cellulolytic anaerobe, produces a complex extracellular cellulolytic organelle called the cellulosome. The cellulosome contains various depolymerizing enzymes that are arrayed on a protein scaffold and effectively degrades complex cellulosic substrates. During cellulose fermentation, the bacterium evolves hydrogen at a high rate. Analysis of its genome sequence reveals the existence of at least three putative hydrogenase genes (CtHydA1, CtHydA2 and CtHydA3) central to hydrogen metabolism. Furthermore, at least 20 genes are potentially related to hydrogen metabolism. The bacterium is thus remarkably versatile in employing various enzymes, some of which are potentially novel, for hydrogen metabolism, indicating the significance of this biological process. Yet little is known concerning the pathway for hydrogen production and the underlying regulatory mechanism/network that control these hydrogenase and the related genes as well as cellulolytic process and other metabolic pathways in the organism.

Due to the novelty of CtHydA3, it was selected for the initial study. Genes encoding CtHydA3, a ferredoxin-like protein (Ct_3004), and three FeFe-hydrogenase maturation proteins (CtHydE, CtHydF, and CtHydG) have been cloned into three plasmids and co-transformed into *E. coli* strain Rosetta (DE3) and BL21 (DE3) for heterologous expression. In addition, a 6X His-tag sequence was fused to either the C- or N-termini of CtHydA3. Protein immunoblots confirmed the expression of the C-terminus His-tagged CtHydA3 (73 kDa band) in *E. coli* Rosetta (DE3), but not in *E. coli* BL21 (DE3), likely due to differences in codon usage between *C. thermocellum* and *E. coli*. The expression is further corroborated by a five-fold increase in *in vitro* hydrogenase activity in the soluble cell extract of the recombinant Rosetta strain, mediated by reduced methyl viologen. However, no difference in *in vivo* hydrogen production was detected in the recombinant Rosetta strain, suggesting an inability of the recombinant hydrogenase to contribute to the host's hydrogen metabolism. The C-terminal His-tagged protein failed to bind to a TALON metal affinity column to facilitate its purification. Work is underway to express the N-terminal tagged His-CtHydA3 and explore its affinity purification and characterization.

To identify transcription factors controlling metabolic pathways, we developed an affinity purification method by immobilizing promoter DNA sequences to a solid support. DNA-binding proteins from the *C. thermocellum* cell lysate, obtained by growing on cellobiose or crystalline cellulose and eluted from the affinity columns, were identified by the MALDI-TOF or LC-MS-MS techniques. Several transcription factor candidates were identified. In a reversed approach, we expressed in *E. coli* various putative DNA-binding proteins found in the genome of *C. thermocellum*. Their target DNA-binding sites will be screened by using a DNA microarray we designed and confirmed by EMSA (electrophoretic mobility shift assay).

The studies will provide important insights into the pathway and regulatory mechanism/network controlling hydrogen metabolism and cellulolysis as well as other pertinent metabolic pathways in this very intriguing cellulolytic and thermophilic bacterium, which catalyzes the rate-limiting cellulose-degradation reaction in a single-step process of biomass conversion (or CBP, Consolidated Bioprocessing). Detailed understanding of the pathway and regulatory mechanism/network will ultimately provide rationales for engineering, alternating, or deregulating the organism for biomass conversion to liquid and hydrogen fuels.

134

Filling Knowledge Gaps in Biological Networks: Integrated Global Approaches to Understand Biofuel Metabolism in *Chlamydomonas reinhardtii*

Venkataramanan Subramanian,^{1,2} Alexandra Dubini,¹ Claudia Catalanotti,³ Wenqiang Yang,³ Christopher H. Chang,¹ Ambarish Nag,¹ David M. Alber,¹ Peter A. Graf,¹ Glenn Murray,² Kwiseon Kim,¹ Arthur R. Grossman,³ Matthew C. Posewitz,² and **Michael Seibert**^{1*} (mike.seibert@nrel.gov)

¹National Renewable Energy Laboratory, Golden, Colo.; ²Colorado School of Mines, Golden, Colo.; and ³Carnegie Institution of Washington, Stanford, Calif.

Project Goals: Development of photobiological, biofuels-production processes, a key component of DOE's renewable energy mission, would be accelerated by improved understanding of cellular metabolism and its regulation. Currently, a lack of comprehensive experimental data hinders the development of reliable metabolic models that have robust predictive capabilities. Therefore, we are employing high-performance computing to (1) estimate parameters that delimit the space of stable solutions for experimentally constrained metabolic models, (2) explore network capabilities in silico, and (3) integrate experimental, systems biology data to verify and refine metabolic models. Biochemical reactions are modeled with the fundamental independent variables being enzyme concentrations and Michaelis-Menten parameters. Through iterative model building and an understanding of cellular metabolism obtained from high throughput "omics" data, we are constructing metabolic models that link individual enzyme reactions and the activities of specific metabolic pathways with the production of biofuels in the green alga, *Chlamydomonas reinhardtii*. High-performance simulation and optimization will predict metabolic outputs based on the kinetic parameters governing individual reactions, identify metabolic limitations, and predict specific manipulations that are likely to improve biofuel outputs. The work is envisioned as an important contribution toward strengthening our knowledge of energy-related biosystems.

