

**Sections:**

**Systems Biology for DOE Energy Missions: Bioenergy**

Biofuels: Analytical and Imaging Technologies,  
Engineering, and Production

Systems Biology and Metabolic Engineering Approaches  
for Biological Hydrogen Production



U.S. DEPARTMENT OF  
**ENERGY**

Office of Science

# Joint Meeting 2011

## Genomic Science Awardee Meeting IX

and

## USDA-DOE Plant Feedstock Genomics for Bioenergy Awardee Meeting

[Revised: April 14, 2011]

**Crystal City, Virginia**  
**April 10-13, 2011**

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

<http://genomicscience.energy.gov>

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

and assure the traceability of intellectual property. JBEI has developed an integrated ELN system based on commercially successful, off-the-shelf products. The organization of this system will be presented. Our initial deployment to a pilot user community has yielded encouraging results, which will be discussed. We are continuing with improvements to the system and development of support for additional user interface formats.

This work conducted by the Joint BioEnergy Institute was supported by the Office of Science, Office of Biological and Environmental Research, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

## 68

### DNA Assembly Design with j5 and DeviceEditor Biocad Tools

J. Chen, R.D. Rosengarten, D. Densmore, T.S. Ham, J.W. Thorne, J.D. Keasling, and **N.J. Hillson\*** (njhillson@lbl.gov)

Joint BioEnergy Institute, Fuels Synthesis Division, Lawrence Berkeley National Laboratory, Emeryville, Calif.

<http://www.jbei.org>,  
<http://jbei-exwebapp.lbl.gov/j5>

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 have developed two on-line software tools, j5 and DeviceEditor, that automate the design of sequence agnostic, scar-less, multi-part assembly methodologies and translates them to robotics-driven protocols. Given a target library to construct, the software provides automated oligo, direct synthesis, and cost-optimal assembly process design, and integrates with liquid-handling robotic platforms to set up the PCR and multi-part assembly reactions. This work reduces the time, effort and cost of large-scale cloning and assembly tasks, as well as enables research scales otherwise unfeasible without the assistance of computer-aided design tools and robotics.

This work conducted by the Joint BioEnergy Institute was supported by the Office of Science, Office of Biological and Environmental Research, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

## Biofuels: Analytical and Imaging Technologies, Engineering, and Production

## 69

### Engineering Bacterial Surfaces to Display Cellulosomes for Biofuel Production

**Robert Clubb\*** (rclubb@mbi.ucla.edu)

University of California, Los Angeles

**Project Goals: Engineer surface of *Bacillus subtilis* to display multi-enzyme cellulolytic protein complexes that can efficiently degrade biomass into fermentable sugars.**

To cost-efficiently produce biofuels, improved methods are needed to convert lignocellulosic biomass into fermentable sugars. One promising approach is to degrade biomass using cellulosomes, surface displayed multi-cellulase containing complexes present in cellulolytic *Clostridium* and *Ruminococcus* species. In this study we created genetically modified strains of *B. subtilis* that display on their surface heterologous proteins and protein-protein complexes. Proteins containing the appropriate cell wall sorting signal are covalently anchored to the peptidylglycan by co-expressing them with the *B. anthracis* sortase A (SrtA) transpeptidase (SrtA). Greater than 300,000 heterologous proteins per cell are displayed in strains in which the WprA cell wall protease has been deleted. A two-component minicellulosome was constructed that consists of a cell wall attached scaffoldin protein that non-covalently binds to the CelA endoglucanase from *C. thermocellum*. Unlike the wild-type organism, *B. subtilis* displaying the minicellulosome robustly grow on acid-treated cellulose by degrading it into its component sugars. Importantly, the cells exhibit greater cellulolytic activity than several previously reported in vitro and yeast displayed minicellulosomes. *B. subtilis* has a robust genetic system and is currently used in a wide range of industrial processes. Thus, grafting more elaborate multi-enzyme containing minicellulosomes onto the surface *B. subtilis* may yield more potent cellulolytic bacteria that can be used to degrade biomass.

## 70

## Employing Cyanobacteria for Biofuel Synthesis and Carbon Capture and Storage

Christer Jansson<sup>1\*</sup> (PI) (CGJansson@lbl.gov), Hoi-Ying Holman<sup>1,2</sup> (Co-PI), Ping Hu<sup>1</sup> (Co-PI), Nina Kamennaya,<sup>1</sup> Laura Mahoney,<sup>1</sup> and Elizabeth Holman<sup>1,3</sup>

<sup>1</sup>Earth Sciences Division (ESD), Lawrence Berkeley National Laboratory, Berkeley, Calif., <sup>2</sup>Advanced Light Source (ALS), and <sup>3</sup>School of Chemistry, University of California, Berkeley

**Project Goals:** In our two inter-linked research projects, CyanoFuels and CyanoCarbon, we explore the potential for employing cyanobacteria in biofuel synthesis, and biological carbon capture and storage (CCS), respectively. We exploit the two major modes of CO<sub>2</sub> uptake exhibited by cyanobacteria, the photosynthetic conversion of CO<sub>2</sub> to biomass and metabolites, as well as the biomineralization of CO<sub>2</sub> and Ca<sup>2+</sup> to calcium carbonate (CaCO<sub>3</sub>) (Figure 1). We utilize synchrotron radiation Fourier transform infrared (SR-FTIR) spectromicroscopy as a non-invasive technique for metabolic fingerprinting of individual live cells in real time.

**Background:** Cyanobacteria are photoautotrophic Gram-negative bacteria that carry out oxygenic photosynthesis. Many are diazotrophic and thus can assimilate not only CO<sub>2</sub> but also N<sub>2</sub> from the atmosphere. Cyanobacteria occupy a wide array of terrestrial, marine, and freshwater habitats, which include extreme environments such as hot springs, deserts, bare rocks, and permafrost zones. In their natural environments, some cyanobacteria are often exposed to the highest rates of UV irradiance known on our globe. Cyanobacteria generally thrive in high CO<sub>2</sub> levels and are considered attractive systems for CO<sub>2</sub> capture from flue gas. Since many cyanobacteria are halophilic and, strains for biofuel production or carbon capture and storage (CCS) can be cultured in either marine waters, saline drainage water, brine from petroleum refining industry, or CO<sub>2</sub> injection sites, thereby sparing freshwater supplies. A large number of strains are thermophilic and thus tolerate high temperatures characteristic of flue gas. Also, being bacteria, cyanobacteria are amenable to homologous recombination, which allows rapid site-directed mutagenesis, gene insertions, replacements and deletions in a precise targeted and predictable manner.

In our two inter-linked research projects, CyanoFuels and CyanoCarbon, we explore the potential for employing cyanobacteria in biofuel synthesis, and biological carbon capture and storage (CCS), respectively. We exploit the two major modes of CO<sub>2</sub> uptake exhibited by cyanobacteria, the photosynthetic conversion of CO<sub>2</sub> to biomass and metabolites, as well as the biomineralization of CO<sub>2</sub> and Ca<sup>2+</sup> to calcium carbonate (CaCO<sub>3</sub>) (Figure 1). We utilize synchrotron radiation Fourier transform infrared (SR-FTIR) spectromicroscopy as a non-invasive technique for metabolic fingerprinting of individual live cells in real time.

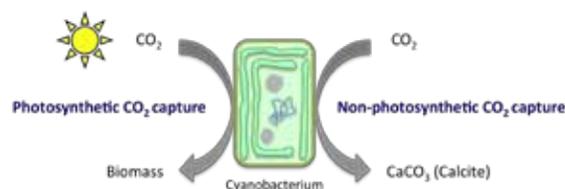


Figure 1. Two modes of CO<sub>2</sub> uptake in cyanobacteria.

**CyanoFuels:** We are using the cyanobacterium *Synechocystis* PCC6803 (*S. 6803*) as a model to study phototrophic biosynthesis of free fatty acids (FAs) and liquid biofuels such as FA alkylesters (biodiesel), alkanes, and isoprenoids. Cyanobacteria synthesize a wide array of hydrocarbons, including terpenes such as carotenoids and hopanoids, as well as linear and branched alkanes and alkenes, like branched methyl- and ethylalkanes which are unique to cyanobacteria. Biosynthesis of free FAs is of interest since they can be used for downstream chemical processing to biofuels and also because cyanobacterial biosynthesis of alkanes with specified chain lengths may require free FAs as an intermediate metabolite. Using SR-FTIR with multivariate analysis of *S. 6803* strains engineered for free FA accumulation, we were able to demonstrate that the metabolic state of a strain could be fingerprinted at the single cell level. We conclude that SR-FTIR spectromicroscopy offers the opportunity to be employed as a high-throughput diagnostic tool for characterization and screening of genetically engineered cells.

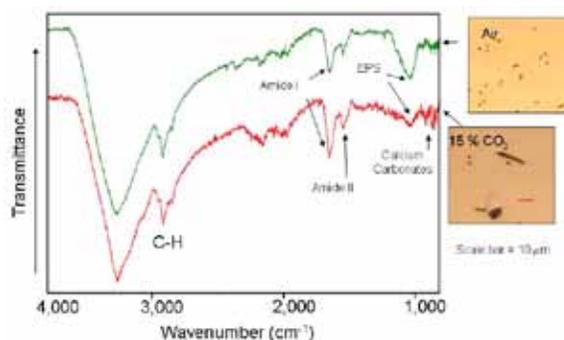


Figure 2. SR-FTIR spectra of calcification in the cyanobacterium *Synechococcus* 8802.

**CyanoCarbon:** Biomineralization of atmospheric or dissolved CO<sub>2</sub> by calcium carbonate (CaCO<sub>3</sub>) precipitation (calcification) is a common phenomenon in marine, freshwater, and terrestrial ecosystems and is a fundamental process in the global carbon cycle. Formation and deposition of CaCO<sub>3</sub> occur at the outer cell surface, either at the proteaceous surface layer (S layer) or in the exopolysaccharide substances (EPS). The cyanobacterial CO<sub>2</sub> concentrating mechanism (CCM) is often a critical part of the calcification process. Biological or biomimetic strategies for point-source CCS of flue gas based on cyanobacterial calcification offer potential strategies for reducing anthropogenic CO<sub>2</sub> emissions. However, much research is urgently needed to further our understanding of the biochemical and physical

processes in cyanobacteria that promote calcification. We are studying calcification in several lacustrine and marine cyanobacteria under different CO<sub>2</sub> levels, e.g. 15% CO<sub>2</sub> typical of flue gas. We use SR-FTIR to monitor CaCO<sub>3</sub> crystallization and morphology, and to follow the metabolic shifts in cells grown under increasing CO<sub>2</sub> concentrations (Figure 2). We apply SIMS to study the cell surface properties and their influence in the calcification process.

## 71 *Synechococcus* Sp. Pcc 7002: A Robust Cyanobacterial Platform For Systems Biology and Biofuels Engineering

D.A. Bryant<sup>1\*</sup> (dab14@psu.edu), A.E. Konopka,<sup>2</sup> G.E. Pinchuk,<sup>2</sup> S.M. Stolyar,<sup>2</sup> M.C. Posewitz,<sup>3</sup> and A.S. Beliaev<sup>2</sup>

<sup>1</sup>Dept. of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park; <sup>2</sup>Biological Sciences Division, Pacific Northwest National Laboratory, Richland, Wash.; and <sup>3</sup>Colorado School of Mines, Dept. of Chemistry and Geochemistry, Golden

**Project Goals: The PNNL Biofuels Scientific Focus Area (BSFA) will conduct fundamental research on cyanobacteria with specific emphasis on pathways of carbon, nitrogen, and redox metabolism that consume reductant and conserve energy produced by photosynthetic light reactions. Consistent with DOE BER Genomic Sciences Program goals, the long-term objective of the BSFA is to develop a predictive understanding of metabolic subsystems and regulatory networks involved in solar energy conversion to biofuel precursors or products. Toward this goal, we will interrogate fluxes through metabolic subsystems that include (i) photosynthetic energy conservation and reductant generation, (ii) CO<sub>2</sub> accumulation, fixation, and reduction, (iii) biosynthesis of metabolic intermediates and monomers, and (iv) macromolecular synthesis. Our studies will exploit two complementary model cyanobacterial systems, *Cyanothece* sp. strain ATCC 51142 and *Synechococcus* sp. strain PCC 7002, and address discrete scientific hypotheses.**

The PNNL Biofuels Scientific Focus Area (BSFA) conducts fundamental research on cyanobacteria with specific emphasis on pathways of carbon and redox metabolism that consume reductant produced by photosynthetic light reactions. The long-term objective is to develop a predictive understanding of metabolic subsystems and regulatory networks involved in solar energy conversion to biofuel precursors or products. This requires a systems-level understanding of metabolic modules (see Figure 1) at levels beyond the biophysics of photosynthesis and the primary production of fixed CO<sub>2</sub>. Optimizing biofuel production by cyanobacteria also requires an improved understanding of the physiological constraints and the regulatory controls affecting “downstream” carbon partitioning between biomass, metabolite storage pools, and desired product synthesis. Therefore, understanding the factors constraining

maximum rates of carbon processing, including production of autocatalytic macromolecules (i.e., RNA and protein) is necessary to identify regulatory mechanisms and devise strategies to overcome and manipulate these constraints. A systems approach to these issues is being undertaken, exploiting the fast-growing cyanobacterium, *Synechococcus* sp. PCC 7002, which has the added benefit of a well-developed genetic system. Our experimental strategy dissects growth-limiting factors by introducing external (i.e., light and CO<sub>2</sub> availability) and internal perturbations (i.e., gene knock-outs and knock-ins) that affect the internal state of the cell and, then investigating system performance in response to variable growth rates imposed by chemostat culture. The effect of these perturbations is analyzed by assaying the state and activity of key markers within the metabolic subsystems and allows calculation of flux distributions through the system that, in combination with activity measurements are used for estimating the performance of metabolic modules.

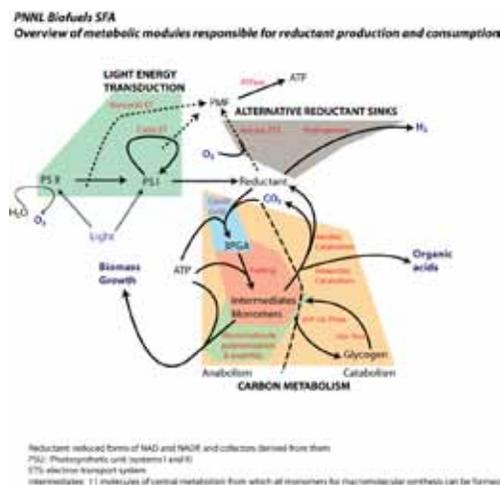


Figure 1. Metabolic modules of cyanobacteria. A central focus is the formation of reductant and conservation of energy from the coupled operation of Photosystems I and II (light energy transduction module), consumption of reductant via CO<sub>2</sub> reduction in the Calvin cycle or via other reductant sinks, and energy consumption (primarily the Calvin cycle or macromolecule polymerization) or storage (as glycogen). Anoxic or oxic catabolism of stored glycogen can generate ATP or ATP + reductant, respectively. “Carbon metabolism” is comprised of three principal submodules essential for biomass synthesis; CO<sub>2</sub> fixation; fueling reactions that produce the 11 precursor metabolites and ~75-100 building blocks from which all cellular molecules are formed; and macromolecule polymerization; in addition, the carbon metabolism entails glycogen synthesis and catabolism.

An experimentally tractable system is important for rapid progress and *Synechococcus* 7002 presents an attractive platform for functional genomics and biotechnological applications through metabolic engineering. This cyanobacterium has several useful and remarkable properties: it grows much faster than other well-characterized cyanobacteria; it tolerates extremely high light intensities; it is highly resistant to reactive oxygen and nitrogen species; it grows over a very wide range of salinity; it is easily manipulated genetically by natural transformation; and it can be grown photomixotrophically, photoheterotrophically, or heterotrophically on glycerol. The relatively small 3.4-Mb genome has been

completely sequenced and comprises a 3.0 Mb chromosome and 6 plasmids, which vary in size from 4.8 to 186 kb and in copy number (up to ~10 times the chromosome number for pAQ1). Using neutral-site integration platforms within these plasmids, a flexible and effective gene expression system has been developed. Various promoters, including  $P_{cpcBA}$ ,  $P_{psbA1}$ ,  $P_{ntcABCD}$ , and  $P_{cpcC}$ , allow gene expression at differing levels. Two of these systems allow regulatable expression using either N-source or light wavelength.

These advances have established *Synechococcus* 7002 as an exceptional platform for developing a more informed understanding of how metabolic networks are regulated and integrated in cyanobacteria. A thorough understanding of the fundamental mechanisms regulating metabolic pathway flux and photosynthetic activity will inform iterative metabolic engineering approaches designed to improve the synthesis/accumulation of products relevant to DOE bioenergy missions. We have over-expressed a variety of genes, including those encoding soluble and membrane proteins, complex multi-subunit proteins, and various tagged and fusion proteins. Based on transcript levels determined by cDNA sequencing, the *ldhA* gene, encoding D-lactate dehydrogenase, was overproduced by ~3000-fold relative to transcript levels in wild-type cells when expression was driven by the *cpcBA* promoter. The resulting strain excreted D-lactate and acetate continuously for up to 58 days; this led to the accumulation of up to 20 mM D-lactate and nearly 80 mM acetate in the external medium. This system has also been used in complementation analyses, in the production of novel carotenoids, in the expression of a heterologous, oxygen-tolerant [NiFe]-hydrogenase from *Ralstonia eutropha* (13 genes encoded by 14 kb) and in producing YFP-protein fusions for protein localization studies. Mutants lacking various fermentative enzymes can increase hydrogen production or redirect metabolites during dark anoxic conditions. Mutants lacking the ability to synthesize glycogen can be osmotically manipulated to store and then rapidly excrete up to 200  $\mu\text{g/ml}$  glucosylglycerol and substantial amounts of sucrose.