The goal of this project, jointly funded by the DOE Computational Biology and SciDAC Programs, is to develop a means to globally map, *in silico*, all biological pathways in *Chlamydomonas reinhardtii* (*Chlamydomonas* throughout) that can impact the production of H₂ and other biofuels. *Chlamydomonas* is the first alga for which a fully sequenced genome was available (thanks to the JGI), and this organism has recently emerged as the prototype for investigating the regulation of basic metabolism, particularly fermentative processes. *Chlamydomonas* has a complex anaerobic metabolic network that can be induced in the dark and produces H₂ along with other fermentation products such as formate, acetate, ethanol, and CO₂. Previous studies have focused on determining genes (by microarray and RT-PCR analysis) that were differentially regulated as the result of shifting cultures of the CC-425 parental strain from aerobic growth to dark anaerobiosis [1]. Anoxia led to differential expression of genes involved in fermentation metabolism, and specifically caused accumulation of transcripts encoding pyruvate formate lyase (*PFL1*) and pyruvate:ferredoxin oxidoreductase (*PFR1*). Moreover, *Chlamydomonas* synthesized the metabolites formate, acetate, and ethanol in the ratio 1:1:0.5. Recent experiments have shown that in a strain lacking proteins involved in assembling an active hydrogenase enzyme (*hydEF* mutant), the metabolite ratio shifted to 2:1:1, and succinate accumulated instead of H₂. Interestingly, levels of transcripts encoding proteins involved in fermentation also changed in the mutant relative to parental cells following the imposition of anoxic conditions; the *PFL1* mRNA accumulated to a higher level and the *PFR1* mRNA to a lower level in the mutant than in parental cells [2]. These results have allowed us to generate a physiological model that highlights the flexibility of fermentation metabolism and H₂ production in *Chlamydomonas*. Additional mutants (e.g., ΔPFL , ΔFMR , $\Delta HYDA2$, $\Delta HYDA2:\Delta PFL$, $\Delta HYDA2:\Delta MME4$) have been identified this year using a HTP screening procedure, and the analyses of these mutants are enabling us to further refine our knowledge of fermentation metabolism and its regulation in *Chlamydomonas*. Furthermore, transcript, protein and metabolite analyses of various single and double mutants (e.g., $\Delta HYDA1$, $\Delta ADH1$, $\Delta PDC3$, $\Delta HYDA1:\Delta HYDA2$, $\Delta PFL1:\Delta ADH1$) are already underway.

To integrate the results described above with various experimental results discussed below, we have developed metabolic models which, when simulated using our recently developed software toolkit, enable us to make informed decisions about the best ways to link experimentally measurable parameters to the biochemical reactions and their regulation in *Chlamydomonas*. The multi-compartment model includes carbon metabolism (glycolysis, tricarboxylic acid cycle, starch metabolism, Calvin-Benson Cycle), oxidative phosphorylation, and fermentative metabolism. Steady-state kinetic relationships are expressed as thermodynamically consistent ordinary differential equations. Merging of component pathway models, incorporation of ionization reactions, transformation to C++, compilation of the data, and linking the information to high-performance executable programs is largely automated using our *High-Performance Systems Biology Toolkit* (HiPer SBTK) [3]. The programs being used

are for parameter sampling, data fitting, and local or global optimization within a defined space of kinetic parameters and/or enzyme concentrations. Job configuration is possible through an auxiliary graphical interface or by direct editing of simple text. The code has been modularized, permitting facile incorporation of new techniques for high-dimensional sampling, optimization, and model integration. Key future developments will include investigation of model phase space with respect to initial conditions, and model expansion to include the photosynthetic light reactions and the effects of changes in light intensity on cellular metabolism.

Our current biological research, using global transcriptomic- and proteomics-based approaches, is supporting extensive computational analyses; this research is focused on elucidating the regulation of algal fermentative pathways and identifying various pathways that either directly or indirectly impact H_2 and biofuels production. We recently used 2D-differential gel electrophoresis (DIGE) and shotgun mass spectrometry to identify, quantify and compare proteins present under anoxic, H_2 -producing conditions [dark and sulfur-deprived] to those present under oxic, non- H_2 -producing conditions. Preliminary DIGE results indicate that the levels of 189 proteins are similar under oxic and anoxic conditions, 41 proteins are higher during anoxic growth and 69 proteins are higher during oxic growth. Furthermore, mass-spectrometry-based, shotgun-proteomic experiments (using the LTQ-Orbitrap system) with whole cell protein extracts detected 1485 proteins under anoxic conditions, 853 proteins under sulfur-deprivation conditions, and 1664 proteins under oxic conditions; analyses of this dataset is underway. Finer resolution will be achieved in the future by proteomic dissection of specific subcellular compartments. In-depth comparative analyses of the specific proteins identified under the various conditions will be discussed.

In summary, high-throughput 'omics' techniques are being used to input transcript, protein, and metabolism knowledge (from both parental and mutant strains) into computational models that explore metabolism (and its flexibility) in the green alga, *Chlamydomonas*. The coupling of experimental results, mutant analyses, and *in silico* modeling is expected to improve our understanding of the complexity of *Chlamydomonas* metabolic networks, including ways in which the cells adjust to changing environmental conditions (i.e., by modulating metabolite fluxes), and how blocking steps in specific metabolic pathways alter metabolite flow. This information as a whole will suggest critical strategies for engineering *Chlamydomonas* metabolism for more efficient production of H_2 and/or other algal biofuels.

References

1. Mus, F., Dubini, A., Seibert, M., Posewitz M.C., and Grossman A.R. (2007) "Anaerobic acclimation in *Chlamydomonas reinhardtii*: Anoxic gene expression, hydrogenase induction and metabolic pathways", *J. Biol. Chem.* 282 (35), 25475-25486.
2. Dubini, A., Mus, F., Seibert, M., Grossman, A.R., and Posewitz M.C., (2009) "Flexibility in anaerobic metabolism as revealed in a mutant of *Chlamydomonas reinhardtii* lacking hydrogenase activity", *J. Biol. Chem.* 284, 7201-7213.
3. Chang, C.H., Graf, P., Alber, D.M., Kim, K., Murray, G., Posewitz, M., and Seibert M., (2008) "Photons, Photosynthesis, and High-Performance Computing: Challenges, Progress, and Promise of Modeling Metabolism in Green Algae", *J. Phys. Conf. Ser.* 125: 012048.

135

Development of Biologically-Based Assays to Study Rate-Limiting Factors in Algal Hydrogen Photoproduction

Alexandra Dubini,¹ Matthew Wecker,² and Maria L. Ghirardi^{1*} (maria_ghirardi@nrel.gov)

¹National Renewable Energy Laboratory, Golden, Colo. and ²GeneBiologics, LLC, Boulder, Colo.

Project Goals: See below.

Photobiological H_2 production from water is a clean, non-polluting and renewable technology. Although the potential light conversion efficiency to H_2 by biological organisms is theoretically high (about 10%), the system is currently limited by biochemical and engineering constraints. These limitations include (but are not restricted to) the extreme O_2 sensitivity of the biological hydrogenases and the low availability of reductants to the hydrogenase due to the existence of competing metabolic pathways.

To address the O_2 sensitivity issue, our research is developing a new, biologically-based assay to screen large microbial populations for improved H_2 -production properties. This novel assay is based on the H_2 -sensing system of the photosynthetic bacteria *Rhodospirillum rubrum*. The H_2 -sensing system is being optimized as a green fluorescence protein-based reporter-assay for heterologous hydrogen production by the bacteria. It will then be used to screen for O_2 tolerant [FeFe]-hydrogenases generated through directed-evolution techniques. The hydrogenases of *Clostridium acetobutylicum*, *Chlamydomonas reinhardtii* and *Bacteroides thetaiotaomicron*, along with their respective assembly proteins have been introduced into broad host range vectors and are being shuttled into *R. rubrum*.

To address the issue of competitive metabolic pathways with H_2 production, we have started using a yeast two hybrid. A complex library of 10^7 preys has been constructed and 8 different baits are being used to screen the library. Those baits include two hydrogenases and the 6 potential partner ferredoxins. We wish to deconvolute the hydrogenase interactors *in vivo* by fishing preys from the library and hopefully discover which of one the ferredoxins is the direct electron donor. We also aim at understanding the ferredoxin interaction network as they are key players in the hydrogenase pathway and are electron donor to many other competitive pathways. The screens are in progress and will help us having a better idea of the metabolic network involved in the hydrogen production pathway.

136

Transcriptome and DNA Methylome Analysis of Algae Using Ultra-High-Throughput Sequencing

David Casero Díaz-Cano,¹ Shawn Cokus,¹ Madeli Castruita,² Steven Karpowicz,² Sabeeha Merchant,² and Matteo Pellegrini¹ (matteop@mcd.ucla.edu)

¹Department of Molecular, Cell and Developmental Biology and ²Department of Chemistry and Biochemistry, University of California, Los Angeles

Project Goals: Develop algal biofuels.

High-throughput sequencing has undergone remarkable increases in efficiency over the past few years. Some of the most efficient machines are those produced by Illumina, which sequence tens of billions of bases per week. We are currently using these sequencers in to study the transcriptome of algae such as *Chlamydomonas reinhardtii*, a unicellular eukaryote in the plant lineage, has been exploited in the laboratory over the last 50 years as a model organism for the study of eukaryotic photosynthesis. The advent of massively parallel short read sequencing technology opens the door to (near) full coverage of the *Chlamydomonas* transcript map via deep sequencing of mRNAs. To evaluate the potential of Illumina's Solexa technology for a) generating a whole transcriptome for *Chlamydomonas*, b) identifying differentially expressed genes, and c) reconstructing gene models *de novo*, we analyzed RNAs isolated from a variety of conditions. We have verified that these libraries may be used to quantitatively estimate transcript fold changes in different conditions using existing gene models. We are also developing a new annotation pipeline using only the short read sequencing data, and have shown that these approaches allow us to accurately reconstruct gene models.

137

Cell-Free Synthetic Pathway Biotransformations (SyPaB) for Producing Hydrogen and Even Fixing Carbon Dioxide

Y.-H. Percival Zhang^{1,2*} (ypzhang@vt.edu), Jonathan Mielenz,^{2,3} and Mike Adams^{2,4}

¹Biological Systems Engineering Dept., Virginia Tech, Blacksburg; ²DOE BioEnergy Science Center (BESC), Oak Ridge, Tenn.; ³Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tenn.; and ⁴Dept. of Biochemistry and Molecular Biology, University of Georgia, Athens

Project Goals: Demonstrate a new direction for synthetic biology — cell free synthetic pathway biotransformation (SyPaB) Produce 12 moles of hydrogen from low-cost cellulosic materials for the first time Solve several challenges

for cell-free SyPaB Propose a novel pathway for CO₂ fixation by using electricity

Cell-free synthetic pathway biotransformations (SyPaB), a new direction of synthetic biology, is to implement complicated biochemical reaction network by assembling a number of purified enzymes and coenzymes.³ As compared to microbial fermentation, SyPaB has several advantages: (i) high product yields without synthesis of cell mass or formation of by-products; (ii) great engineering flexibility (i.e., easy assembly and control when the building blocks are available); (iii) high product titer; (iv) fast reaction rate; and (v) broad reaction condition. But SyPaB suffers from several obvious weaknesses, such as no ability for self-duplication, costly enzymes, enzyme deactivation, costly and labile coenzymes, and so on.

We have designed the non-natural pathways that completely oxidize starch or cellodextrins by using water as an oxidant for generation of **12 mol of hydrogen per mol of glucose equivalent**^{4,5} (Fig. 1). These catabolic pathways comprise substrate phosphorylation mediated by phosphorylases, pentose phosphate pathway, and hydrogenesis mediated by hydrogenase. Also, we have increased overall hydrogen production rate by 10 fold through pathway optimization, high temperature and high substrate concentrations.⁴ The above reaction is a unique entropy-driven chemical reaction, i.e., low-temperature heat energy is absorbed and is converted to chemical energy – hydrogen that we can utilize for the first time. Also, we propose to use renewable carbohydrate as a hydrogen carrier to solve hydrogen storage challenge (Fig. 2).⁶

The opinion that SyPaB is too costly for producing low-value biocommodities are mainly attributed to the lack of stable standardized building blocks (e.g., enzymes or their complexes), costly labile co-enzymes, and replenishment of enzymes and co-enzymes. The economical analyses clearly suggest that developments in stable enzymes or their complexes as standardized parts, efficient coenzyme recycling, and use of low-cost and more stable biomimetic coenzyme analogues, would result in much lower production costs than do microbial fermentations because the stabilized enzymes have more than three orders of magnitude higher weight-based total turn-over numbers than microbial biocatalysts, although extra costs for enzyme purification and stabilization are spent³ (Fig.3).

Fig. 4 clearly suggest that the ultimate hydrogen production costs would be as low as \$1.50 per kg, where carbohydrate accounts for 80% of the final product price². Developing thermostable enzymes with TTN_w of > 100,000 are easily reached based on our experiences.^{7,8} For example, we have obtained three recombinant thermophilic building blocks — #2 *Clostridium thermocellum* phosphoglucomutase⁷, #4 *Thermotoga maritima* 6-phosphogluconate dehydrogenase,⁸ and #11 *T. maritima* fructose bisphosphatase expressed in *E. coli*, all of which have TTN_w of > 200,000 at ~60 °C. The above results suggest that discovery and utilization of highly stable thermophilic enzymes from extremophiles that have known genomic sequences are highly operative. Recycling NAD with TTN of 1,000,000 has been reported and the

use of less costly biomimetic NAD will be more economically promising, as shown in Fig. 4.^{6,9}

In addition to high-yield generation of hydrogen from biomass sugars, we have designed a novel artificial photosynthesis pathway that can utilize electricity to fix CO₂ for producing ethanol and amylose (Fig. 5). When this process is implemented on large scales, it would solve several challenges for sustainability, such as CO₂ fixation, electricity storage, food production, transportation fuel production, water conservation or maintaining an ecosystem for space travel (the concept paper is under review for publication).