We are also in the process of generating four distinct pAQ1-based constructs for expressing enzymes that will allow the synthesis of either fungible fuel feedstocks or biopolymer monomers in *Synechococcus* sp. PCC 7002. These include medium chain (C10-14) thioesterases, an alkane biosynthesis operon, a  $\beta$ -caryophyllene synthase (sesquiterpene), and an oleoyl-12-hydroxylase (biopolymer monomer). These enzymes have been shown to generate the desired product when transgenically expressed in a variety of other organisms, however, product yields are typically low. Regulatory and metabolic constraints are also likely to emerge in *Synechococcus* sp. PCC 7002 that limit the overall conversion efficiencies and produce unintended effects on photosynthetic efficiencies. Systematic efforts are therefore required to examine in detail the metabolic and regulatory consequences of introducing exogenous metabolic pathways into *Synechococcus* sp. PCC 7002. These efforts are anticipated to result in the production of the desired product, but are also likely to reveal limitations that must be addressed by subsequent research to improve yields.

## 72

### Constraint-Based Modeling of Cyanobacterial Metabolism

J.L. Reed<sup>1\*</sup> (reed@engr.wisc.edu), T. Vu,<sup>1</sup> S.M. Stolyar,<sup>2</sup> G.E. Pinchuk,<sup>2</sup> O.V. Geydebekht,<sup>2</sup> E.A. Hill,<sup>2</sup> L. Kucek,<sup>2</sup> J. De Ingeniis,<sup>3</sup> A. Osterman,<sup>3</sup> J. Hamilton,<sup>1</sup> A.E. Konopka,<sup>2</sup> D.A. Bryant,<sup>4</sup> and **A.S. Beliaev**<sup>2</sup>

<sup>1</sup>University of Wisconsin-Madison; <sup>2</sup>Pacific Northwest National Laboratory, Richland, Wash.; <sup>3</sup>Burnham Institute for Medical Research, La Jolla, Calif.; and <sup>4</sup>The Pennsylvania State University, University Park

#### Project Goals: same as abstract above

Oxygenic photosynthetic microorganisms possess a unique ability to convert light energy into chemical energy using water as an electron donor for the reduction of electron carriers (ferredoxin and NADP+). Despite the wealth of information on the mechanistic details of bacterial photosynthesis and supramolecular complexes involved in photosynthesis and CO<sub>2</sub> fixation, a system-level understanding of photosynthetic metabolism is yet to be achieved. Phototrophic microorganisms display varying levels of light conversion efficiencies, which ultimately translate into different rates of electron transfer, ATP/NAD(P)H production, and growth. Understanding the origin of these properties will provide new fundamental insights that can be widely applied to the development of photosynthetic systems for biofuel development. To increase biofuel production using photosynthetic strains we first need to improve our understanding of energy and reductant partitioning in photoautotrophs. This requires connecting energy- and reductant-generating reactions with reactions involved in the 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 and reductant partitioning in cyanobacteria. An important outcome of the project will be the development of predictive tools, i.e., genome-scale metabolic and regulatory models, which will provide a platform for integrating all knowledge and experimental data generated within the project. The models will also have the ability to serve as an in silico tool for intelligently manipulating photosynthetic microorganisms to act as catalysts for solar energy conversion and will potentially allow development of a highly efficient biofuel production process.

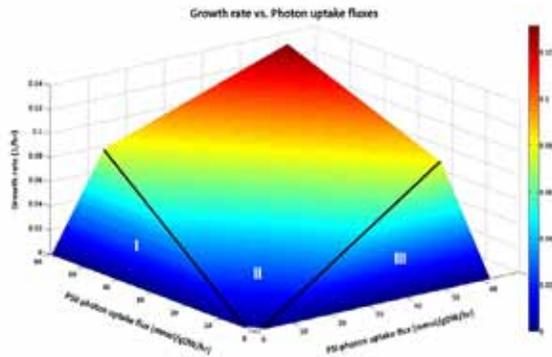


Figure. *In silico* prediction of the effect of photon uptake fluxes on growth rate. The 3D plot shows the solution space of growth rate vs. photon uptake fluxes. Region I and III are single-substrate limiting regions, while region II is dual-substrate limiting region.

We have built a genome-scale metabolic network for *Cyanobacterium* sp. strain 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 806 genes, 585 metabolites, and 664 reactions accounting for common pathways such as central metabolism, respiration, nucleotide and amino acid biosynthesis, as well as those that are more unique to cyanobacteria such as photosynthesis, carbon fixation, and cyanophycin production. Photosynthesis was modeled as a set of three sequential reactions that involve photosystem II, the cytochrome *b<sub>6</sub>f* complex, and photosystem I. This was done in order to study how different wavelengths of light and separate photosystem activities affect electron flow through photosynthesis and respiration, nitrogen source utilization, cellular growth, and hydrogen production (Figure 1). Using batch data collected with different qualities of light, we have used the model to estimate energy requirements for *Cyanobacterium* 51142. We have also used the model to investigate how fluxes through the different photosystems affects cellular growth and to estimate the fluxes through linear and cyclic photosynthesis pathways, cytochrome oxidases, NADH dehydrogenases, Mehler reactions, and ferredoxin-NADP reductase. 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 *Cyanobacterium* 51142 to nitrogen limitation and light limitation imposed on photosystems I and II (Table 1). Biomass composition, gene expression, and protein expression measurements were also carried out for these two chemostat conditions. These experimental datasets have been incorporated into the model to improve flux predictions for both the nitrogen and light-limited chemostat conditions. For these simulations, fluxes were favored through reactions whose proteins were detected, while fluxes through reactions catalyzed by poorly expressed genes were avoided.

Table 1. Model Predicted Growth Rates vs. Experimentally Determined Growth Rates Under Variable Light Quality and Intensity

Batch #	Photon flux through PSII (mmol/gDW/h <sup>-1</sup> )	Photon flux through PSI (mmol/gDW/h <sup>-1</sup> )	Experimental growth rate (h <sup>-1</sup> )	Predicted growth rate (h <sup>-1</sup> )	Error, %
3-A	35.0	34.6	0.079	0.074	6.3
4-A	26.4	53.6	0.080	0.085	6.3
6-A	15.5	19.0	0.035	0.035	0
9-A	26.0	15.6	0.041	0.043	4.9
11-A	13.6	33.4	0.051	0.049	3.9

We have additionally developed a draft metabolic model for *Synechococcus* sp. strain PCC 7002, a fast-growing, non-nitrogen-fixing cyanobacterium which is 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 biofuel development. The *Synechococcus* 7002 model contains 614 genes and 552 reactions. Initial comparisons between the reconstructed metabolic networks of *Cyanobacterium* 51142 and *Synechococcus* 7002 suggested that both networks share a significant number of pathways. Some differences between the two models involve reactions associated with acetate metabolism, adenosyl-homocysteine recycling, and glycoaldehyde transport. Once the model for *Synechococcus* 7002 is complete, we will apply metabolic engineering algorithms to identify genetic strategies for improving chemical and biofuel production.

Early Career Program  
Speaking Wednesday 11:10 a.m.

## 73

### Applying the Biology of Brown Rot Fungi to Consolidated Bioprocessing

Jonathan S. Schilling\* (schilling@umn.edu)

University of Minnesota

**Project Goals: Resolve the reaction partitioning mechanisms of brown rot fungi so that they can be applied to bioconversion of lignocellulosic feedstocks.**

**Hypothesis:** During brown rot, oxidative pretreatments occur ahead of enzymatic saccharification, spatially, and the fungus partitions these reactions using gradients in pH, lignin reactivity, and plant cell wall porosity. These can be recreated without the fungus present for faster bioconversion, integrating otherwise incompatible steps.

Consolidated bioprocessing (CBP) of lignocellulose combines enzymatic sugar release (saccharification) with fermentation, but pretreatments remain separate and costly. In nature, lignocellulose-degrading brown rot fungi

consolidate pretreatment and saccharification, likely using spatial gradients to partition these incompatible reactions. To characterize this relevant biological system, my objectives are to (1) physically sample wood degraded by the brown rot fungus *Postia placenta*, spatially map the coincident locations of pretreatment and saccharification reactions, and correlate with pH and lignin chemistry; (2) to image pH and porosity at the fungus-plant interface and layer this data with images showing cellulase ingress; and (3) map, along the active hyphal front, the co-occurring expression of iron reductases associated with pretreatment and of cellulase involved in saccharification. These are spatially focused goals. Therefore, my respective approaches involve either small-scale, spatially resolved characterization (Objective 1) or appropriately resolved microscopy (Objectives 2 and 3). Small-scale physical sample analysis includes traditional wet chemical characterization, coupled both with spin-trap adduct recovery of hydroxyl radicals produced by the fungus and with C<sup>13</sup>-labeled tetramethyl ammonium hydroxide thermochemolysis for specific brown rot lignin modifications. For microscopy, I am using fluorescence lifetime imaging (FLIM) with confocal detection for pH measurements, cryo-transmission electron microscopy (TEM) with electron tomography for porosity measures, a complementary scanning transmission X-ray approach for porosity, and traditional TEM with immunolabeling to track cellulase ingress. I am also planning to co-localize chitinous fungal biomass, imaged using a traditional WGA-FITC dye, with fluorescence in-situ hybridization to measure mRNA transcribed from iron reductase and endoglucanase DNA sequences, made possible by the recent DOE JGI-funded annotation of the *P. placenta* sequence. Overall, this research will help resolve how brown rot fungi consolidate oxidative pretreatments with enzyme-based saccharification, so that we might better understand and exploit natural synergies between bioconversion steps currently approached as separate, distinct steps.

74

Student Oral Presentation—Tuesday

### Biomimicry of the Fungal Consolidation of Biomass Pretreatment with Saccharification

Justin T. Kaffenberger\* (kaffe002@umn.edu) and  
Jonathan S. Schilling

Department of Bioproducts and Biosystems Engineering,  
University of Minnesota, St. Paul

<http://faculty.bbe.umn.edu/schilling>

**Project Goals:** The goal of this project is to incorporate systems biology to elucidate the brown-rot fungal mechanism for biomass degradation. As a proven evolved biological means of consolidating chemical pretreatment and saccharification of biomass, this mechanism can guide process design in the consolidation of commercial biomass conversion.

Efficient breakdown of ligno-cellulosic material prior to fermentation for bio-fuel production is a key step in optimiz-

ing bio-fuel production. Current industrial practice involves chemical pretreatment and acid or enzymatic saccharification to prepare ligno-cellulosic biomass for fermentation. The commercial viability of bio-fuels hinges in large part on production cost. One approach to reducing production cost is process consolidation. Consolidation not only eliminates the capital costs associated with the additional processing steps, but can also trim operational costs by reducing total reaction time and down time between steps.

In nature, the brown-rot fungus *Postia placenta* successfully performs both chemical pretreatment and enzymatic saccharification of biomass in a consolidated manner. As evidenced by its recently sequenced genome, *P. placenta* produces a limited suite of cellulases, lacking genes for exoglucanases. Despite this limitation, it is still capable of rapidly degrading wood. Brown-rot fungi quickly reduce the degree of polymerization of wood with little weight loss through a reduction-oxidation pathway that yields highly reactive hydroxyl radicals through a chelator-mediated fenton reaction. While hydroxyl radicals would readily damage its cellulases, the fungus manages to conduct both reaction types, enzymatic and oxidative, in the same general location. Examination of the spatial and temporal relationship of these two reaction systems as brown-rot fungal degradation of wood progresses will be presented.

75

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

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

University of Florida, Gainesville

**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. Because so many ions are present, we analyzed standard compounds that are normally present in wood. Full-scan and MS<sup>n</sup> spectra were obtained for  $\beta$  1,4-glucan, 4-O-methylglucuronxylyan,  $\beta$ -glucan, starch and microcrystalline cellulose. MS<sup>2</sup> spectra from intact wood and holocellulose show ions characteristic of cellulose and 4-O-methylglucuronxylyan. The MS image of  $m/z$  1319 a [glu]<sup>6</sup> oligomer characteristic of cellulose shows a uniform signal in young *Populus* stems; however, MS<sup>2</sup> demonstrates two isobaric ions at  $m/z$  1319, one corresponding to cellulose ( $m/z$  995) localized to the secondary xylem and phloem bundles and one additional ion ( $m/z$  1275) localized to the bark (Figure 1). Thus, the use of the linear ion trap (LIT) and MS<sup>n</sup> fragmentation are required to increase the selectivity and interpret the complex spectra for mapping the distribution of cellulose, lignin and hemicelluloses within the biomass during pretreatment and hydrolysis.

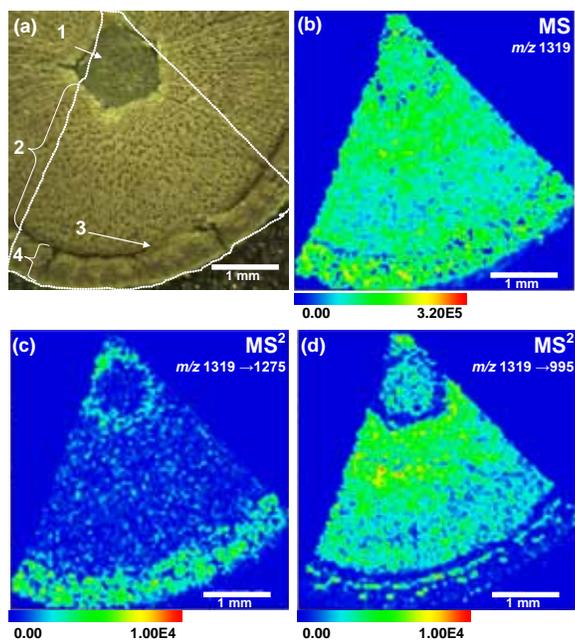


Figure 1. a) optical image, b) MS of  $m/z$  1319, c) MS<sup>2</sup> of  $m/z$  1319-1275, d) MS<sup>2</sup> of  $m/z$  1319 to 995

**Magnetic Resonance Microscopy:** Excellent image quality is obtained from *Populus* wood and bagasse samples using T2 and diffusion weighted modes. New images were collected with state of the art rf microcoils at 8  $\mu$ m resolution. Figure 2 shows the improved signal to noise ratios provide much better images.

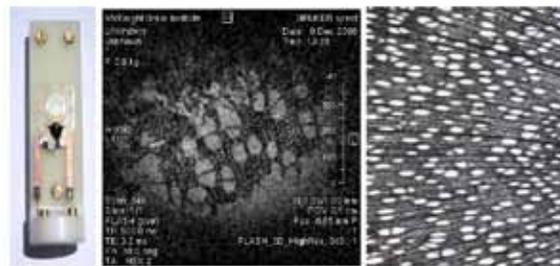


Figure 2. Using the new rf microcoils (left) this figure shows an MR microimage (8 microns, 4hrs 33 min) of a wood chip sample (middle) with example optical histology 5X objective.

For quantitative diffusion imaging, a series of scans are acquired with diffusion encoding in several different directions. From this data, the calculated diffusion tensor reflects the diffusion anisotropy within materials. Within the wood chip, diffusion occurs preferentially along, rather than across, the xylem/phloem vessels. Figure 3 shows the resultant fractional isotropy (FA) map that is quantitatively generated. Blue represents no anisotropy (as seen in the water surrounding the wood chip), while within the chip there is considerable anisotropy (red and yellow speckling). Note the blue regions in the wood chip, which represent the isotropic diffusion of the relatively unrestricted water within the xylem vessels.

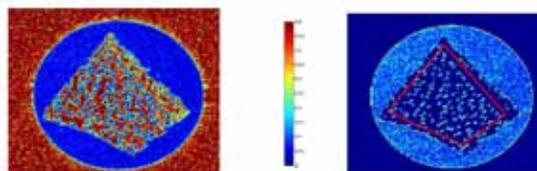
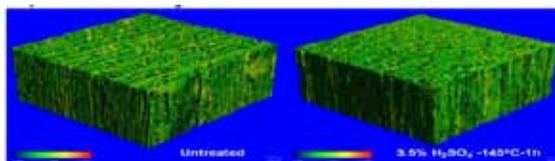


Figure 3. Wood chip FA map with the vessels are clearly defined (blue holes). TR/TE=2000/35.6, FOV=0.6 cm, in-plane resolution=31  $\mu$ m, slice thickness = 1.0 mm, 19 averages, diffusion time ( $\Delta$ )=25 msec, b-value =900 s/mm<sup>2</sup>, temp=40°C. A similar FA map (right), with all the fiber and ray cells segmented out. Increased anisotropy can be seen at the edges of the xylem.