References

1. Y.-H.P. Zhang, *Microbe* 4 (2009) In press.
2. Y.-H.P. Zhang, J.-B. Sun, A.-P. Zeng, J.-J. Zhong, *Curr. Opin. Microbiol.* (2010) Invited/In preparation.
3. Y.-H.P. Zhang, *Biotechnol. Bioeng.* Accepted (2010) <http://dx.doi.org/10.1002/bit.22630>.
4. X. Ye, Y. Wang, R.C. Hopkins, M.W.W. Adams, B.R. Evans, J.R. Mielenz, Y.-H.P. Zhang, *ChemSusChem* 2 (2009) 149-152.
5. Y.-H.P. Zhang, B.R. Evans, J.R. Mielenz, R.C. Hopkins, M.W.W. Adams, *PLoS One* 2 (2007) e456.
6. Y.-H.P. Zhang, *Energy Environ. Sci.* 2 (2009) 272-282.
7. Y. Wang, Y.-H.P. Zhang, *J. Appl. Microbiol.* 108 (2010) 39-46.
8. Y. Wang, Y.-H.P. Zhang, *Microb. Cell Fact.* 8 (2009) 30.
9. J.D. Ryan, R.H. Fish, D.S. Clark, *ChemBioChem* 9 (2008) 2579-2582.

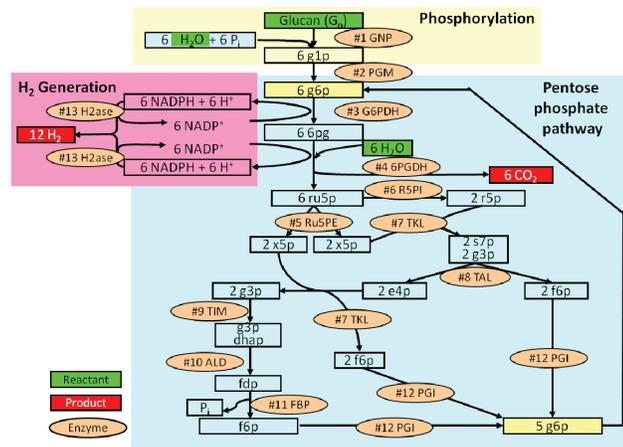


Fig. 1. The non-natural SyPaBs for high-yield hydrogen generation from starch or cellulosic materials [4, 5].

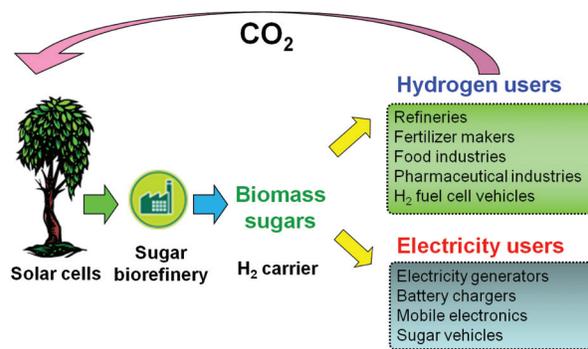


Fig. 2. Scheme of the hydrogen economy based on renewable carbohydrates [1].

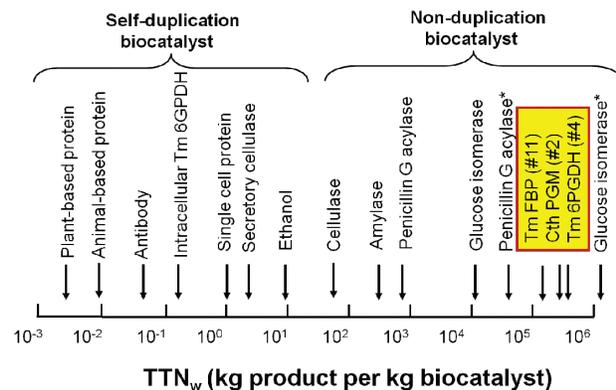
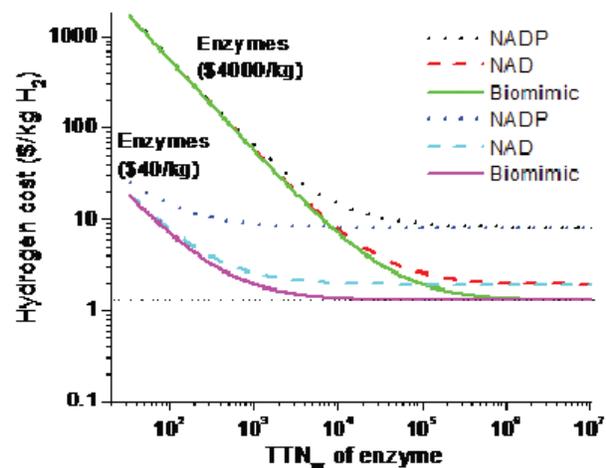


Fig. 3. Comparison of weight-based total turn-over number (TTNW) of self-duplication living biocatalysts and non-duplication enzymes [1, 3].



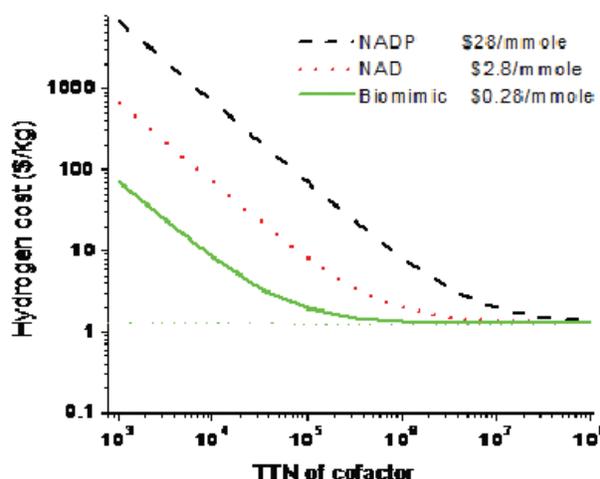


Fig. 4. The estimated hydrogen production costs in terms of turnover number of enzymes and co-enzymes. Carbohydrate prices are \$0.18 per kg [2].

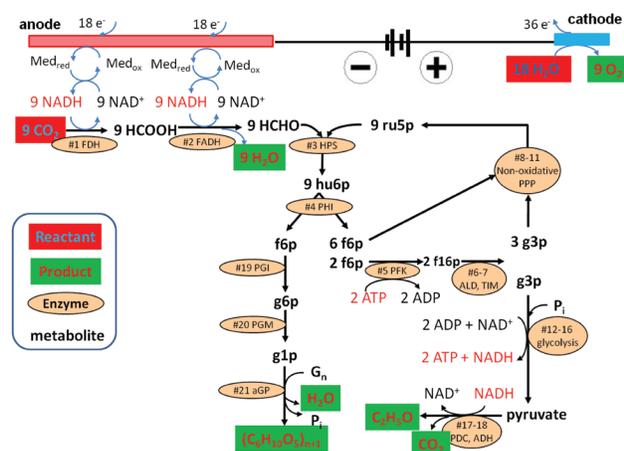


Fig. 5. Artificial photosynthesis that can fix CO₂ by using electricity for producing biofuels and food (under review).

submitted post-press

BioBricks Without Borders: Investigating a Multi-Host BioBrick Vector and Secretion of Cellular Products

Elisabeth Linton* and Charles D. Miller (charles.miller@engineering.usu.edu)

Dept. of Biological and Irrigation Engineering, Utah State University, Logan

Project Goals: This research was carried out as part of the 2009 iGEM competition. The overall goal of this project is to use synthetic biological engineering for improved expression and recovery of cellular products. More specific

objectives include the conversion of a multi-host vector into a BioBrick compatible format to facilitate expression of BioBricks in variety of organisms and the design BioBrick parts to encourage export of cellular compounds out of the cytoplasm.

The aim of the Utah State University iGEM project is to develop improved upstream and downstream processing strategies for manufacturing cellular products using the standardized BioBrick system. A BioBrick-compatible broad-host vector would facilitate exploitation of advantageous characteristics of various organisms beyond *E. coli*, such as the ability to photosynthetically assimilate carbon. Multi-host vectors were investigated to enable the use of BioBrick constructs in organisms like *Pseudomonas putida*, *Rhodobacter sphaeroides*, and *Synechocystis* PCC6803. For this portion of the project, vector pCPP33 was successfully converted to a BioBrick-compatible format. Following expression, product recovery poses a difficult and expensive challenge. Product purification commonly represents more than half of the total production expense. To counter this problem, secretion-promoting BioBrick devices were constructed through genetic fusion of signal peptides with protein-coding regions. Specifically, phasin protein was targeted for membrane translocation due its binding interaction with polyhydroxyalkanoates (PHAs), which are microbially-accumulated biodegradable plastics. Successful secretion of phasin protein holds potential to lead to an improved recovery mechanism for PHAs. The secretion of green fluorescent protein was studied in parallel due to its ease of detection. A genetic library of more than 50 BioBrick parts has been constructed to carry out this study. Current results indicate that many of these parts are functional and can be used to test production and recovery of cellular products.

Computing for Bioenergy

138

Bayesian Computational Approaches for Gene Regulation Studies of Bioethanol and Biohydrogen Production

Charles E. Lawrence¹ (charles_lawrence@brown.edu), Lee Newberg,^{2,3} William Thompson,¹ and Lee Ann McCue^{4*}

¹Center for Computational Molecular Biology and the Division of Applied Mathematics, Brown University, Providence, R.I.; ²The Wadsworth Center, New York State Dept. of Health, Albany; ³Dept. of Computer Science, Rensselaer Polytechnic Institute, Troy, N.Y.; ⁴Pacific Northwest National Laboratory, Richland, Wash.

<http://www.brown.edu/Research/CCMB/>

Project Goals: We are extending our phylogenetic Gibbs sampling algorithms to reconstruct the joint posterior space of the ancestral states of regulatory motifs and

developing point estimates and confidence limits for these discrete high-dimensional objects. We are also applying our existing models and technologies, along with the above modifications, to clades of alpha-proteobacterial species, to identify regulatory mechanisms and reconstruct the ancestral states of the regulatory networks for the efficient fermentation of sugars to ethanol and the production of biohydrogen.

Decreasing America's dependence on foreign energy sources and reducing the emission of greenhouse gases through the development of biofuels are important national priorities. These priorities have catalyzed research on cellulosic ethanol as a clean, renewable energy source to replace fossil fuels, and biohydrogen as a carbon-free energy carrier. Turning these biofuels into viable alternative energy sources requires further research into the degradation of cellulose and fermentation of the resulting sugars, and the metabolic and regulatory networks of biohydrogen production. The genomes of many of the microbial species capable of these processes have been sequenced by the GTL and other programs, and many more are expected soon. These sequence data provide a wealth of information to explore nature's solutions for the production of biofuels. In particular, among the over 170 α -proteobacterial species with genome sequence data available are several species with metabolic capabilities of interest, including efficient fermentation of sugars to ethanol and the ability to produce hydrogen. Understanding the regulatory mechanisms and complex interplay of metabolic processes in these species is key to realizing the promise of biofuels. Thus, our research goal is to identify the ensemble of solutions that have been explored by the α -proteobacteria to regulate the metabolic processes key to biofuel production.

The solution space explored by these species spans three scales: molecular (genes/gene products), cellular (genomic), and communities (clades). On the finest scale, the catalytic steps of biofuel production (ethanol or hydrogen) are performed by individual enzymes that are the products of individual genes, each regulated by a set of *cis* and *trans* elements. At the cellular level, the expression of the gene products is often coordinated via a set of *trans* elements (transcription factors) that interact with all or most genes in these pathways to form a regulatory unit called a regulon. Furthermore, the collection of species that encode the genes form clades which have explored a catalytic and regulatory space on an evolutionary time scale. We are developing probabilistic models to represent these multiscale processes, 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.

Specifically, we are characterizing the gene/clade interface by extending our phylogenetic Gibbs sampling algorithms to reconstruct the joint posterior space of the ancestral states of regulatory motifs, and developing point estimates and confidence limits for these discrete high-dimensional objects. We are also applying our existing models and technologies, along with the above modifications, to clades of

α -proteobacterial species, to identify regulatory mechanisms and reconstruct the ancestral states of the regulatory networks for the efficient fermentation of sugars to ethanol and the production of biohydrogen.