**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. 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.



76

## Towards Real-Time and High Throughput Molecular Characterization of Microbial Deconstruction of Biomass

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

Lawrence Berkeley National Laboratory, University of California, Berkeley

**Project Goals: Our project aims to (1) develop a high throughput methodology to monitor and characterize key chemical and biological processes during microbial deconstruction of plant biomass and conversion to bio-fuels directly, and (2) apply the technology to investigate enzyme-microbe synergy during biomass deconstruction.**

Synchrotron radiation-based Fourier transform infrared (SR-FTIR) spectromicroscopy is a label-free, non-invasive molecular technique that couples the high brightness of synchrotron radiation with the high throughput and vast analytical capabilities of FTIR spectrometers. One key challenge to our effort to apply SR-FTIR spectromicroscopy to bioenergy research has been the richness of infrared spectral features arising from the complex systems of lignocellulose biomass, the biomass degraders, and their interactions. More specifically, we are facing an extremely large amount of data from the multidimensional hyperspectral imaging, where the number of data points is on the order of billions. A suite of highly efficient methods is needed to systematically identify the evolving patterns and features of different chemical components. Furthermore, the dynamics of the infrared signatures of many important biochemical molecules during the deconstruction and conversion of plant biomass to biofuels have yet to be characterized fully. To improve confidence in interpretation of spectral data, we are conducting well-controlled experiments using model compounds and organisms. In parallel, we utilize quantum chemical simulation and modeling to guide our analysis of infrared measurements. During the first phase of our effort, we investigated model systems important to biofuel production, such as *Clostridium* actions on bioenergy crops. We also compared the biomass deconstruction processes between the *Clostridium* cellulosomes and the living *Clostridium* cells. A comparison between the bacteria/plant biomass and cellulosomes/plant biomass systems reveals direct molecular evidence of microbe-enzyme synergy. We found that the high throughput SR-FTIR technology under development at the Lawrence Berkeley National Laboratory can potentially aid in the optimization of microbes for an accelerated bioprocessing of plant biomass into biofuels in the future.

77

## Proteomic Measurements of Redox Sensing and Post-Translational Modifications in Cyanobacteria

T.C. Squier<sup>1\*</sup> (thomas.squier@pnl.gov), D.J. Bigelow,<sup>1</sup> R.N. Brown,<sup>1</sup> D.A. Bryant,<sup>2</sup> B. Chen,<sup>1</sup> S. Datta,<sup>1</sup> N. Fu,<sup>1</sup> A.E. Konopka,<sup>1</sup> M.S. Lipton,<sup>1</sup> W. Qian,<sup>1</sup> S. Stolyar,<sup>1</sup> D. Su,<sup>1</sup> A.T. Wright,<sup>1</sup> and **A.S. Beliaev<sup>1</sup>**

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

**Project Goals: The PNNL Biofuels Scientific Focus Area (BSFA) will conduct fundamental research on cyanobacteria with specific emphasis on pathways of carbon, nitrogen, and redox metabolism that consume reductant and conserve energy produced by photosynthetic light reactions. Consistent with DOE BER Genomic Sciences Program goals, the long-term objective of the BSFA is to develop a predictive understanding of metabolic subsystems and regulatory networks involved in solar energy conversion to biofuel precursors or products. Toward this goal, we will interrogate fluxes through metabolic subsystems that include (i) photosynthetic energy conservation and reductant generation, (ii) CO<sub>2</sub> accumulation, fixation, and reduction, (iii) biosynthesis of metabolic intermediates and monomers, and (iv) macromolecular synthesis. Our studies will exploit two complementary model cyanobacterial systems, *Cyanothece* sp. strain ATCC 51142 and *Synechococcus* sp. strain PCC 7002, and address discrete scientific hypotheses.**

As part of PNNL's Scientific Focus Area "Genome-Enabled Studies of Photosynthetic Microorganisms for Bioenergy Applications" this project is applying and developing approaches for live cell imaging that in conjunction with advanced separations, mass spectrometry and traditional biochemical techniques are being used to identify redox sensors of environmental change and post-translational protein modifications associated with intracellular signaling. Specifically, we seek to identify sequence-specific motifs within individual protein sensors in *Synechococcus* PCC 7002 that are targets for post-translational modification (PTM). We are investigating the central hypothesis that site-specific PTMs modulate cellular pathways in response to environmental change, and that their identification is critical to an ability to construct predictive models of metabolism needed for bioenergy applications. Investigated PTMs include glycosylation, phosphorylation, and redox-dependent alterations involving cysteine (e.g., reversible dithiol formation, glutathione adduct formation, and oxidation to form sulfenic acid). Using existing and newly developed affinity-based methods, we have fractionated cells and using advanced separations and mass spectrometry identified the following targeted PTMs in cellular proteins. All modified peptides are analyzed with tandem mass spectrometry and identified

with SEQUEST using the mass of the phosphoryl or glycosyl group as the modification on the specific target amino acids.

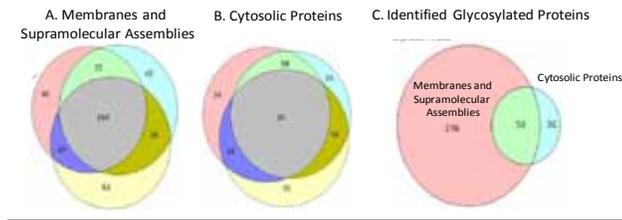


Figure 1. Identification of Glycosylated Proteins. 406 glycosylated proteins in *Synechococcus* 7002. Identified by LC-MS/MS following enrichment using a lectin column. Reproducibility (A, B) of sample identification is shown for (A) membrane and (B) cytosolic protein fractions. (C) Summary of identified glycol modifications identified from three replicate measurements.

**Glycosylation:** To identify how enzyme-mediated glycosylation may regulate metabolism, we have used cell fractionation and lectin affinity chromatography to purify glycosylated proteins from *Synechococcus* 7002 cell lysates. Isolated glycoproteins were digested with trypsin followed by high-resolution LC-MS/MS. A total of 330 proteins (containing  $\geq 2$  confidently identified peptides) were identified in the membrane fraction, whereas 130 proteins were found in the cytosolic fraction, with very high reproducibility from three separate replicates (Figure 1). Identified proteins include well studied examples of known glycosylated proteins, including the S-layer like proteins (SYNPCC7002\_A2813, SYNPCC7002\_A1634, and SYNPCC7002\_A1178). A total of 406 different proteins were identified (Figure 1C). Identification of the majority of glycosylated proteins within the membrane fraction is consistent with the hypothesized functional role of glycosyl-transferase proteins in regulating protein localization and cellular function. Identified proteins are being mapped onto their respective cellular pathways, with the goal of correlating changes in glycosylation patterns with environmentally-induced changes in cellular function.

**Phosphorylation:** Phosphoester linkages on serine, threonine, and tyrosine protein side-chains are known to regulate cellular metabolism and intracellular signaling, and are routinely identified at PNNL using metal affinity chromatography (IMAC) enrichment. To characterize phosphorylation in *Synechococcus* 7002 and understand their relationship to biomass production, we implemented an automatic IMAC system for high-throughput phosphopeptides enrichment. In this system, samples are being enriched on a normal flow “large IMAC” column with optimized binding using an autosampler controlled by in-house developed software. The system allows for large quantities of starting material and effectively removes nonspecific-bound peptides using optimized washing protocol. Initial testing of the new automated platform was carried out for *E. coli* phosphoproteome measurements and we achieved an average absolute mass accuracy of 0.9 ppm. Out of 960 total detected phosphopeptides, we identified 147 unique peptides from 119 *E. coli* proteins which included 31 unique phosphopeptides previously identified in the literature. An increased coverage of putative phosphopeptide species provided by the automated

system will be critical to our ability to characterize phosphorylation patterns in *Synechococcus* 7002, and, subsequently our understanding of regulatory mechanisms involved in central carbon and energy metabolism.

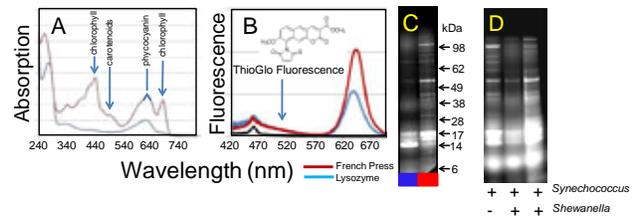


Figure 2. Identification of Thiol-Reactive Proteins within Thylakoids of *Synechococcus*. (A) Absorption spectra, (B) fluorescence emission spectra, and (C) SDS-PAGE showing thiol reactive proteins for lysates from *Synechococcus* before (blue) and after (red) enrichment for thylakoid membranes.

**Cysteine Modification:** We hypothesize that dithiol “switches” (e.g., Cys-Xxx<sub>n</sub>-Cys, n < 6) activated by thioredoxin-dependent pathways regulate electron flux and carbon sequestration to modulate energy partitioning in *Synechococcus* PCC 7002. In addition, modifiers of redox-dependent regulation are likely to include reversible modifications of cysteine previously identified to regulate intracellular metabolism, including S-nitrosylation, glutathiolation, and the formation of sulfenic acid. Identification of redox-dependent dithiols involves two complementary approaches. Using newly synthesized chemical probes, visualization, down-stream purification, and subsequent LC-MS/MS identification of all thiol-reactive proteins are facilitated. Initial results indicate that the majority of thiol-reactive proteins in *Synechococcus* 7002 are located within the thylakoid membrane (Figure 2). These results are consistent with expectations based on homologies with higher plants, where redox-dependent dithiol formation is linked to the regulation of energy metabolism. Identification of these thiol reactive proteins will provide important insights regarding possible relationships between environmental stress, cellular metabolism, and cell growth. Specifically, we hypothesize that modulation of these redox dependent control mechanisms is critical to the maintenance of cell function and will minimize oxidative stress related to redox imbalance. Complementary in vivo approaches to monitor redox-active dithiol reactivity will use recently synthesized cell-permeable monoarsenic probes that selectively bind to reduced disulfides in close proximity to trap available dithiols in living cells prior to cell lysis. The identification of proteins in *Synechococcus* 7002 that undergo disulfide exchange in response to changes in cellular conditions (e.g., light intensity, nutrient availability) involve incubation of these new cell-permeable reagents. Following reaction with proximal thiols, cells are lysed and following reduction of internal disulfides with TCEP, the bound probe will be released permitting capture and identification of redox active proteins by commonly used thiol-capture affinity methods operational at PNNL. Parallel imaging measurements will assess how environmental conditions affect cellular redox

state monitored as reduced dithiols both in model systems and isolated natural microbial communities to facilitate our understanding of possible relationships between environmental changes, cellular redox state, and intracellular metabolism.

## 78

### Comparative Molecular Imaging Analysis of *Brachypodium distachyon* and Its Mutants

Liang Chen,<sup>1\*</sup> Ludmila Tyler,<sup>2,3</sup> Andrew N. Tauscher,<sup>1</sup> John Vogel,<sup>2</sup> and Hoi-Ying N. Holman<sup>1</sup> (hyholman@lbl.gov)

<sup>1</sup>Lawrence Berkeley National Laboratory, Berkeley, Calif.; <sup>2</sup>USDA-ARS Western Regional Research Center, Albany, Calif.; <sup>3</sup>University of California Berkeley, Albany

**Project Goals: Our project aims to (1) develop and apply Synchrotron-Radiation-based FTIR imaging and analysis techniques for high spatial resolution molecular characterization of differences among *Brachypodium distachyon* lines, and (2) develop and provide a high-throughput approach that can help link chemical information to gene functions in *B. distachyon*.**

The grass *Brachypodium distachyon* (*Brachypodium*) is an important model system for bioenergy grasses (e.g. switchgrass and *Miscanthus*) owing to its small genome size (272Mbp), compact physical stature, short lifecycle, self-fertility and transformability. To date, ~10,000 T-DNA insertional mutant lines have been generated for functional genomics studies. The ability to rapidly compare compositions, structures and functions of different *Brachypodium* lines and their mutants is critical to establish causal links between gene functions and chemical composition. Synchrotron-Radiation-based FTIR (SR-FTIR) can play an important role in determining this information because SR-FTIR is a nondestructive label-free spectroscopic method that derives chemical information directly from characteristic vibrational frequencies of various chemical bonds with diffraction-limited spatial resolution. The fast speed and non-invasive nature of this analytical method make it especially suitable for development into a high-throughput chemical analysis tool for *Brachypodium* studies. As investigators of the Berkeley Synchrotron Infrared Structural Biology (BSISB) program, we aim to develop and apply the SR-FTIR spectro-imaging technique together with multivariate statistics to the rapid chemical analysis of the *Brachypodium* cell wall and the comparative analysis of different *Brachypodium* lines and their mutants.

During the first phase of our technology research and development effort, we used six genetically diverse *Brachypodium* lines selected from a larger collection of 187 natural accessions (Figure 1). As a first step, we performed bulk FTIR measurement on ground stems from the six lines. With Principal Component Analysis-Linear Discriminant Analysis (PCA-LDA) software tools specifically developed

for the *Brachypodium* project, we found that the lines could be clearly separated into two groups by their mid-infrared signatures (Figure 2), suggesting underlying compositional and/or structural differences among the lines. Based on this promising initial result, we then used SR-FTIR to obtain high-resolution infrared images of thin sections of these six *Brachypodium* lines. This study reveals the spatial heterogeneity in the chemical composition of the plants. Using comparative image analysis methods, we found that there are significant differences in the spatial distribution of cell wall composition between the six *Brachypodium* lines. The genetic basis of these phenotypic differences will be evaluated by making crosses between lines with different compositions/patterns and identifying the genes responsible using a map-based approach. With our optimized methods we also plan to examine the compositional changes in *Brachypodium* cell wall mutants identified by near infrared spectroscopy in the Vogel lab. Our methods will provide the foundation for future studies linking chemical information to gene functions in *Brachypodium*. Our long-term goal is to standardize and enable full automation of our analytical system for high through-put screening experiments for multiple natural and engineered energy crops.

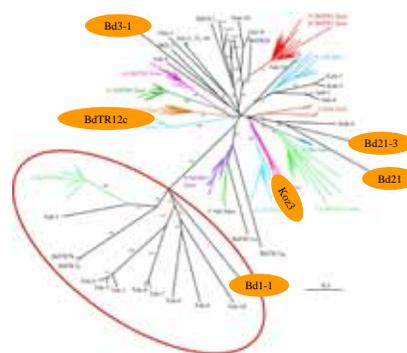


Figure 1. Genotypic diversity in select inbred lines. An unrooted Neighbor-Joining consensus tree of 187 lines based on 100 shared allele bootstrap trees constructed using 43 SSR markers. The lines in the red circle share a number of phenotypic characters (small seeds, nearly hairless lemmas, long vernalization requirements). The lines used in this study are highlighted by orange ovals. Note that Bd21 is the line sequenced for the reference genome. The other five lines used in this study were recently resequenced through the DOE JGI Community Sequencing Program.

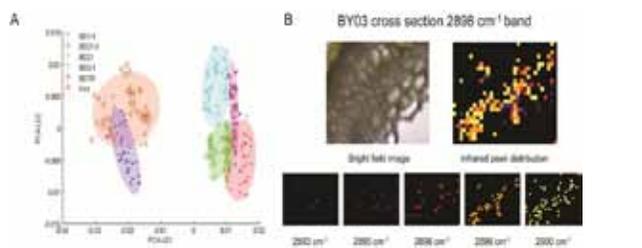


Figure 2. Chemical analysis of *Brachypodium* samples using mid-IR. (A). The PCA-LDA score plot of samples from 6 *Brachypodium* lines. 45 spectra are collected from each line. Spectra data are collected from 600  $\text{cm}^{-1}$  to 4000  $\text{cm}^{-1}$ . The spectra in the region between 800  $\text{cm}^{-1}$  and 1200  $\text{cm}^{-1}$  are used as input for PCA-LDA analysis. The "Bd1-1" and "Koz-3" samples are clearly separated from the other 4 types of samples. (B). The distribution and shift of the 2898  $\text{cm}^{-1}$  band in the cross section of a *Brachypodium* sample thin slice. Each peak position is represented by a different color.

## 79

### Neutrons Reveal Structural Changes During Switchgrass and Poplar Biomass Pretreatment

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

<sup>1</sup>Oak Ridge National Laboratory, Oak Ridge, Tenn.; and <sup>2</sup>Institute of Paper Science and Technology, Georgia Institute of Technology, Atlanta

Program: SFA Biofuels

**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. This Scientific Focus Area in Biofuels seeks to develop and demonstrate the "Dynamic Visualization of Lignocellulose Degradation by Integration of Neutron Scattering Imaging and Computer Simulation" for multiple-length scale, real-time imaging of biomass during pretreatment and enzymatic hydrolysis. This will provide fundamental information about the structure and deconstruction of plant cell walls that is needed to drive improvements in the conversion of renewable lignocellulosic biomass to biofuels.**

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

Here, we present results from the dilute acid pretreatment method used to break down switchgrass and poplar biomass. Results are from both ex-situ experiments as well as a new in-situ reaction cell which allows direct SANS monitoring of changes during pretreatment.