139

Sugar-Salt and Sugar-Salt-Water Complexes: Structure and Dynamics of Glucose – $\text{KNO}_3 - (\text{H}_2\text{O})_n$

Madeleine Pincu^{1,2*} (mpincu@uci.edu), Brina Brauer,² Robert Benny Gerber,^{1,2} and Victoria Buch^{2†}

¹The Fritz Haber Center for Molecular Dynamics, Chemistry Institute, the Hebrew University, Jerusalem, Israel and ²Dept. of Chemistry, University of California, Irvine

[†]Deceased

Project Goals: To investigate structure, dynamics, and energetics of saccharides and their interactions with ions and biological osmolytes in the presence of microhydration.

Molecular dynamics (MD) simulations are carried out for the complex of glucose with KNO_3 and for complexes of the type glucose – $\text{KNO}_3 - (\text{H}_2\text{O})_n$, for $n \leq 11$. Structure and dynamic properties of the systems are explored. The MD simulations are carried out using primarily the DLPOLY/OPLS force field, and global and local minimum energy structures of some of the systems are compared with ab initio MP2 calculations. The main findings include: (1) Complexation with KNO_3 leads to an “inverse anomeric effect,” with the β -glucose complex more stable than the α -glucose by ~ 1.74 Kcal/mol. (2) As temperature is increased to 600K, the KNO_3 remains undissociated in the 1:1 complex, with the K^+ hooked to the equilibrium site, and the NO_3^- bound to it, undergoing large-amplitude bending/torsional motions. (3) For $n \geq 3$ water molecules added to the system, charge separation into K^+ and NO_3^- ions takes place. (4) For $n = 11$ water molecules all hydroxyl groups are hydrated with the glucose adopting a surface position, indicative of a surfactant property of the sugar. (5) Comparison of DLPOLY with MP2 structure predictions indicates that the empirical force field predicts global and local minimum structures reasonably well, but errs in giving the energy rankings of the different minima. Implications of the results to effects of salts on saccharides are discussed.

submitted post-press

A Multi-Scale Approach to the Simulation of Lignocellulosic Biomass

Goundla Srinivas,¹ Yana Kholod,² Loukas Petridis,¹ James W. Evans,² Monica Lamm,² Theresa L. Windus,² Xiaolin Cheng^{1,3*} (chengx@ornl.gov), **Mark S. Gordon**,² and **Jeremy C. Smith**^{1,3}

¹Oak Ridge National Laboratory, Oak Ridge, Tenn.;

²Ames Laboratory, Iowa State University, Ames; and

³University of Tennessee, Knoxville

Project Goals: In concert with the imminent increase in the Department of Energy's (DOE) leadership supercomputing power to the petaflop range, the objective of this project is to develop multiscale methods for extending the time- and length-scales accessible to biomolecular simulation on massively parallel supercomputers. This project also aims to apply the developed multiscale approaches to obtain an understanding of the structure, dynamics and degradation pathways of extended cellulosic and lignocellulosic materials. Information from multiscale simulation, when closely integrated with experiment, will provide fundamental understanding needed to overcome biomass recalcitrance to hydrolysis.

The research involves the development of multiscale simulation methods and their application to solve critical problems needed for understanding the bottleneck in cellulosic ethanol production: the recalcitrance to hydrolysis of lignocellulosic biomass. The multiscale methodologies span from accurate quantum-chemical techniques, needed to understand critical local interactions in biomass, to atomistic and coarse-grained simulations, needed to approach systems-level phenomena. The codes developed will be parallelized for efficient use on petascale supercomputers. The physical simulation models of lignocellulosic biomass derived using the multiscale approaches will serve as a basis for interpreting an array of biophysical experiments, and when closely integrated with experiments, will eventually lead to a description of the physicochemical mechanisms of biomass recalcitrance to hydrolysis, and thus aiding in developing a strategy to overcome the recalcitrance. Our progress presented here mainly focus on: adapting the quantum mechanical fragment molecular orbital (FMO) method to the study of cellulosic/lignocellulosic biomass, developing an adaptive fast multipole based Poisson Boltzmann (PB) electrostatic solvation model for nano-scale biomolecular systems and self-consistently coarse-graining of cellulose/lignocellulose force fields.

This research is funded by the Genomic Science Research Program, Office of Biological and Environmental Research (BER), and the Scientific Discovery through Advanced Computing (SciDAC) program, U. S. Department of Energy (DOE), currently under FWP ERKJE84.

submitted post-press

Processivity of Cellobiohydrolase, Cellulose Structure, and Advanced Methods for Petascale Molecular Dynamics

James F. Matthews,¹ Xiaolin Cheng,² Malin Bergenstrahle,³ Eric H. Knoll,¹ Mark Williamson,⁴ Edward C. Uberbacher,² John W. Brady,³ Charles L. Brooks III,⁵ Linghao Zhong,⁶ Mark R. Nimlos,¹ Michael E. Himmel,¹ and **Michael F. Crowley**^{1*} (Michael.Crowley@nrel.gov)

¹National Renewable Energy Laboratory, Golden, Colo.; ²Oak Ridge National Laboratory, Oak Ridge, Tenn.; ³Cornell University, Ithaca, N.Y.; ⁴San Diego Supercomputer Center, San Diego, Calif.; ⁵University of Michigan, Ann Arbor; and ⁶Pennsylvania State University, Mont Alto

Project Goals: The SciDAC project is focused on molecular modeling of molecular and macromolecular systems that are key to the understanding and designing of cellulose degrading technologies acting on biomass. The project is enabling scientific investigation of cellulose structure and properties and enzymatic mechanism (processivity) for degrading cellulose to sugars as a raw material for biofuels. Speed, new sampling methods, and simulation size (scaling) are the aims of code development and performance work with the additional creation of validation suites for code integrity. The main thrust is to enable the specific needs of modeling to produce the methods necessary for thermodynamic, kinetic, and mechanistic insight into cellulose conversion processes at a speed which will make this research possible, mainly through the use of high-performance computers at the petascale level and advanced sampling methods.

The mechanism by which Cel7A cellobiohydrolase degrades cellulose is not currently known nor is the structure and molecular properties of cellulose microfibrils as found in biomass. The degrading of cellulosic biomass to sugars is primarily accomplished both in nature and in biorefineries by enzymes such as Cel7A; the understanding of both the structure of cellulose and of the mechanism of cellulose-degrading enzymes is critical to improving the technology of producing biofuels from biomass. There have been several barriers to studying these molecular systems with modeling, primarily the lack of reliable force fields for cellulose and the lack of highly scalable molecular dynamics programs that can treat these systems and have the particular thermodynamic sampling abilities and force field features to answer the questions which are unique to this difficult problem. Our progress presented here is fivefold: cellulose force field evaluation and characterization, molecular dynamics code validation, enabling highly parallel programs to utilize cellulose force fields, solving parallel bottlenecks in existing codes, and designing highly parallel implementations of sampling methods for studying enzymatic mechanisms and cellulose morphology and decrystallization.

The SciDAC project “Understanding Processivity of Cellobiohydrolases” is supported by the Office of Biological and Environmental Research and OASCR in the DOE Office of Science.

Small Business Innovation Research (SBIR)

140

Genetic Analysis of Cellulose Degradation by *Clostridium phytofermentans*

Andrew C. Tolonen^{1*} (tolonen@genetics.med.harvard.edu), Amanda C. Chilaka,^{1,2} and **George M. Church¹**

¹Dept. of Genetics, Harvard Medical School, Boston, Mass. and ²Dept. of Biology, Northeastern University, Boston, Mass.

Project Goals: See below.

Microbial cellulose degradation is a central part of the global carbon cycle and has potential for the development of inexpensive, carbon neutral biofuels from non-food crops. The major roadblock to the use of cellulosic biomass as a biofuel feedstock is the recalcitrance of cellulosic fibers to breakdown into sugars. *Clostridium phytofermentans* grows on both of the two main components of plant biomass, cellulose and hemicellulose, by secreting enzymes to cleave these polysaccharides and then fermenting the resulting hexose and pentose sugars to ethanol. In order to breakdown cellulose biomass, *C. phytofermentans* has a repertoire of 161 carbohydrate-active enzymes (CAZy), which include 108 glycoside hydrolases spread across 39 families.

Broadly, our goal to understand the genetic mechanisms that permit to *C. phytofermentans* to efficiently convert cellulosic biomass to ethanol. To enable targeted gene inactivation in *C. phytofermentans*, we show that interspecific conjugation with *E. coli* can be used to transfer a plasmid into *C. phytofermentans* that has a resistance marker, an origin of replication that can be selectively lost, and a designed group II intron for efficient, targeted chromosomal insertions without selection. We applied these methods to inactivate Cphy3367, a β -1,4-glucanase in glycoside hydrolase family 9 (GH9). Cellulolytic *Clostridia* usually have numerous genes for GH9 proteins: the *C. thermocellum* ATCC 27405 genome has 16 GH9 genes, *C. cellulolyticum* H10 has 13 GH9 genes, and *C. cellulovorans* has 5 GH9 genes. In contrast, *C. phytofermentans* has only a single GH9-encoding gene, *cphy3367*. The *C. phytofermentans* strain with an intron insertion in *cphy3367* (strain AT02-1) grows normally on some carbon sources such as glucose, cellobiose, and hemicellulose, but has lost the ability to degrade cellulose (Fig 1). Although *C. phytofermentans* up-regulates the expression of numerous enzymes to breakdown cellulose, this process thus relies upon a single, key hydrolase, Cphy3367. Generally,

these results show that targeted gene inactivation can be used to identify key enzymes for the breakdown of biomass by *C. phytofermentans*. Future genetic studies of in *C. phytofermentans* will untangle the roles of additional hydrolases in cellulose degradation.

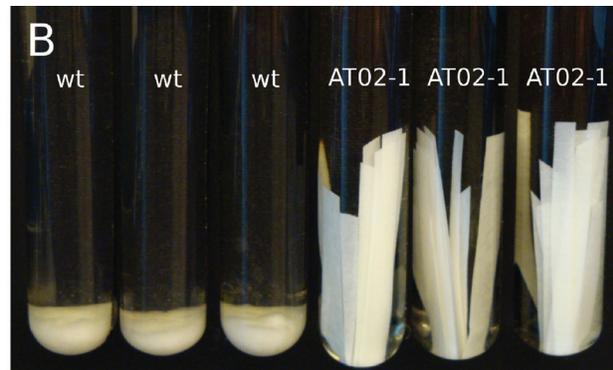
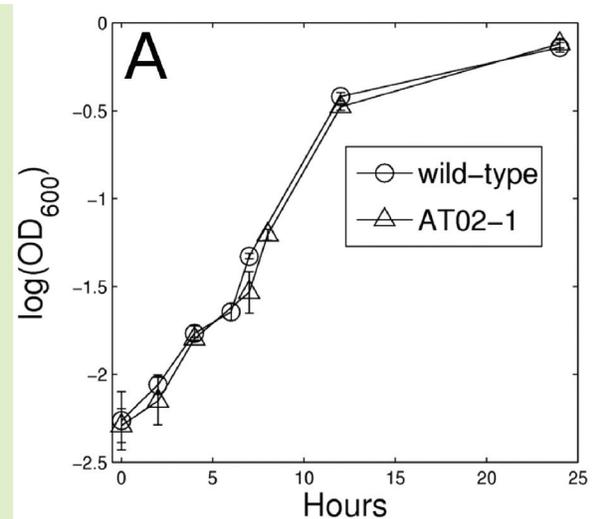


FIG 1 *C. phytofermentans* strain with disruption of *cphy3367* (AT02-1) had similar growth rates as wild-type on glucose A, but had lost the ability to degrade filter paper cellulose B. Growth curves are means of triplicate cultures. Error bars show one standard deviation and are smaller than the symbols where not apparent.

141

Microbioreactor Technology for Obligate Anaerobes

Harry Lee (harrylee@pharyx.com) and Paolo Boccazzi* (boccazzi@pharyx.com)

Pharyx, Inc., Boston, Mass.

<http://www.pharyx.com>

Project goals are 1) determine microbioreactor designs that will support anaerobic inoculation and fermentation, 2) identify optimal materials for fabricating anaerobic bioreactors, 3) determine the range of process parameters where microbioreactor data corresponds to serum tubes and stirred tank fermentors, 4) monitor enzyme activity on-line.