From our ex-situ studies, the similar, basic trends in structural and morphological changes in switchgrass and poplar biomass during dilute acid pretreatment are: (1) increase in the small-scale structure which can be related to the crystalline core cross-section; (2) decrease in the interconnectivity of the biopolymers; (3) formation of additional distinct structures at length scales 50-200 Å that are due to formation of lignin aggregates which first appears around 125-150°C; and (4) at length scales larger than 1000 Å, no change in the smooth domain boundaries until 60 min at 160°C. On the other hand, the inherent differences in the chemical composition, structure and morphologies of biomass grass (switchgrass) when compared to trees (poplar) cause many differences in the details of the structural changes during dilute acid pretreatment, such as: (1) larger increase in the small-scale structure than a single crystalline microfibril indicating coalescence of more neighboring crystalline microfibrils; (2) larger lignin aggregate size formation and growth; and (3) rougher surface morphology of the micron-sized cell walls.

The results from the in-situ study demonstrated the appearance of a distinct additional structure of  $R_g \sim 30$  Å after attaining 120°C. A maximum temperature of 180°C was used for this study. The size of the additional structure, after 30 min at 180°C was  $R_g \sim 700$  Å and thereafter was larger than our accessible particle-size length scales. If the additional structure is interpreted as lignin aggregates, these undergo much larger growth during the in-situ process and the first appearance of the lignin aggregates occurs at a slightly lower temperature (120°C) when compared to the ex-situ study, above 125°C.

## 80

### Integration of Small-Angle Neutron Scattering and Computational Modeling to Investigate the Structures of Lignin Aggregates and the *Trichoderma reesei* Cellobiohydrolase I

Loukas Petridis<sup>1\*</sup> (petridisl@ornl.gov), Sai Venkatesh Pingali,<sup>1</sup> Volker Urban,<sup>1</sup> William T. Heller,<sup>1</sup> Jeremy C. Smith,<sup>1</sup> Hugh M. O'Neill,<sup>1</sup> Marcus Foston,<sup>2</sup> Dean A. Myles,<sup>1</sup> Art Ragauskas,<sup>2</sup> Barbara R. Evans,<sup>1</sup> and **Brian H. Davison**<sup>1</sup>

<sup>1</sup>Oak Ridge National Laboratory, Oak Ridge, Tennessee; and <sup>2</sup>Institute of Paper Science and Technology, Georgia Institute of Technology, Atlanta

Program: SFA Biofuels

**Project Goals: Lignocellulosic biomass comprises the vast majority of biomass on Earth and has the potential to play**

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

Computational modeling and small-angle neutron scattering (SANS) were used to examine the molecular architecture of lignin across a broad range of length scales. Similarly, this combined approach was used to understand the solution structure of cellobiohydrolase I from *Trichoderma reesei*, which breaks down cellulose into cellobiose, and how the enzyme responds to changes in pH.

Here, we report on two studies using this combined approach to better understand lignin and the cellobiohydrolase I (Cel7A), the major cellulase enzyme component from the fungus *T. reesei*. Lignin, a major polymeric component of plant cell walls, forms aggregates *in vivo* after pretreatment of lignocellulosic biomass for ethanol production. The aggregates are thought to reduce ethanol yields by inhibiting enzymatic hydrolysis of cellulose. Extensive atomistic molecular dynamics (MD) simulations using models built with information provided by chemical analysis, <sup>13</sup>C- and <sup>1</sup>H/<sup>2</sup>H- NMR studies and SANS were used to demonstrate that the surfaces of the aggregates are highly irregular and characterized by a surface fractal dimension that is invariant across length scales from ~1–1000Å. The simulations reveal extensive water penetration of the aggregates and heterogeneous chain dynamics corresponding to a rigid core with a fluid surface. The detailed multiscale structure revealed here should aid in understanding biomass recalcitrance to hydrolysis and in feedstock engineering efforts to improve biofuel yield. The solution structure of Cel7A was studied as a function of pH using SANS and computational modeling. The enzyme, which consists of a large catalytic core connected to a much smaller cellulose binding domain by a flexible linker, has a pH optimum of 4.5. The SANS data demonstrate that a conformational change takes place as the pH is lowered from 7.0 to 4.2. The spatial relationship between the cellulose binding domain and the catalytic core changes between pH 7.0 and 5.3. At pH 4.2, the enzyme transitions to a flexible conformation that is intermediate to a tightly-folded structure and a disordered state. The results suggest that the increased flexibility observed in the structure of the enzyme is necessary for optimum activity of the enzyme. The insight provided by this work could aid in the engineering of enzymes optimized for converting lignocellulosic biomass into biofuels.

80A‡

submitted post-press

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

Brian G. Saar,<sup>1</sup> Yining Zeng,<sup>2</sup> Christian W. Freudiger,<sup>1</sup> Yu-San Liu,<sup>2</sup> Michael E. Himmel,<sup>2</sup> **Shi-You Ding**<sup>2\*</sup> (Co-PI: Shi.you.Ding@nrel.gov), and **X. Sunney Xie**<sup>1\*</sup> (PI: xie@chemistry.harvard.edu)

<sup>1</sup>Department of Chemistry and Chemical Biology, Harvard University and <sup>2</sup>Biosciences Center, BioEnergy Science Center, National Renewable Energy Laboratory

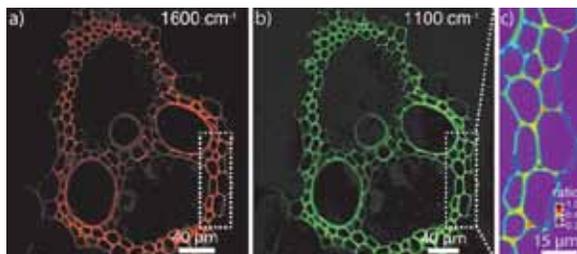
**Project Goals:** The objective of this project was to capitalize on and to further develop Coherent Raman Scattering (CRS) microscopy in order to characterize and image lignin and polysaccharides in plant cell walls, and to monitor plant cell wall deconstruction in real time in biomass conversion processes from a variety of agricultural products. The specific aims were (1) construction of a new generation of CRS microscope with frequency modulation that offers much improved sensitivity in order to selectively image cell wall constituents in plant cell walls. (2) Use of the new microscope to monitor and study the dynamic structural and compositional changes of plant cell wall macromolecules during their chemical and enzymatic degradation *in situ*. (3) Development of a near-field CRS microscope with nanometer spatial resolution in order to probe the structure and dynamics of plant cell wall processing.

We demonstrate that stimulated Raman scattering (SRS) microscopy<sup>1,2</sup>, a new imaging method, allows real time observation of biomass conversion processes. Current analytical methods, such as gas chromatography–mass spectrometry, electron or scanning-probe microscopy, and fluorescence microscopy, have restrictions. Microscopy based on infrared absorption offers chemical specificity, but spatial resolution is limited by long infrared wavelengths, and penetration depth into aqueous plant samples is small. Raman microspectroscopy offers label-free chemical contrast with high resolution and chemical specificity. However, the Raman scattering effect is weak, and long pixel dwell times are required for imaging plant materials. This means that real-time imaging is challenging, requiring a long data collection time. SRS microscopy offers chemical contrast based on the intrinsic Raman vibrational frequencies in a sample with much shorter imaging time and easier spectral identification. SRS superseded CARS microscopy<sup>3</sup>, another non-linear vibrational imaging method our group has developed in the past. The SRS imaging technique for studying the conversion process *in situ* offers chemical specificity without exogenous labels, non-invasiveness, high spatial resolution, and real-time monitoring capability. We demonstrate its utility for the study of biomass conversion.<sup>4</sup>

## References

## (publications supported by this project)

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



Imaging of lignin and cellulose with SRS microscopy. a) SRS image at  $1600\text{ cm}^{-1}$  of a vascular bundle, showing the lignin distribution with a red intensity grade. b) SRS image of the same vascular bundle as in (a), showing the cellulose distribution at  $1100\text{ cm}^{-1}$  with a green intensity grade. Both (a) and (b) were obtained with a 50 ms pixel dwell time. These images can be acquired simultaneously using the two-color SRS instrument. c) Ratio of the lignin divided by the cellulose signal at higher magnification, obtained from the region surrounded by the dotted line in (a) and (b).

## Systems Biology and Metabolic Engineering Approaches for Biological Hydrogen Production

# 81

### A Systems Biology Approach to Energy Flow in $\text{H}_2$ -Producing Microbial Communities: Multi-Scale Measurement of Metabolic Function in Complex Microbial Ecosystems

Xavier Mayali<sup>1\*</sup> (mayali1@llnl.gov), Kevin J. Carpenter<sup>1\*</sup> (carpenter37@llnl.gov), Javier Ceja-Navarro<sup>2\*</sup> (JNavarro@lbl.gov), Peter K. Weber,<sup>1</sup> Shalini Mabery,<sup>1</sup> Paul D. Hoepflich,<sup>1</sup> Ulas Karaoz,<sup>2</sup> Todd Z. DeSantis,<sup>2</sup> Meredith M. Blackwell,<sup>3</sup> Stephanie Gross,<sup>3</sup> Eoin L. Brodie,<sup>2</sup> and Jennifer Pett-Ridge<sup>1\*</sup>

<sup>1</sup>Chemical Sciences Division, Lawrence Livermore National Laboratory, Livermore, Calif.; <sup>2</sup>Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, Calif.; and <sup>3</sup>Department of Biological Sciences, Louisiana State University, Baton Rouge

**Project Goals:** The objective of this research is to develop an integrated analysis of energy flow in complex microbial communities for the purpose of optimizing biofuel production. We will combine biogeochemical, stable isotope probing, metatranscriptomic and computational approaches, with the aim of understanding nutrient cycling and biofuel (specifically,  $\text{H}_2$ ) production in microbial communities. Our ultimate goal is the development of multi-scale models that can predict ecological and biochemical relationships within multi-trophic microbial systems.

A fundamental goal of microbial ecology is to understand the biogeochemical role of individual microbial taxa in their natural habitat. This is a complex problem because most microbes remain uncultivated and most microbial communities are very diverse. Thus direct analysis of isolated strains results in a skewed and possibly non-representative ex-situ view of functional capabilities while culture-independent (metagenomic) approaches mostly provide hypotheses as to which of the many biogeochemical processes organisms or communities perform.

To directly measure functional roles of uncultivated microbes, we have developed a multi-scale approach. Beginning at the community level, we probe for organisms capable of assimilating substrates in pathways of interest using an approach termed "chip-SIP", a high-sensitivity, high-throughput stable isotope probing (SIP) method performed on a phylogenetic microarray. We incubate microbial communities with isotopically ( $^{13}\text{C}/^{15}\text{N}$ ) labeled substrates, hybridize community rRNA to a microarray, and measure isotope incorporation—and therefore assimilation—by

secondary ion mass spectrometer imaging (NanoSIMS). Using this approach we quantified amino acid, nucleic acid and fatty acid incorporation by 81 taxa in a marine microbial community. These data enabled the visualization of carbon flow that constrains the functional roles of specific microbial taxa (e.g. substrate generalists vs. specialists) (Fig. 1A).

Our next approach is to directly image substrate assimilation by individual microbial cells to identify the location of activity within a natural system. For this aspect we are currently focusing on protist-bacterial symbioses within the hindgut of lower termites (Fig. 1B). Protists are an extremely diverse, abundant, and important group of microbes with capabilities of interest in alternative energy and climate amongst others. For example, protists of the lower termite hindguts convert lignocellulose into simple carbohydrates while also producing hydrogen gas. Many protist species have tens or hundreds of thousands of bacterial symbionts whose function is poorly understood or unknown. Here we report the first results in exploring these protist-bacterial symbioses using in situ stable isotopic labeling ( $^{13}\text{C}$  enriched feeding substrates and  $^{15}\text{N}$  enriched air) followed by analysis of individual protists and their symbiotic bacteria with NanoSIMS.

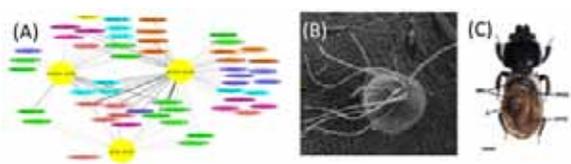


Figure 1. (A) Network diagram showing the uptake and assimilation of nucleic acids, amino acids and fatty acids by a community of microorganisms. (B) Lower termite hindgut protist. (C) Wood-feeding passalid beetle digestive tract.

Our third complementary approach is the analysis of microbial community gene expression in conjunction with micron-scale measurements of physicochemical gradients to provide fundamental physiological context. For this aspect we also focus on wood-ingesting insects (Passalid beetles – Fig. 1C) and the metabolic partitioning through the insect's digestive tract to perform efficient lignocellulose metabolism that results in the production of methane, hydrogen and other potential biofuels. We have adapted mRNA enrichment procedures to deal with the abundance of bacterial, archaeal and eukaryal rRNA prior to meta-transcriptome sequencing. Gene expression data are then used to inform focused isotope-labeling experiments.

Our goal is to combine these approaches into a workflow across the same microbial system to yield a comprehensive model of the flow of energy in these systems.

## 82

### Hydrogen Producers and Consumers in Photosynthetic Microbial Mats Identified by Combining Biogeochemical and Molecular Analysis with Single Cell Techniques

Luke Burow,<sup>1,2</sup> Dagmar Woebken,<sup>1,2</sup> Brad M. Bebout,<sup>2</sup> Lee Prufert-Bebout,<sup>2</sup> Tori M. Hoehler,<sup>2</sup> Steven W. Singer,<sup>3</sup> Alfred M. Spormann,<sup>1</sup> Jennifer Pett-Ridge,<sup>4</sup> and Peter K. Weber<sup>4\*</sup> (weber21@lbl.gov)

<sup>1</sup>Department of Civil and Environmental Engineering, Stanford University, Stanford, Calif.; <sup>2</sup>Exobiology Branch, NASA Ames Research Center, Moffett Field, Calif.; <sup>3</sup>Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, Calif.; and <sup>4</sup>Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory, Livermore, Calif.

**Project Goals:** We are developing a new method that will provide correlated oligonucleotide, functional enzyme, and metabolic image data, thereby linking identity to function in situ in complex microbial communities.

Oligonucleotide and functional enzyme labels, along with biomass labeled by incorporation of stable isotope tracers, will be imaged at the single cell level by nanometer-scale secondary ion mass spectrometry (NanoSIMS). This method, NanoSIP, will improve on the success of previous stable isotope probing (SIP) methodologies in linking microbial identity and function by preserving critical spatial information and allowing finer temporal resolution than previous approaches. NanoSIP will allow us to place coupled genetic and functional information within the context of the physical structure, chemical and physical microgradients, and microbial assemblages of complex communities. NanoSIP will be broadly relevant to DOE's missions in bioenergy, carbon cycling and bioremediation. We will develop our method using *Microcoleus* and *Lynxgryba* cyanobacterial mats, which are relevant to bioenergy because they are ecological counterparts to the cyanobacterial systems considered for biohydrogen production.

Hypersaline, photosynthetic microbial mats are diverse communities of microorganisms that can produce hydrogen ( $\text{H}_2$ ). The hydrogen ecology of these mats provides a challenging test bed for single cell methods and has the potential to provide insights into microbial interactions relevant to industrial-scale biofuels production. Our investigations of microbial mats to date have focused on characterizing the controls on hydrogen production and consumption, including nitrogen fixation, using a combination of biogeochemical and molecular techniques along with high-resolution imaging mass spectrometry by NanoSIMS. In our poster, we will present data for a microbial mat from Elkhorn Slough, CA that is dominated by *Microcoleus chthonoplastes*. Our data indicate that *M. chthonoplastes* is the primary  $\text{H}_2$  producer, and that sulfate reducing bacteria (SRB) are the primary  $\text{H}_2$  consumers. The dominant cyanobacterial nitrogen fixers are

members of a novel clade (UD3), which show significant variability in fixation rates at the single cell level (Fig. 1).

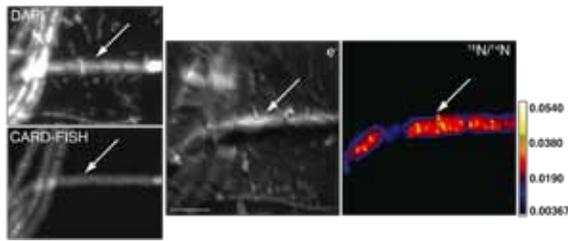


Figure 1. Correlated fluorescence (DAPI and CARD-FISH) and NanoSIMS (secondary e<sup>-</sup> and <sup>15</sup>N/<sup>14</sup>N) images of a microbial mat sample incubated with <sup>15</sup>N<sub>2</sub> and hybridized with a CARD-FISH probe targeting the 16S rRNA of the proposed cyanobacterial clade UD3. The arrow points to a UD3 filament that had a high rate of nitrogen fixation relative to the adjacent UD3 filaments and other microbes (DAPI) based on its enrichment in <sup>15</sup>N (<sup>15</sup>N/<sup>14</sup>N ~0.03). NanoSIMS analysis showed significant variability in N-fixation rates.

Nitrogen fixation has been hypothesized to be a major driver of hydrogen production in microbial mats because of the strong correlation between N<sub>2</sub>-fixation and hydrogen production in field samples. We performed manipulation experiments to differentiate among potential mechanisms of H<sub>2</sub> production. Suppression of N<sub>2</sub>-fixation by addition of excess ammonium to mats did not change the amount of H<sub>2</sub> or organic acids produced, demonstrating that H<sub>2</sub> production occurs independently of N<sub>2</sub>-fixation and likely via fermentation. Depriving mats of daytime sunlight resulted in a ~20-fold decrease in the subsequent nighttime H<sub>2</sub> production, suggesting that photosynthate is the energy source for H<sub>2</sub>-producers. Homogenization of the mat to disruption microbial associations increased H<sub>2</sub> production ~2.5-fold, indicating that physical proximity of H<sub>2</sub>-consumers to H<sub>2</sub>-producers is an important factor in H<sub>2</sub> consumption. Addition of sodium molybdate to inhibit sulfate-reducing bacteria (SRB) also increased net H<sub>2</sub> production ~2.5-fold, implying that SRBs are important H<sub>2</sub>-consumers. Pyrosequencing of small subunit rDNA of the upper strata of nighttime samples revealed the presence of more than 60 phyla across all domains of life (Fig. 2A). Analysis of rRNA:rDNA pyrotags, although not definitive, suggest that OTUs affiliated with Cyanobacteria (including Oscillatoriales) and Chloroflexi (including Chloroflexaceae) are the dominant active microorganisms (Fig 2B). H<sub>2</sub>-evolving NiFe H<sub>2</sub>-ases associated with *M. chthonoplastes* dominated clone libraries derived from the upper strata using degenerate primer sets that targeted a diversity of phototrophic organisms. Tiled H<sub>2</sub>-ase microarray data (~44K probes targeting ~1.2K H<sub>2</sub>-ases) show that the *M. chthonoplastes* H<sub>2</sub>-ases are up-regulated 3-fold at night compared to the day.

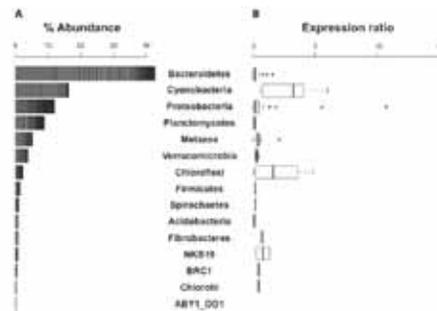


Figure 2. (A) More than 63 phyla across all domains of life were present in the upper 2 mm strata as determined by rDNA sequencing (15 most abundant phyla shown). (B) OTUs belonging to Cyanobacteria and Chloroflexi had relatively higher rRNA:rDNA ratios (expression ratios), suggesting that organisms in these phyla are more active relative to organisms in other phyla present in the mats.

While our manipulations indicate that N-fixation is not the direct driver of H<sub>2</sub> production, the possibility still remained that H<sub>2</sub> production and N-fixation are performed by the same dominant microbe, *M. chthonoplastes*. However, <sup>13</sup>C-NaHCO<sub>3</sub> and <sup>15</sup>N<sub>2</sub> incubations of mat samples followed by high-resolution secondary ion mass spectrometry by NanoSIMS showed that *M. chthonoplastes* has high rates of C-fixation but no detectable N-fixation. Sequencing of the expressed genes encoding for the nitrogenase reductase (*nifH*) showed that 36% of the sequences belonged to a previously unknown cyanobacterium. An enrichment culture of this cyanobacterium (UD3) allowed us to link this *nifH* sequence clade to a 16S rRNA sequence, and 16S rRNA sequence analysis showed that UD3 is only distantly related to any sequence in the databases (max. 92% sequence identity). Catalyzed reporter deposition fluorescence *in situ* hybridization (CARD-FISH) with probes specific for this novel cyanobacterium and subsequent NanoSIMS analyses showed that 69% of the UD3 related population in the microbial mats incorporated <sup>15</sup>N<sub>2</sub> (Fig. 1).

## 83 Developing Techniques for Bioenergy Research Using Plant and Microbial Cultures

Catherine Lacayo<sup>1\*</sup> (lacayo2@llnl.gov), Yongqin Jiao<sup>1\*</sup> (jiao1@llnl.gov), Ali Navid,<sup>1</sup> Ben Stewart,<sup>1</sup> Mona Hwang,<sup>1</sup> Shi-You Ding,<sup>2</sup> Alexander Malkin,<sup>1</sup> Jennifer Pett-Ridge,<sup>1</sup> and **Michael P. Thelen**<sup>1</sup>

Lawrence Livermore National Laboratory, Livermore, Calif.; and <sup>2</sup>National Renewable Energy Laboratory, Golden, Colo.

**Project Goals:** Our research aims to develop an integrated analysis of energy flow in complex microbial communities. We are combining biogeochemical, stable isotope probing, metatranscriptomic and computational approaches, to understand nutrient cycling and biofuel (H<sub>2</sub>) produc-

tion production in complex microbial communities. A comprehensive understanding of such communities is needed to develop efficient, industrial-scale processes for microbial H<sub>2</sub> production and lignocellulose degradation. Experimentally, we are focusing on natural models for H<sub>2</sub> production and degradation of plant/microbial biomass: H<sub>2</sub>-producing cyanobacterial mats and insect hindgut communities, and polysaccharide biogenesis and deconstruction in plant cells and microbial mats. Our ultimate goal is the development of multi-scale models that can predict ecological and biochemical relationships within multi-trophic microbial systems. Our work includes both technique development and biological research components.

To gain a better understanding of the mechanisms of plant cell wall breakdown, and to accelerate efforts to convert lignocellulosic materials to biofuels, we examined two distinct elements involving degradation followed by fermentation. To visualize deconstruction of natural plant lignocellulose, we monitored the enzymatic degradation of the wall of wood cells in culture. For biofuel production, we investigated the interaction of two disparate bacteria in a co-culture system that degrades cellulose and produces hydrogen gas.

To monitor the response of plant cell walls to chemical, enzymatic and microbial degradation, we are imaging single cultured wood cells from the plant *Zinnia elegans* using both optical and topographical methods. We have evaluated the effects of exogenous enzymes, such as fungal cellulases and hemicellulases, on the cell wall as a function of time. Loss of cellulose in the wall was measured using fluorescent probes, such as carbohydrate-binding modules (CBMs) from *Clostridium thermocellum*. Using atomic force microscopy, we observed structural changes consistent with the gradual removal of cellulose fibrils from the wall.

As a consolidated bioprocessing approach, cellulose-degrading organisms are often co-cultured with organic acid-consuming microbes. However, the metabolic interactions between the organisms are poorly understood. In a co-culture containing *Clostridium cellulolyticum* and *Rhodospirillum rubrum*, we analyzed the kinetics of cellular growth as well as carbon fluxes in both organisms using a constraint-based metabolic model. We found that acetate, lactate, ethanol and pyruvate are likely the major metabolites involved in carbon transfer, with pyruvate being the most inhibitory for the growth of *C. cellulolyticum*. Model-based analyses of *R. rubrum* examined the effect of pyruvate and suggested possible ways to improve pyruvate consumption.

LLNL-ABS-469072. This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

## 84

### Altered Sequences and Modifications to Proteins from Organisms of Interest to the Environment and Energy Production

Rachel R. Ogorzalek Loo<sup>1,2\*</sup> (rloo@mednet.ucla.edu), Hong Nguyen,<sup>1,3</sup> Deborah Francoleon,<sup>1,3</sup> Housna Mouttaki,<sup>5</sup> Joseph A. Loo,<sup>1,3</sup> Lars Rohlin,<sup>1,4</sup> Robert Gunsalus,<sup>1,4</sup> Sabeeha Merchant,<sup>1,3</sup> and Michael McInerney<sup>5</sup>

<sup>1</sup>UCLA-DOE Institute for Genomics and Proteomics; <sup>2</sup>Department of Biological Chemistry, <sup>3</sup>Department of Chemistry and Biochemistry, <sup>4</sup>Department of Microbiology, Immunology, and Molecular Genetics, University of California-Los Angeles; and <sup>5</sup>Department of Botany and Microbiology, University of Oklahoma  
<http://www.doe-mbi.ucla.edu/overview>

#### Project Goals:

- Define *Chlamydomonas reinhardtii* proteomics
- Characterize proteome of *Syntrophus aciditrophicus*
- Elucidate surface glycoproteins of methanogens and link to morphology

#### Introduction

Living organisms transcribe and translate genetic information many ways and often embellish the products co- and post-translationally. The average error rate for translating proteins in normally growing mammalian and bacterial cells, 3x10<sup>-4</sup> misreads per codon, was determined by elegant radiolabeling experiments performed 2-3 decades ago. Related experiments established that the error rate increased with exposure to high Mg<sup>2+</sup> concentrations, low temperature, high pH, or various organic solvents. Starvation for certain amino acids in animal and bacterial cells elevates the level of translation errors to one readily visualized as a trail of spots on 2-D gels, an effect referred to as “stuttering”. An associated phenomenon, recoding, describes the local reprogramming of mRNA translation to discard standard translational rules and decode non-canonically. The products of this stimulated process, proteins incorporating shifts in reading frame, bypassed nucleotide segments, and/or unexpected amino acids have been observed in all 3 kingdoms and are known to contribute importantly to posttranscriptional regulation of gene expression. Environmental organisms are often challenged severely and likely to respond in these unusual ways, but the resulting proteomic changes are invisible to many experimental strategies. Proteomic methods can be tailored to reveal and characterize these unanticipated modifications and sequence alterations.

#### Methods

*Syntrophus aciditrophicus*, a gram-negative anaerobe surviving at the limit of thermodynamic feasibility (McInerney *et al.*, *PNAS* 2007, 104, 7600-7605), was cultivated anaerobically on crotonate or benzoate. Alternatively it was co-cultured with *Methanospirillum hungatei* on crotonate, benzoate, or cyclohexane carbonate. Proteins were separated by 2D gel

electrophoresis; protein-containing spots were trypsin-digested and analyzed by LC-MS/MS.

Streptavidin affinity chromatography, SDS-PAGE and near-western blotting were used to visualize and capture biotin-tagged surface-exposed proteins, which were trypsin-digested, and analyzed by LC-MS/MS. Lectin blotting suggested sugars present in glycosylated bands. Concanavalin A lectin affinity chromatography was used to enrich the S-layer protein. ConA eluates were digested with trypsin, GluC and AspN individually and double-digested with immobilized Pronase or Proteinase-K. Glycopeptides were enriched by doubled ConA affinity chromatography, ERLIC (HILIC/WAX) fractionation, C12 fractionation, and nonspecific protease treatment. Glycopeptides were measured by LC-MS/MS (QTOF) and ESI-MS/MS (ECD, IRMPD or CAD with an LTQ-FT Ultra). Normal phase and epifluorescence microscopy were used to visualize cell surface glycosylation.

## Results

In the proteome of *Syntrophus aciditrophicus* we found many examples of translational frameshifting or recoding. De novo sequencing MS/MS spectra unassigned by automated protein and 6-frame translated DNA searches revealed multiple instances where >1 sequence appeared to arise from a single mRNA transcript; e.g., 6 unpredicted variants of the peptide TATPDQAQEAHAFIR in phosphoglycerate mutase. For the predicted peptide R.AQNLHDAK, the peptides AQNLHDAK, LGAQNLHDAK, and pyroQSAQNLHDAK were observed. All together, 14 peptides did not match the phosphoglycerate mutase sequence in any one reading frame.

Standard searching of LC-MS/MS data from a digested spot containing branched chain aminotransferase assigned 27/67 spectra, but after completing our analysis, 61/67 spectra were assigned, revealing 14 peptides unmatched in any one reading frame.

*Methanosarcina mazei* and *Methanosarcina acetivorans* surface-exposed proteins and modifications were identified, employing methods specially developed for compatibility with their fragile cells. *Methanosarcina* S-layer proteins MM1976 and MA0829, do not correspond to those predicted by bioinformatics. Unusual glycosylation patterns were also recovered.

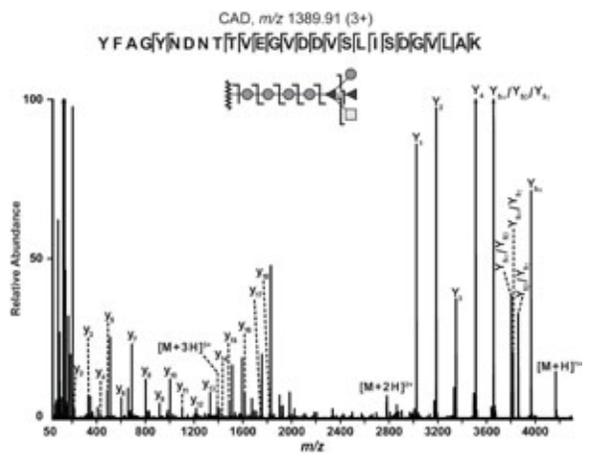


Figure 1. Mass spectrometry analysis of a glycopeptide in *Methanosarcina mazei* MM1364. Glycan symbol code: gray circle, Hexose; black triangle, Deoxyhexose; gray square, N-acetyl Hexosamine; wiggly line, peptide.

# 85

Student Oral Presentation—Monday

## Novel Hydrogen Production Systems Operative at Thermodynamic Extremes

Michael J. McInerney,<sup>1</sup> Ralph S. Tanner,<sup>1</sup> Jessica R. Sieber,<sup>1</sup> Kimberley L. Thomas,<sup>1\*</sup> Matthew W. Maune,<sup>1</sup> Gregory Hurst,<sup>2</sup> Lars Rohlin,<sup>3</sup> and Robert P. Gunsalus<sup>3\*</sup> (robgs@mimicbio.ucla.edu)

<sup>1</sup>University of Oklahoma; <sup>2</sup>Oak Ridge National Laboratory; and <sup>3</sup>University of California, Los Angeles  
<http://www.mimicbio.ucla.edu/faculty/gunsalus/>

**Project Goals:** The goals of this collaborative project are to develop and utilize new research strategies to address Genomic Science program needs in the area of biohydrogen 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 H<sub>2</sub> 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). H<sub>2</sub> production from fatty and aromatic acids requires energy input. The *S. wolfei* genome 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 growth. Interestingly, three cytoplasmic formate

dehydrogenases and one cytoplasmic hydrogenase appear to be NADH-linked because the respective gene clusters contain NADH:quinone oxidoreductases E and F subunit genes. The gene arrangement suggests that *S. wolfei*, like anaerobes known to produce high molar ratios of hydrogen from glucose, may produce H<sub>2</sub> and/or formate from NADH by an electron bifurcation mechanism. The proteome of *S. wolfei* grown syntrophically and axenically was analyzed to determine proteins involved in key redox reactions.

Whole cell-derived peptide mixtures were analyzed with two-dimensional liquid chromatography/tandem mass spectrometry (2D LC-MS-MS) via the MudPIT approach. Proteomic analysis identified ~1090 polypeptides in all three growth conditions. Thirty-four polypeptides, including a zinc-dependent dehydrogenase with a GroES domain and a putative membrane-bound hydrogenase, were detected only in syntrophically grown *S. wolfei*, e.g., with *Methanosprillum hungatei* JF1 on butyrate or crotonate. Multiple systems for interspecies electron transfer and reverse electron transport including a confurcating hydrogenase and a novel FeS oxidoreductase thought to serve as an ETF:quinone oxidoreductase were detected. Multiple beta-oxidation enzymes and several electron transfer flavoproteins were detected in each growth condition. The number of proteins involved in coenzyme and amino acid transport and metabolism increased under syntrophic conditions while the number of proteins involved in lipid transport and metabolism increased during crotonate axenic growth. *S. wolfei*, expressed multiple enzyme systems for fatty acid metabolism, interspecies electron transfer, and energy conservation under all growth conditions. Only a relatively few proteins (34) were unique to the syntrophic lifestyle.

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 generates H<sub>2</sub> at concentrations up to 17% with glucose. It is a member of the proposed phylum 'Synergistetes', a poorly studied microbial group composed of many environmental strains for which little is understood at the ecological, biochemical, and physiological levels.

*A. hydrogeniformans* has fermentation yields > 3 H<sub>2</sub>/glucose, and grows with and produces high levels of H<sub>2</sub> from several other carbohydrates, at least nine different amino acids, and complex substrates like Casamino acids, tryptone and yeast extract. Strain OS1 also generates H<sub>2</sub> from undefined substrates like raw sewage and other waste streams that are potential commercial substrates for H<sub>2</sub> production. Labeling, metabolomic, and enzymatic analyses showed that OS1 degrades glucose by a modified Embden-Myerhof-Parnas (EMP) where a ferredoxin-dependent, glyceraldehyde-3-phosphate ferredoxin oxidoreductase (GAPOR) is used instead of a NAD<sup>+</sup>-linked glyceraldehyde-3-phosphate dehydrogenase. Pyruvate formed by this pathway is then metabolized by pyruvate:ferredoxin oxidoreductase.

These results show that OS1 uses a ferredoxin-dependent metabolism to generate high levels of H<sub>2</sub>. The OS1 genome is approximately 2.38 MB in size with 2,557 orfs, and has a GC content of 46.5%. The current assembly consists of 151 contigs with about 2.2 MB contained in the top three contigs. Machine annotation and manual curation are

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 H<sub>2</sub>-consuming methanogen, it suggests the ability for a more complex alternative lifestyle.

## 86

### Biohydrogenesis in the Thermotogales

Andrew Frock<sup>1\*</sup> (adfrock@ncsu.edu), Amanda Petrus<sup>2\*</sup> (akpetrus@gmail.com), Derrick White<sup>3\*</sup> (derrick.white@yahoo.com), Raghuveer Singh<sup>3\*</sup> (raghuveer28m@yahoo.com), Andrew Loder,<sup>1</sup> Paul Blum,<sup>3</sup> Kenneth Noll,<sup>2</sup> and Robert Kelly<sup>1</sup>

<sup>1</sup>Dept. of Chem. and Biomolecular Engineering, North Carolina State University, Raleigh; <sup>2</sup>Department of Molecular and Cell Biology, University of Connecticut, Storrs; and <sup>3</sup>Beadle Center for Genetics, University of Nebraska, Lincoln

All contributed equally to this work.