Anaerobic microorganisms have evolved biochemical pathways that can be exploited for industrial applications. These include the ability to breakdown environmental pollutants for bioremediation, the breakdown of cellulose into simple sugars for biofuels, and the production of specialty chemicals. However, there remains a tremendous challenge to the scale-up of bioenzymatic activities to industrial processes. While systems biology approaches and metabolic engineering promise to contribute to our understanding of these systems, a key bottleneck is in conducting controlled experiments to ground these approaches with high quality data. Thus far, experiments are frustrated by the laborious set-up and operation of stirred tank bioreactor systems, which for anaerobic microbiology is further encumbered by the requirement of an anaerobic environment. The absence of easy to use systems also holds back more traditional microbiology approaches such as mutagenesis and screening and directed evolution.

We are developing a parallel bioreactor system, based on microfluidic integration technology and disposable microbioreactor modules, with application specific customizations for anaerobic fermentation. These customizations are aimed to enable up to 32 simultaneous anaerobic fermentations under controlled conditions, with online monitoring of growth kinetics and other phenotypes such as enzyme activity. A unique feature of this system is the ability to operate it in ambient air through careful inoculation port and reactor and control module design, or to operate it within an anaerobic bag, taking advantage of its compact size.

Preliminary anaerobic fermentations in microbioreactor devices are shown in Figure 1. Cell growth in the three bioreactor chambers compared to the uninoculated medium control is clear, and exponential growth can be seen for $t < 5$ h, indicating obligate anaerobes can be cultured in plastic bioreactors fabricated using our microfluidic integration platform. However, for the *Butyrivibrio fibrisolvens* and *Clostridium acetobutylicum* fermentations, optical density measurements were confounded by the gas bubbles generated by the microorganisms. This was confirmed visually during the fermentation and evidenced in the photographs,

shown in Figure 2, of the devices at the end of the fermentation.

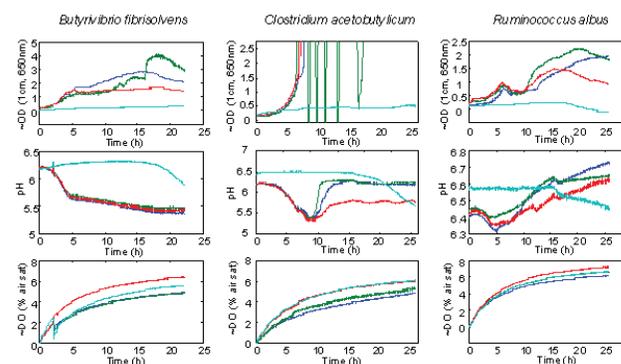


Figure 1. Online sensor measurements of dissolved oxygen (DO), pH and optical density (OD) for anaerobic fermentations of three obligate anaerobes. Growth studies were conducted in our anaerobic bioreactor modules which contained 4 growth chambers each. Light blue lines represent the uninoculated controls, other 3 lines represent 3 independent reactor replicates in each cassette. All strains were grown in modified peptone yeast extract glucose (MPYG) medium (Atlas RM, 2004. Handbook of microbiological media. CRC, Boca Raton). Measurements were taken every 60 seconds.

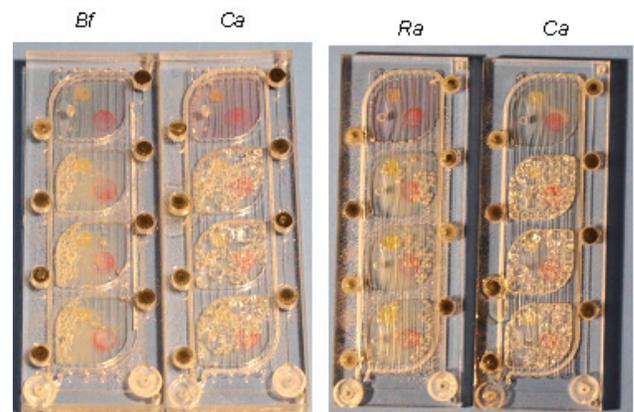


Figure 2. Photographs of integrated bioreactor devices. Pink circles are dissolved oxygen sensors, yellow circles are pH sensors, black plugs are butyl rubber injection ports. A stream of nitrogen flowed through the serpentine channels that surround the chamber are biologically generated gas. Bf (*Butyrivibrio fibrisolvens*), Ca (*Clostridium acetobutylicum*), Ra (*Ruminococcus albus*). First growth reactor from top of each cassette contained uninoculated medium.

142

A Genome-Wide Perspective on the Regulation of Plant Carbohydrate Conversion to Biofuels in *Clostridium phytofermentans*

Elsa Petit,¹ Tom Warnick,¹ Greg Latouf,^{1,6} James Hayes,^{2,5} Amy Biddle,^{1*} Zhiyi Sun,¹ Danny Schnell,^{2,4,5} **Sue Leschine**,^{1,4,5} and **Jeff Blanchard**^{1,4,5} (blanchard@microbio.umass.edu)

¹Dept. of Microbiology, ²Dept. of Biochemistry, ³Graduate Program in Molecular and Cellular Biology, ⁴The Massachusetts Institute for Biofuels Research, and ⁵Institute for Cellular Engineering, University of Massachusetts, Amherst and ⁶Qteros Inc.

Project Goals: Derive genetic models of the conversion of cellulosic biomass to alcohols in *C. phytofermentans* by microarray analysis of the fermentation of simple sugars contained in feedstocks relevant to the agricultural and forestry industries.

The economic costs of degrading cellulosic biomass currently hold back widespread use of plant biomass from agricultural and forestry wastes as a domestic renewable alternative to gasoline. *Clostridium phytofermentans* ferments all major components of the plant cell wall to ethanol and is an emerging model organism for understanding the direct conversion of plant biomass to fuels. The *C. phytofermentans* genome contains an abundance of carbohydrate degradation enzymes, carbohydrate transporters and transcriptional regulators. In order to build a foundation for engineering improvements on industrially relevant feedstocks, microarray experiments were carried out on a number of purified plant cell wall carbohydrates. The results demonstrate that *C. phytofermentans* regulates the stoichiometry of the plant degradative and assimilatory machinery in response to substrate availability. Because of the modularity of the sugar degradation and assimilation systems and the lack of a cellulosome, further improvement in biomass degradation through genetic manipulation is relatively straightforward.