#### Project Goals:

1. Examine the regulation of substrate catabolic proteins and pathways as this relates to carbon partitioning, disposition of reducing power, and H<sub>2</sub> generation in *Thermotoga maritima* (Tma).
2. Dissect catabolic and regulatory pathways using genetic approaches based on past success with other hyperthermophiles.
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 biohydrogenesis.

The bacterial order Thermotogales consists of obligate fermentative anaerobes that grow optimally in the range of 65-80°C. These bacteria are characterized by their unique "toga"-like outer membrane, which plays a role in breaking down a wide range of complex polysaccharides. Thermotogales are also capable of producing hydrogen with a high yield approximately twice that of mesophilic bacteria (3-4 mol H<sub>2</sub>/mol glucose compared to 1-2). Study of all aspects of Thermotogales physiology would be furthered by the development of a genetic system. The objectives of this project are stated above under "Project Goals."

To investigate these objectives, genetic, biochemical and functional genomic methods were used and revealed considerable new information about this deep understudied phylum of hydrogen evolving organisms.

A genetic system for Tma has been developed and provided fundamental information about recombination, competence and the minimal promoter for transcription. The system includes three genetic markers, methods for transformation, cell line genotyping, and cell banking. Initial studies using

the genetic system focused on nonessential genetic targets in anticipation of studies focusing on metabolic genes. Three types of Tma recombinants have been constructed: gene replacements exchanging wild type for mutant (1) or mutant for wild type genetic alleles (2), and cell lines with chromosomally integrated disruption constructs (3). In each case, PCR and DNA sequencing was used to verify genotypes of clonal cell lines. (1) Gene replacement involving repair of a 128 nt *pyrE* deletion rescued uracil auxotrophy and provided the first estimate of regions (500 bp) required for homologous recombination in Tma. (2) Gene replacement of the DNA gyrase subunit encoded by *gyrB*, introduced a synthetic *gyrB* allele arising from a G to A transition mutation at nt 401 conferring resistance to novobiocin. Incorporation of flanking synonymous third position codon changes (nt 399, 405) was used to discriminate against spontaneous drug resistance mutations and provided the first estimate of recombination rates in Tma (102/ug DNA). (3) Integration of an arabinose isomerase (*araA*) disruption cassette was accomplished by single crossover. This recombinant was produced using a *groES* promoter fused to a thermostable kanamycin marker and provided the first demonstration of a minimal sequence for transcription initiation in Tma. Currently, the genetic system is being used to target genes involved in carbon catabolism (xylose isomerase), organic acid excretion (*ackA*, *ldh*), hydrogen formation (*hydA*), hydrogen consumption (*hydA*, *hydB*, *hydC*; bifurcating hydrogenase), and several non-metabolic targets.

The most unique and visually striking aspect possessed by the Thermotogales is the outer toga structure of the cell envelope. The toga presents an interesting subject on the merit of its structural novelty within the prokaryote domain. This structure associates closely with the cell membrane and by ballooning at the poles creates a notable periplasmic space. It serves as the interface at which these organisms interact with their environment. It is also proposed to act as an extracellular matrix for the organization of enzymes and proteins, allowing the utilization of insoluble carbon sources. Therefore understanding the composition of the toga will provide essential information regarding carbohydrate utilization and the role of this outer envelope. Studies on the identification of proteins present in the toga are currently being run on maltose-grown cells. Future studies will include cells grown a variety of carbon sources to gauge their effect on protein speciation in the toga.

To investigate the functionality and composition of the toga, samples were isolated by several techniques for proteomic analysis. In our first technique, outer envelope fractions were collected from lysed cell samples by sucrose gradient fractionation. Toga sheath material collected from fractions of higher sucrose densities yielded a sample with at least fifteen protein bands visualized by SDS-PAGE analysis. This sample contains bands within the molecular weight range of the two verified structural proteins present only in the toga, *Ompα* and *Ompβ*. These fractions are being sequenced by mass spectrometry to identify toga proteins. This sample was also further purified to obtain the band matching the weight of the *Ompβ* trimer and is currently being sequenced

to identify the gene encoding *Ompβ*, which was not identified in the published genome annotation.

Outer envelope extraction from intact cells was carried out in three separate methods; (1) "trypsin shaving" of peptides from proteins with portions exposed on the cell surface, (2) biotinylation of exposed cell surface proteins using sulfo-NHS-S-biotin followed by avidin chromatography purification and (3) an organic extraction utilizing a chloroform/methanol solution that extracts surface proteins while leaving the cell membrane intact. This approach ensures that outer envelope proteins are being collected without contamination by cytoplasmic or cell membrane proteins.

Genome sequences are available for six members of the Thermotoga genus, and although some of these species are very closely related, key differences can impact carbohydrate utilization and hydrogen production. For example, Thermotoga species generally use ATP-binding cassette (ABC) transporters for sugar transport, but a single phosphotransferase system (PTS) transporter is found in *T. naphthophila* and *T. sp. RQ2*. During growth on a mixture of monosaccharides, *T. sp. RQ2* showed a preference for fructose (the putative PTS substrate) that was not observed for *T. neapolitana*, *T. petrophila*, or *Tma*. These highly similar species offer a unique opportunity to directly investigate the role that the transport mechanism plays in energy conservation and hydrogen production. Analysis of gene transcription during growth on glucose and a mix of complex polysaccharides for four Thermotoga species (*Tma*, *T. neapolitana*, *T. petrophila*, and *T. sp. RQ2*) showed that although responses to these growth substrates were mostly conserved from one species to another, *T. petrophila* exhibited a unique glucose-dependent induction of xylan/xyloside utilization genes.

Interspecies interactions and population dynamics during mixed culture are also being investigated. A real-time PCR assay has been developed to quantify the relative amounts of the same four species present in a mixed culture. Under the conditions initially tested, *T. sp. RQ2* dominates the culture after a relatively short period of time. This observation may be attributable to a faster maximum growth rate, but interspecies interactions have not been ruled out as a factor. Mixed culture experiments up to this point have been performed in batch culture, so it remains to be seen how the population dynamics are affected by a controlled growth rate in a chemostat culture.

## 87

## Pathways and Regulatory Network of Hydrogen Production from Cellulose by *Clostridium thermocellum*

Jonathan Millen,<sup>1</sup> Ji Hye Jo,<sup>2</sup> Shiv Thammannagowda,<sup>2</sup> Yuan Zhang,<sup>1</sup> Ryan Saverins,<sup>1</sup> Pin-Ching Maness,<sup>2</sup> Mary Lipton,<sup>3</sup> and J.H. David Wu<sup>1\*</sup> (davidwu@che.rochester.edu)

<sup>1</sup>University of Rochester, N.Y.; <sup>2</sup>National Renewable Energy Laboratory, Golden, Colo.; and <sup>3</sup>Pacific Northwest National Laboratory, Richland, Wash.

**Project Goals:** The overall objective of this research is to understand H<sub>2</sub> 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.

*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 hydrogenases (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. The versatility indicates 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.

To probe the physiological functions of these three hydrogenases, we performed protein immunoblots with *C. thermocellum* 27405 (acquired from ATCC) cultured in cellobiose medium. We observed the expression of CtHydA1 and CtHydA2, but not CtHydA3. Sequencing the CtHydA3-encoding gene uncovered a mutation con-

verting an amino acid residue to a stop codon and yielding a truncated protein. This mutation is confirmed by the RT-PCR data. Moreover, a second mutation was found in its upstream gene, Ct\_3004 encoding an [FeS]- protein, the putative redox partner of CtHydA3. This mutation was also confirmed by RT-PCR. The mutations yield an inactive chimeric protein fusing Ct\_3004 [FeS]-protein with the N-terminus of the CtHydA3 (mutant). However, none of these mutations were detected in another stock culture of the 27405 strain. At similar stages of growth, we consistently detected more hydrogen production in WT than mutant, assayed both *in vivo* (10% more) and *in vitro* (25% more), the latter mediated by reduced methyl viologen. The results suggest that CtHydA3 contributes to hydrogen production *in vivo*. At similar stages of growth, the mutant produced slightly more ethanol than WT, presumably by redirecting the cellular flux toward the reduced end product in lieu of hydrogen.

To identify transcription factors controlling hydrogen 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, eluted from the affinity columns, were identified by the MALDI-TOF or LC-MS-MS techniques. Several transcription factor candidates were identified. One of them binds to the promoter region of CtHydA1. The binding was confirmed by EMSA (electrophoretic mobility shift assay). Furthermore, the EMSA results indicate that binding of the protein to the promoter sequence is inhibited by NADH. Additional EMSA work using various oligonucleotides corresponding the promoter region revealed that the binding site consists of a 22 bp sequence containing a palindrome. Searching the entire *C. thermocellum* genome for the palindromic binding site sequences, we identified many potential binding sites in the intergenic sequences. EMSA analysis of these potential binding sites show that at least 5 of these promoter regions are indeed recognized by the regulatory protein. These results indicate that CtHydA1 expression is regulated by a transcription circuit that responds to the cellular redox status.

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.

## 88

## A Systems Approach to Uncovering the Metabolic and Regulatory Networks Surrounding Hydrogen Production in *Pyrococcus furiosus*

Michael P. Thorgersen\* (mthorger@uga.edu), Gerrit J. Schut,\* Gina L. Lipscomb, Jennifer R. Huddleston, Yixuan Zhu, Karen Stirrett, William J. Nixon, Janet Westpheling, Michael W.W. Adams, and **Robert A. Scott**

PI: Robert A. Scott (rscott@uga.edu)  
Co-PIs: Michael W. W. Adams, and Janet Westpheling  
Presenters: Michael P. Thorgersen (mthorger@uga.edu), and Gerrit J. Schut (gerti@uga.edu)

Department of Biochemistry and Molecular Biology,  
University of Georgia, Athens  
<http://adams.bmb.uga.edu/>

**Project Goals: The goal of SAPHyRe (Systems Approach to Probing Hydrogen Regulation) is to characterize the regulatory network surrounding hydrogen production pathways in *Pyrococcus furiosus* as it relates to carbon and nitrogen sources, metal availability, and oxidative stress. The outcome of this project will serve two purposes: 1) it will bring us one step closer to utilizing *P. furiosus* in development of alternative energy sources and 2) it will serve as a model methodology for investigating the regulatory pathways of hydrogen production in other organisms.**

The efficient production of hydrogen ( $H_2$ ) by the hyperthermophilic archaeon *Pyrococcus furiosus* is controlled by a complex regulatory and metabolic network. Hydrogen production can therefore be affected by environmental changes such as different carbon sources, the presence of electron acceptors such as sulfur, stress from reactive oxygen species, and decreased metal availability. The goal of our research is to characterize this network and identify key control points for  $H_2$  production. We developed a genetic system for *P. furiosus* and constructed several mutant strains with gene deletions predicted to have an impact on hydrogen production either directly or indirectly. *P. furiosus* contains one membrane bound ferredoxin-dependent hydrogenase and two cytosolic NADPH-dependent hydrogenases. Studies of hydrogenase deletion mutants show that the membrane bound hydrogenase functions as the major  $H_2$  producing enzyme, while the cytosolic hydrogenases only play minor roles in hydrogen metabolism. The expression of the hydrogenase genes is greatly decreased in the presence of  $S^0$ , and previous *in vitro* studies identified the redox active transcriptional regulator SurR to be involved in this response. *In vivo* studies confirmed the key role for SurR in activation of the hydrogenase operons in the absence of  $S^0$  and in mediating the primary response to  $S^0$  that results in a metabolic shift from production of  $H_2$  to  $H_2S$ . Also, a

connection between dioxygen ( $O_2$ ) stress and decreased  $H_2$  production in *P. furiosus* was found. *P. furiosus* redirects electron flow away from  $H_2$  production to deal with oxidative stress, and we identified an enzyme (FdpA) that appears to scavenge  $O_2$  *in vivo*. Using genetic technologies in *P. furiosus* is leading to a better understanding of the regulatory pathways influencing  $H_2$  production and will allow fine-tuned control at the environmental and genetic levels for improved  $H_2$  production.

## 89

## Development of *Cyanotheca* as a New Model Organism for Photobiological Hydrogen Production

Anindita Bandyopadhyay,<sup>1\*</sup> Jana Stöckel,<sup>1</sup> Lawrence Page,<sup>1</sup> Xueyang Feng,<sup>2</sup> Yinjie Tang,<sup>2</sup> Sujata Mishra,<sup>3</sup> Xiaohui Zhang,<sup>3</sup> Louis A. Sherman,<sup>3</sup> Uma K. Aryal,<sup>4</sup> Stephen J. Callister,<sup>4</sup> Janani I. Shutthanandan,<sup>4</sup> Carrie D. Nicora,<sup>4</sup> Ljiljana Pasa-Tolic,<sup>5</sup> Dave W. Koppenaal,<sup>5</sup> Richard D. Smith,<sup>4</sup> and **Himadri B. Pakrasi<sup>1,2</sup>**

<sup>1</sup>Depts. of Biology and <sup>2</sup>Energy, Environmental and Chemical Engineering, Washington University, St. Louis, Mo.; <sup>3</sup>Dept. of Biological Sciences, Purdue University, West Lafayette, Ind.; <sup>4</sup>Biological Sciences Division; and <sup>5</sup>Environmental Molecular Science Laboratory, Pacific Northwest National Laboratory, Richland, Wash.  
<http://sysbio.wustl.edu/Pakrasi/projects/hydrogen.php>

**Project Goals: The aim of this project 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  $H_2$  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:* We established a two stage system for photobiological  $H_2$  production in the unicellular cyanobacterium *Cyanotheca* ATCC 51142. Using this system, we demonstrated high rates of nitrogenase-mediated photobiological  $H_2$  production under aerobic conditions (150  $\mu$ moles of  $H_2$ /mg Chl.h).  $H_2$  production was dependent on the cellular glycogen reservoir and could be enhanced with an external carbon source like high  $CO_2$  and glycerol. A batch culture of *Cyanotheca* 51142 in the presence of glycerol could produce more than 900 ml of  $H_2$  per liter of culture over a period of two days. These rates are several folds higher compared to the rates reported for any other wild type model  $H_2$  producing strain.  $H_2$  production in *Cyanotheca* 51142 was largely driven by light even in the presence of glycerol, suggesting that the observed rates were not solely due to glycerol fermentation. However,

photosystem II did not seem to be involved in the process as observed from studies in the presence and absence of the photosystem II inhibitor DCMU (3-(3,4-dichlorophenyl)-1,1-dimethylurea). The strain is being analyzed further at the systems level to unravel the mechanism of H<sub>2</sub> production and to obtain insights into possible ways of further improving yields. H<sub>2</sub> production has also been optimized in six other sequenced *Cyanotheca* strains. Analysis of nitrogenase activity and H<sub>2</sub> production in these strains revealed the ability of aerobic nitrogen fixation and H<sub>2</sub> production in five of the six *Cyanotheca* strains. *Cyanotheca* PCC 7425 exhibited H<sub>2</sub> production only under anaerobic incubation conditions. These results will be presented and discussed in detail.

*Cyanotheca* ATCC 51472 and PCC 7822 were assayed for H<sub>2</sub> production as well as for the storage of carbon (in the form of glycogen, PHA and EPS) in detail. *Cyanotheca* 7822, but not 51472, produces copious amounts of EPS and PHA and both strains make substantial levels of glycogen. Cultures were grown under ambient and elevated (0.4%) CO<sub>2</sub> concentrations, as well as in the presence of glycerol. *Cyanotheca* 51472 grows well under high light (400 μmol.m<sup>-2</sup>s<sup>-1</sup>). Both the species had enhanced growth under increased CO<sub>2</sub> and under photomixotrophic conditions (with glycerol), and glycogen content was enhanced in 51472 under increased CO<sub>2</sub> concentration. Glycerol enhanced the nitrogenase activity and glycogen content in both the species. The level of H<sub>2</sub> evolved was positively correlated with higher levels of intracellular glycogen in both *Cyanotheca* strains. In addition, our results demonstrated that *Cyanotheca* sp 7822 was more efficient than 51472 in H<sub>2</sub> production both under photoautotrophic and photomixotrophic conditions. Although both strains produce high levels of H<sub>2</sub> (150-200 μmoles H<sub>2</sub> evolved/mg Chl/h), neither strain is as productive as *Cyanotheca* 51142.

**Comparative Genomics:** Complete genome sequences of six *Cyanotheca* strains (ATCC 51142, PCC 7424, PCC 7425, PCC 7822, PCC 8801, and PCC 8802) are currently available and one more (ATCC 51472) is in the process of completion at the DOE Joint Genome Institute. The sequencing revealed the presence of one linear chromosome in *Cyanotheca* 51142 and 3 linear elements in *Cyanotheca* 7822. This feature is unique to *Cyanotheca* strains compared to other sequenced cyanobacteria and suggest the presence of distinctive metabolic traits in members of this group. A comparison of the genomes of the six *Cyanotheca* strains revealed the presence of several pathways analogous to non-oxygenic microbes in these strains, an observation which complies with their ability to maintain a suboxic intracellular environment for a significant part of a diurnal cycle. These characteristics suggest that the group *Cyanotheca* can be appealing as model organisms for studies pertaining to biohydrogen production.

**Cyanotheca Genetics:** *Cyanotheca* 7822 was successfully transformed using a single-stranded DNA technique and a *nifK* knockout mutant was generated. A stable line of this mutant is being further analyzed. In addition, we constructed a mutant (*DhupL*, a deletion of the gene encoding the large subunit of uptake hydrogenase) in *Cyanotheca* 7822 by

inserting a neomycin/kanamycin antibiotic resistance cassette in the *hupL* gene. The growth rate of the mutant strain was about half the rate of the wild type strain in BG11 medium with nitrate, and the *DhupL* strain grew very poorly in nitrate-free BG11. In contrast to uptake hydrogenase mutants in filamentous cyanobacterial strains we observed no H<sub>2</sub> production and no nitrogenase activity in *DhupL*. The *in vitro* uptake hydrogenase activity is zero in the mutant. The mutant cells are rounder and fatter in morphology and 1.7 times larger in volume relative to the wild type cells. The knockout of *hupL* caused a defect of nitrogenase in terms of conversion of N<sub>2</sub> to NH<sub>3</sub>, as well as a defect in H<sub>2</sub> production. In the unicellular cell, HupL may play an important role in balancing *in vivo* gas metabolism.

**Metabolomic and Fluxomic Studies:** The previous <sup>13</sup>C-assisted metabolism analyses have identified unique metabolic features (*i.e.* citramalate pathway and CO<sub>2</sub> fixation regulations) in *Cyanotheca* 51142. Based on these discoveries, a constraint-based genome-scale flux balance model for *Cyanotheca* 51142 metabolism is under development. The model can be used not only to analyze functional pathways under different cultivation conditions, but also for *in silico* genetic manipulations (*e.g.* knock-out and over-expression) and to predict metabolic behavior in mutant strains. To facilitate our study on other cyanobacterial species, our model development mainly focuses on a general platform for metabolic network reconstruction and flux balance analysis. Such user-friendly and website-based software is able to automatically reconstruct genome scale metabolic network for different cyanobacterial species based on the “KEGG” database. Then the customers can manually redefine functional pathways and flux boundaries. Users can also select their own objective functions (such as maximum biomass productions, minimal enzyme usage, etc). In addition, the software can perform dynamic flux analysis by integrating kinetics and fluxomics via static optimization approach. This software can be potentially used for comparing the physiologies of different cyanobacterial species and providing the guidelines for rational design of metabolic network for CO<sub>2</sub> fixation and H<sub>2</sub> production.

**Proteomic Studies:** The proteomes of six *Cyanotheca* strains (ATCC 51142, PCC 7822, PCC 7424, PCC 7425, PCC 8801 and PCC 8802) were analyzed and compared to identify proteins common to all strains as well as unique proteins characteristic of an individual strain. The observed coverage of predicted proteins (based upon genome annotation) ranged from about 47% of the predicted genome in *Cyanotheca* PCC 7822 to 67% in *Cyanotheca* ATCC 51142. Details of these analyses will be presented and discussed.

Project number - DE-FC02-07ER64694

## 90

Systems Biology of Hydrogen Regulation in *Methanococcus maripaludis*

Sung Ho Yoon,<sup>1</sup> Qiangwei Xia,<sup>2</sup> Erik L. Hendrickson,<sup>2</sup> June Burn,<sup>4</sup> **William B. Whitman**,<sup>3</sup> **Murray Hackett**,<sup>2</sup> **Nitin Baliga**,<sup>1</sup> and **John A. Leigh**<sup>4\*</sup> (leighj@u.washington.edu)

**Co-PIs:** Nitin Baliga,<sup>1</sup> Murray Hackett,<sup>2</sup> and William B. Whitman<sup>3</sup>

<sup>1</sup>Institute for Systems Biology, Seattle, Wash.;

<sup>2</sup>Department of Chemical Engineering, University of Washington, Seattle; <sup>3</sup>Department of Microbiology, University of Georgia, Athens; and <sup>4</sup>Department of Microbiology, University of Washington, Seattle

<http://baliga.systemsbiology.net/enigma/content/hydrogen-regulation>

<http://faculty.washington.edu/leighj/>

<http://depts.washington.edu/mhlab/>

**Project Goals: Generate a regulatory network model for a hydrogenotrophic methanogen. Determine the response of the model species *Methanococcus maripaludis* to changes in hydrogen, nitrogen, and phosphate conditions.**

### 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<sub>2</sub> to reduce CO<sub>2</sub> to methane. 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<sub>2</sub> 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).

### Transcriptome structure

We have determined the transcriptome structure of *Methanococcus maripaludis* using a tiling array. We mapped transcription start and transcription termination sites for 1,025 transcription units. In many cases, we observed conditional activation of promoters inside operons and even inside coding sequences. We identified 29 antisense RNAs, 8 additional non-coding RNAs, 6 new protein-coding genes, and 5 putative overlapping protein-coding genes. A poster on the comparative evolution of transcriptome structures in four different Archaea will be presented separately (S. H. Yoon et al.).

### Regulatory network inference

We have generated quantitative transcriptome data over time courses during four transitions in chemotats: phosphate limitation to hydrogen limitation, nitrogen limitation to hydrogen limitation, high hydrogen to low hydrogen, and

low hydrogen to high hydrogen. Regulatory network inference based on this data is under way. In addition, analysis of steady-state conditions has identified numerous previously unannotated transcripts that are regulated by hydrogen limitation or generally by nutrient limitation.

### Proteomics

As a complement to our previous proteomic analysis (1), we have measured protein abundances in response to nutrient conditions, using the same strain and same samples that were analyzed at the transcriptome level. A comparison of regulatory trends at the transcriptome and proteome levels is under way. An analysis of local false discovery rates in the proteomic data will be presented in a separate poster (M. Hackett et al.). Chemostat samples are being generated for differential analysis of post-translational modifications under all three nutrient limitation conditions, using both computational and chemical approaches.

### Harvesting, quenching, and metabolite measurements

Rapid and non-destructive methods for harvesting cell samples are important for the accurate measurement of metabolite levels and potentially for the detection of conditional post-translational modifications. We have implemented a rapid-cooling method using cold methanol to prevent metabolic perturbations during harvesting. One metabolite we are focusing on is the coenzyme F<sub>420</sub>, an important electron carrier in methanogenesis. We have developed a rapid, anaerobic, acetone extraction method, which we follow with a fluorometric assay to determine levels of oxidized and reduced F<sub>420</sub> in cells.

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.

## 91

## Quantitative Local False Discovery Rates, Deep Sampling and Protein Abundance Change for *Methanococcus maripaludis*

Qiangwei Xia,<sup>1,2†</sup> David A. C. Beck,<sup>3</sup> Tiansong Wang,<sup>2</sup> E.L. Hendrickson,<sup>1</sup> **John A. Leigh**<sup>2</sup> (leighj@u.washington.edu), and **Murray Hackett**<sup>1\*</sup> (mhackett@u.washington.edu)

<sup>1</sup>Department of Chemical Engineering, <sup>2</sup>Department of Microbiology, and <sup>3</sup>eScience Institute, University of Washington, Seattle

†Current address: Regeneron Pharmaceuticals, Inc., Tarrytown, N.Y.

**PI:** John A. Leigh, University of Washington,

**Co-PIs:** Murray Hackett, University of Washington; Nitin Baliga, Institute for Systems Biology; William B. Whitman, University of Georgia

**Project Goals:** Proteomics is a rapidly changing field, and in addition to our primary goals that involve probing questions of gene regulation as they apply to nutrient limitation, methanogenesis and the potential for hydrogen production, we are also interested in improving our analytical procedures so as to get the maximum return on the resources invested. This is especially important with respect to whole cell proteomes where the need to be as comprehensive as possible must be balanced by considerations of cost and available instrument time. To that end we have been pursuing three interrelated 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* and related organisms.
2. To establish just how much sampling is required for spectral counting to become as 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.

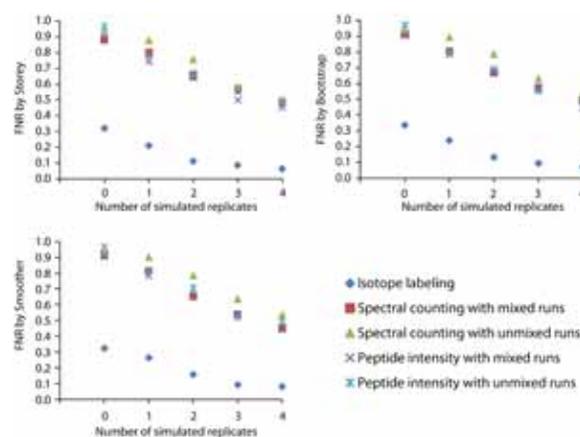
### *Methanococcus maripaludis* Proteomics

Protein abundance ratios were measured using five different approaches for the Archaeon *Methanococcus maripaludis*. Multidimensional capillary HPLC coupled with tandem mass spectrometry was used for analysis of heavy (<sup>15</sup>N) and natural abundance (<sup>14</sup>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 ~91% of the predicted protein-encoding open reading frames. False discovery rates (fdr) and local false discovery rates (lfdr) were compared as complementary approaches to multiple hypothesis testing for quantitative significance.

### Results for calculations using fdr and lfdr theory

At the limit of deep sampling frequency based measurements are competitive but lack the same power to detect relative abundance change when compared with metabolic stable isotope labeling, based on a conventional *t*-test for two sample comparisons assuming unequal variances (see figure below, lower FNR (false negative rate) means higher power, first entries on the left are real data, followed by Monte Carlo simulations of additional replicates. In this example a two-fold change cutoff was used). This result was driven primarily by the large numbers of heavy-light pairs used in the *t*-test calculations, sometimes several thousand for a single calculation for a given ORF observed in a two-state comparison. In general the lfdr calculations were difficult to implement with the *M. maripaludis* nutrient limitation data due to the relatively poor fit of the p-values and (or) z-scores used as inputs and the strict requirements of existing lfdr algorithms with respect discontinuities in the input distributions.



This research was supported by the Office of Science (BER), U.S. Department of Energy, Award No. DE-FG02-08ER64685

## 92

## Development of Biologically-Based Assays to Study Rate-Limiting Factors in Algal Hydrogen Photoproduction

**Maria L. Ghirardi**<sup>1\*</sup> (maria.ghirardi@nrel.gov),  
**Alexandra Dubini**<sup>1</sup> (alexandra.dubini@nrel.gov), and  
**Matthew Wecker**<sup>2</sup> (matt.wecker@nrel.gov)

<sup>1</sup>Biosciences Center, National Renewable Energy Laboratory, Golden, Colo.; and <sup>2</sup>Genebiologics, LLC, Boulder, Colo.

**Project Goals: To develop techniques that will provide a deeper understanding of algal H<sub>2</sub> metabolism and accelerate the development of future photobiological H<sub>2</sub>-producing catalysts and organisms.**

Photobiological H<sub>2</sub> production from water is a clean, non-polluting and renewable technology. The efficiency of light conversion into H<sub>2</sub> by biological organisms is theoretically high (about 10%). However, the system is currently limited by biochemical and engineering constraints, including the extreme O<sub>2</sub> sensitivity of the biological hydrogenases and the low availability of reductants to the hydrogenase due to the predominance of competing metabolic pathways. We address the O<sub>2</sub> sensitivity issue by developing an assay to screen large microbial populations for improved H<sub>2</sub>-production properties with the eventual goal of evolving O<sub>2</sub> tolerance in [FeFe]-hydrogenases. The assay that we have created is based on the H<sub>2</sub> sensing system of the photosynthetic bacteria *Rhodobacter capsulatus* and it is tied to the production of a green fluorescence protein in response to H<sub>2</sub>. The issue of competitive pathways utilizes the yeast two-hybrid assay to measure and deconvolute the interactions between different ferredoxin isoforms present in the green alga, *Chlamydomonas reinhardtii*, with their known electron acceptors. Identified interactions are then quantified by isothermal titration calorimetry.

We have demonstrated that the *Rhodobacter*-based H<sub>2</sub>-sensing assay is remarkably sensitive to H<sub>2</sub>, even at levels below the background level of H<sub>2</sub> in air, thus rivaling the sensitivity of the best known chemical sensors to H<sub>2</sub>. We have shown that H<sub>2</sub>-reporting bacteria, co-cultured with *Chlamydomonas reinhardtii* under anaerobic conditions fluoresce in direct relation to the amount of H<sub>2</sub> produced by the algae. This suggests a useful means of selecting *Chlamydomonas* H<sub>2</sub>-production mutants with improved H<sub>2</sub> production activity. We will focus next on adapting the assay for operation under photobiological H<sub>2</sub>-production conditions and use it to screen for O<sub>2</sub>-tolerant [FeFe]-hydrogenases generated through directed-evolution techniques. The hydrogenases of *Clostridium acetobutylicum* and *C. reinhardtii* and hydrogenase assembly proteins from *Bacteroides thetaio-taomicron* have been introduced into a hydrogenase-negative *R. capsulatus* strain and are being tested for heterologous H<sub>2</sub> production.

To address the issue of competitive metabolic pathways with H<sub>2</sub> production, we have performed a yeast two-hybrid assay to screen for proteins interacting with the 6 ferredoxin (FDX) isoforms and the 2 hydrogenases (HYD) present in *Chlamydomonas reinhardtii* to characterize the pathways involved in shuffling electrons away from or directly to the hydrogenases. The results have yielded a list of genes encoding for different proteins that are going to be the primary targets for future metabolic engineering and used also to understand how the reductants are allocated to the HYDs. Each FDX and HYD has been shown to interact with specific binding partners. We have confirmed those interactions by library screening and pairwise interaction studies. We have now a metabolic network map available from those studies where the FDXs and the HYDs have been assigned a potential function/interaction in the newly discovered pathways. The major information that resulted from this study is showing that FDX1 is interacting with both HYDs, confirming its role in hydrogen photoproduction. The next step is to study those interactions in vitro and to define their affinity binding using the IsoThermal Calorimetry technique. Currently, HYDs and FDX1 are being purified and will be the first candidates to be tested for their binding constant.

### Publications:

1. Matt S.A. Wecker, Jon Meuser, Matthew Posewitz and Maria L. Ghirardi "Design of a new biosensor for algal H<sub>2</sub> production based on the H<sub>2</sub>-sensing system of *R. capsulatus*". J. Int. Hydrogen Energy, submitted.

## 93

## Phototrophic Metabolism of Organic Compounds Generates Excess Reducing Power That Can Be Redirected To Produce H<sub>2</sub> as a Biofuel

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

Department of Microbiology, University of Washington, Seattle

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

Program title: Hydrogen production

**Project Goals: The goals of this project are (i) to use <sup>13</sup>C-metabolic flux analysis and other approaches to identify metabolic factors that influence the phototrophic production of H<sub>2</sub> from organic compounds by *R. palustris* and (ii) to use the resulting information to guide the engineering of *R. palustris* for improved H<sub>2</sub> production characteristics.**

There is currently a pressing need for renewable fuels to negate the adverse impacts of burning fossil fuels. H<sub>2</sub> is a promising fuel, having about three-times the energy content of gasoline. Although most manufactured H<sub>2</sub> comes from fossil fuels, H<sub>2</sub> can also be produced biologically. *Rhodospirillum rubrum* (R. rubrum) is a photosynthetic bacterium that produces H<sub>2</sub> from organic compounds. To address the issue of competitive metabolic pathways with H<sub>2</sub> production, we have performed a yeast two-hybrid assay to screen for proteins interacting with the 6 ferredoxin (FDX) isoforms and the 2 hydrogenases (HYD) present in *Chlamydomonas reinhardtii* to characterize the pathways involved in shuffling electrons away from or directly to the hydrogenases. The results have yielded a list of genes encoding for different proteins that are going to be the primary targets for future metabolic engineering and used also to understand how the reductants are allocated to the HYDs. Each FDX and HYD has been shown to interact with specific binding partners. We have confirmed those interactions by library screening and pairwise interaction studies. We have now a metabolic network map available from those studies where the FDXs and the HYDs have been assigned a potential function/interaction in the newly discovered pathways. The major information that resulted from this study is showing that FDX1 is interacting with both HYDs, confirming its role in hydrogen photoproduction. The next step is to study those interactions in vitro and to define their affinity binding using the IsoThermal Calorimetry technique. Currently, HYDs and FDX1 are being purified and will be the first candidates to be tested for their binding constant.

*seudomonas palustris* uses energy from sunlight and electrons from organic waste to produce H<sub>2</sub> via nitrogenase. In order to understand and improve this process we used <sup>13</sup>C-substrates having various oxidation states to track and compare central metabolic fluxes in non-H<sub>2</sub> producing wild-type *R. palustris* and an H<sub>2</sub>-producing mutant. The pathways by which substrates were oxidized generated excessive amounts of reducing power such that only 40-60% could be used for biosynthesis, depending on the growth substrate. Wild-type cells relied heavily on the CO<sub>2</sub>-fixing Calvin cycle to oxidize the excess reduced electron carriers, using CO<sub>2</sub> produced from the organic substrates by other metabolic reactions. The H<sub>2</sub>-producing mutant used a combination of CO<sub>2</sub> fixation and H<sub>2</sub> production to oxidize excess reduced electron carriers. The majority of electrons for H<sub>2</sub> production were diverted away from CO<sub>2</sub> fixation for all substrates. Microarray and qRT-PCR analyses indicated that this shift of electrons towards H<sub>2</sub> involved transcriptional control of Calvin cycle gene expression. These observations pointed to the Calvin cycle as a convenient single target to disrupt to force more electrons towards H<sub>2</sub> production. Blocking Calvin cycle flux by mutation in the H<sub>2</sub>-producing strain resulted in higher H<sub>2</sub> yields for all substrates. The increase in H<sub>2</sub> yield was proportional to the Calvin cycle flux in the parent strain for most substrates. These results demonstrate how systems level approaches, such as <sup>13</sup>C-metabolic flux analysis, can lead to effective strategies to improve product yield. Furthermore, our results underscore that the Calvin cycle and nitrogenase have important electron-accepting roles separate from their better known roles in biomass generation and ammonia production.

## 94 Large Scale Functional Genomic Analysis of Hydrogen-Producing *Rhodospseudomonas palustris*

Yasuhiro Oda<sup>1\*</sup> (yasuhiro@uw.edu), Christopher D. Armour,<sup>2</sup> Somsak Phattarasukol,<sup>1</sup> Colin R. Lappala,<sup>1</sup> John C. Castle,<sup>3</sup> Matthew C. Biery,<sup>2</sup> Christopher K. Raymond,<sup>2</sup> and **Caroline S. Harwood**<sup>1</sup>

<sup>1</sup>Department of Microbiology, University of Washington, Seattle; <sup>2</sup>NuGEN Technologies, Inc., San Carlos, Calif.; and <sup>3</sup>Institute for Translational Oncology and Immunology, Mainz, Germany

**Project Goals: To use a systems level approach to dissect metabolic and regulatory networks that are critical for nitrogenase-catalyzed hydrogen production by a phototrophic bacterium *Rhodospseudomonas palustris*.**

Hydrogen gas has good potential as an alternative fuel since it is clean burning and has a high energy content. Bacteria can produce hydrogen using hydrogenases or nitrogenases. The photosynthetic bacterium *Rhodospseudomonas palustris* produces copious amounts of hydrogen along with ammonia via the enzyme nitrogenase. Nitrogenase-catalyzed hydrogen production requires large amounts of ATP and electrons,

which *R. palustris* can obtain from sunlight and biomass (such as agricultural and industrial waste), 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 genomic, transcriptomic, and phenotypic data from up to 100 *R. palustris* strains using Bayesian network analysis to identify all genes that are involved in hydrogen production. This would include 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.

The emergence of next generation sequencing technology has opened up new opportunities to obtain large scale systems biology information. The success of high-throughput sequence-based transcriptome profiling depends on the availability of both efficient methods for the construction of high complexity cDNA libraries and computational data analysis tools. Here, we describe molecular and computational tools for high-throughput analysis of bacterial transcriptomes. Our cDNA library construction method utilizes computationally designed hexamers to selectively enrich non-rRNA transcripts during cDNA synthesis. This method requires only 500 ng of total RNA and enables the construction of barcoded cDNA libraries within 6h. When we applied this method to characterize transcriptome profile of *R. palustris* CGA009, on the order of 50% of the total sequencing reads were mapped to non-rRNA transcripts. We also developed a sequence analysis pipeline that provides global mapping statistics, gene-level expression profiles, and chromosomal expression maps as standard outputs. Using our tools, we characterized the transcriptomes of five *R. palustris* strains grown under the conditions of nitrogen-fixing (hydrogen-producing)-high light, nitrogen-fixing-low light, and ammonia-high light. Under these conditions, there were significant differences in growth rates, hydrogen production, and nitrogenase activities among strains tested. Genes that were previously identified as highly expressed under hydrogen-producing conditions by Affymetrix GeneChip were also expressed at higher levels in all strains, including the molybdenum-nitrogenase gene cluster, PII nitrogen regulatory genes, and ammonium transporter genes. Interestingly, levels of expression among these genes varied significantly from one strain to another, indicating that these expression changes may result in the observed phenotypic differences among the strains. We also identified a ferredoxin gene (which may donate electrons to nitrogenase) and several hypothetical genes that were highly expressed among the strains that had higher hydrogen production and nitrogenase activities.

These results demonstrate that our molecular and computational tools are a powerful way to characterize bacterial transcriptomes. We further demonstrated the integration of large scale functional genomic and phenotypic data derived from hydrogen-producing *R. palustris* strains. These studies have revealed novel insights into bacterial transcriptome dynamics and have led to the identification of several genes that may be important for hydrogen production.

95

Student Oral Presentation—Monday

**Xpression: A Complete Pipeline for Analyzing RNA-Seq Data**Somsak Phattarasukol<sup>1\*</sup> (sphatt@u.washington.edu), Matthew Radey,<sup>2</sup> Mitchell Brittnacher,<sup>2</sup> and **Caroline S. Harwood**<sup>1</sup><sup>1</sup>Dept. of Microbiology and <sup>2</sup>Northwest Regional Center of Excellence, University of Washington, Seattle**Project Goals: To use a systems level approach to dissect metabolic and regulatory networks critical for nitrogenase-catalyzed hydrogen production by the phototrophic bacterium *Rhodospseudomonas palustris*.**

RNA-seq is a revolutionary tool for transcriptome analyses. It has a number of advantages over traditional microarray-based technologies (e.g. annotation-independent detection of transcription, improved sensitivity and increased dynamic range), which collectively results in a much more precise measurement of gene expression levels. Despite these clear benefits, wide spread adoption of RNA-seq by researchers is impeded by the lack of an integrated software to analyze the data. Currently, millions of raw sequence reads are generated for each RNA-seq experiment, which makes it impossible to interpret these data without using computational tools.

We have developed Xpression, a complete pipeline for analyzing RNA-seq data generated from the Illumina platform. The workflow starts involves 1) filtering out low-quality sequence reads in raw sequence files, 2) aligning the sequence reads against a reference genome using the Burrows-Wheeler Alignment (BWA) tool, 3) computing statistics of sequence mapping, 4) reporting the number of sequence reads uniquely mapped to a particular region of the genome to indicate expression levels, and 5) generating plots to illustrate expression data (as normalized graphs of sequence read coverage across a genome) in platforms such as Artemis or Integrated Genome Browser.

Note that outputs from Xpression can be conveniently used in further downstream analyses. For example, users can additionally apply statistical softwares such as edgeR, DEseq or baySeq on expression data to identify differentially expressed genes, or open an alignment file in specialized visualization softwares such as Tablet. Also, Xpression can be configured to match certain requirements. For example, users may specify what barcodes are attached to biological samples, the types of library preparation protocol being used and whether strandness of sequencing reads is conserved.

96

**Reconstructing the Metabolic Network of *Rhodobacter sphaeroides* 2.4.1**Saheed Imam<sup>1\*</sup> (sriram@wisc.edu), Safak L. Yilmaz,<sup>2</sup> Ugur Sohmen,<sup>2</sup> Timothy J. Donohue<sup>3</sup> (tdonohue@bact.wisc.edu), and **Daniel R. Noguera**<sup>2</sup> (noguera@engr.wisc.edu)<sup>1</sup>Program in Cellular and Molecular Biology, <sup>2</sup>Department of Civil and Environmental Engineering, and <sup>3</sup>Department of Bacteriology, University of Wisconsin, Madison**Project Goals: The goals of this project include: (i) creating a genome-scale metabolic reconstruction for *R. sphaeroides* 2.4.1; (ii) determination of its biomass composition to generate suitable biomass objective functions for modeling; (iii) validating the generated model using constraint based analysis and comparing predictions to experimental data; and (iv) employing the model in hypothesis generation.**

**Background:** *Rhodobacter sphaeroides* represents one of the best studied members of the group of purple non-sulfur photosynthetic bacteria and serves as an excellent model for the study of the complex and versatile metabolic capabilities of this group of organisms. The ability of *R. sphaeroides* to produce large amounts of hydrogen gas (H<sub>2</sub>), polyhydroxybutyrate (PHB) and fatty acids photoheterotrophically, as well as its ability to utilize atmospheric carbon dioxide (CO<sub>2</sub>) as a carbon source during photoautotrophic growth, make it an excellent candidate of potential use in a wide variety of biotechnological applications. To effectively harness its biotechnological potential, a thorough understanding of its metabolic capabilities is necessary. The goals of this project included: (i) creating a genome-scale metabolic reconstruction for *R. sphaeroides* 2.4.1; (ii) determination of its biomass composition to generate suitable biomass objective functions for modeling; (iii) validating the generated model using constraint based analysis and comparing predictions to experimental data; and (iv) employing the model in hypothesis generation.

**Methods:** Network reconstruction was conducted using all archived *R. sphaeroides* 2.4.1 metabolic and genomic information from KEGG database combined with metaSHARK analysis. The model generated from the reconstruction was used in simulations employing constraint based approaches including flux balance analysis (FBA), flux variability analysis (FVA) and alternate optima analysis.

*R. sphaeroides* cells were grown photosynthetically in continuous culture and the biomass composition of cells determined using established laboratory techniques. Gas production was measured by respirometry, with H<sub>2</sub> and CO<sub>2</sub> contents determined via gas chromatography.

**Results:** The reconstructed metabolic network of *R. sphaeroides* 2.4.1 consists of 796 metabolites, 856 transformation

reactions and 300 transport reactions. The reconstruction accounts for 1095 genes, covering about 25% of the *R. sphaeroides* open reading frames. The subsystem distribution of the reactions in the reconstruction is summarized in Fig 1. The empirically determined biomass composition of photosynthetically growing *R. sphaeroides* (Fig 2), was used in generating a biomass objective function for constraint based analysis. FBA simulations, which allowed for prediction of metabolic flux distributions during aerobic, photoheterotrophic and photoautotrophic growth, showed good qualitative agreement with experimental data (Table 1). Model predictions of growth rate, PHB synthesis, H<sub>2</sub> and CO<sub>2</sub> also showed good agreement with experimental data. Finally the model was also used in the identification of other potential pathways for carbon assimilation and cellular redox balancing, thus serving as a source of testable hypotheses.

**Conclusions:** Our model of *R. sphaeroides* metabolic network serves as an important starting point in our goal of reconstructing and integrating the metabolic, regulatory and signaling networks of this complex microbe. The metabolic model shows good qualitative and quantitative agreement with experimental data and thus should provide an excellent framework for future metabolic modeling of *R. sphaeroides* and other related bacteria.

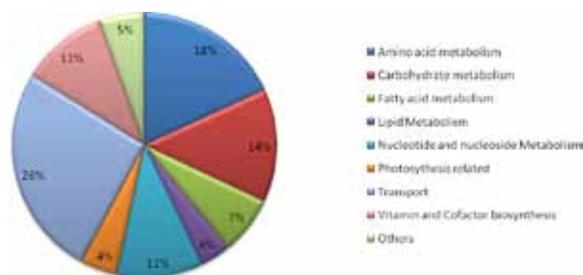


Figure 1. Subsystem distribution of reactions in reconstruction. Pie chart depicts the subsystem distribution of the model's reactions, with the percent contribution of each subsystem of reactions indicated in the corresponding section of the chart. It can be seen that amino acid, carbohydrate and nucleotide metabolism dominate the transformation reactions present in the model.

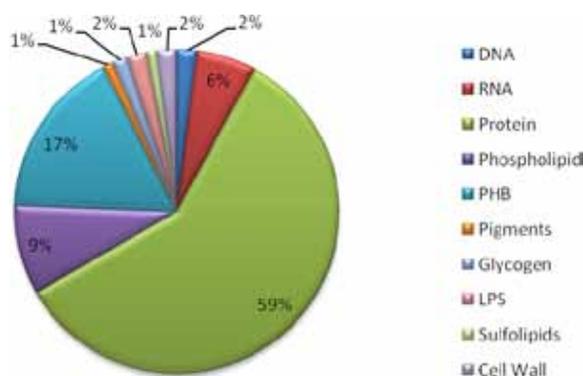


Figure 2. Biomass composition of photosynthetically grown *R. sphaeroides*.

Carbon/nitrogen source	Light	Dark		
		O <sub>2</sub>	DMSO	None
Succinate + NH <sub>3</sub>	+/+	+/-	+/-	-/-
Succinate + Glutamate	+/+	+/-	+/-	-/-
Lactate + NH <sub>3</sub>	+/+	+/-	+/-	-/-
Glutamate only	+/-	+/-	+/-	-/-
CO <sub>2</sub> + H <sub>2</sub> + NH <sub>3</sub>	+/-	-/-	-/-	-/-
CO <sub>2</sub> + H <sub>2</sub> + N <sub>2</sub>	+/-	-/-	-/-	-/-

Table 1: Qualitative assessment of metabolic model. Table summarizes model predictions during aerobic and photosynthetic growth. +/+ Growth and H<sub>2</sub> production predicted; +/- Growth but no H<sub>2</sub> production predicted; -/- No growth

## 97 Systems Level Approaches to Understanding and Manipulating Heterocyst Differentiation in *Nostoc punctiforme*: Sites of Hydrogenase and Nitrogenase Synthesis and Activity

Harry Christman,<sup>1</sup> Elsie Campbell,<sup>1</sup> Douglas Rissler,<sup>1</sup> Brett Phinney,<sup>2</sup> Wan-Ling Chiu,<sup>3</sup> and John C. Meeks<sup>1\*</sup> (jcmeeks@ucdavis.edu)

<sup>1</sup>Department of Microbiology and <sup>2</sup>Genome Center, University of California, Davis; and <sup>3</sup>Department 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, H<sub>2</sub>, is uncoupled from growth. The goals of this project are to apply transcriptomic and proteomic analyses to free-living and symbiotically grown wild type and mutants strains to identify regulatory elements and pathways that can be manipulated.

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. There is an increase in symbiotic nitrogen fixation that parallels the increased heterocyst frequency and about 85-90% of the metabolic end product, ammonium, is excreted to the plant partner. This is the kind of metabolic uncoupling we seek for biohydrogen production. We model heterocyst differentiation in the context of establishing the pattern of

heterocyst spacing, maintaining the pattern during growth on N<sub>2</sub> and symbiotic disruption of the spacing pattern. Establishment of the pattern following combined nitrogen deprivation is further modeled as a two stage process: biased initiation where a cluster of cells sense nitrogen starvation and initiate the differentiation process, followed by competitive resolution through interactions between HetR and PatS, where a single cell in the cluster completes the differentiation and maturation into a mature functional heterocyst. We do not know whether the same or similar regulatory networks are involved in establishment, maintenance and disruption of the pattern. 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 may have been co-opted during symbiotic growth.

**Genetics.** We have isolated an exogenously induced transposon mutant of *N. punctiforme* that displays a heterocyst frequency (> 30% of the total cells) and spacing pattern (multiple singular heterocysts) that is essentially the same as that in symbiotic association. This is the heterocyst frequency and spacing pattern we hypothesize would be an experimental platform for enhanced hydrogen production. The mutant does grow with N<sub>2</sub> as the sole nitrogen source, but at a slower rate than the wild type and with about 30% of the rate of nitrogen fixation (acetylene reduction). Thus, the mutant physiology does not parallel that of the symbiotic growth state. The targeted gene encodes a protein unique to heterocyst-forming cyanobacteria that we have designated PatN. PatN has four distinct domains: a 30 amino acid cytoplasmic region, a transmembrane domain, a glutamine rich coiled-coil domain in the periplasm, and a signal sequence. Our working hypothesis is that PatN may be involved in the translocation of a negative acting heterocyst differentiation element, such as, for example, the RGSGR C-terminal pentapeptide of PatS. We are setting up a microscope for fluorescence recovery after photobleaching (FRAP) analysis to examine translocation properties of the mutant relative to the wild type. The RGSGR pentapeptide is also present in HetN, which is modeled to be involved in maintenance of the heterocyst spacing pattern. The putative HetN protein in *N. punctiforme* lacks the RGSGR motif. Bioinformatic analysis revealed two additional proteins encoded by *N. punctiforme* as containing the RGSGR motif. Single deletion mutants do not result in a multiple contiguous heterocyst (Mch) spacing pattern as is characteristic of *hetN* mutants of *Anabaena* sp. strain PCC 7120. Moreover, *patS* deletion mutants of *N. punctiforme* do not show an extensive Mch pattern. These results imply considerably more cross-talk or complementation in negative acting regulatory elements in *N. punctiforme* compared to *Anabaena* 7120. To confirm this suggestion we are constructing multiple deletion mutants.

**Transcriptomics.** To more rigorously analyze time course microarray data, we have adapted the Bayesian Analysis of Time Series (BATS) software. BATS analyses have reduced the number of statistically significant genes that are differentially expressed during heterocyst differentiation in the wild type and mutant strains. The reanalysis confirmed that

the wild type (511 genes) and mutants impaired in HetR (1150 genes) and HetF (722 genes) (two positive acting regulatory elements) function have different patterns and numbers of genes differentially transcribed during nitrogen starvation. We interpret these data to verify that the primary role of HetR and HetF, directly or indirectly, is to influence gene transcription. Conversely, other than *patN* itself, no other genes are altered in their patterns of differential transcription in the *patN* mutant. This observation implies that PatN has a structural or functional as opposed to transcriptional role in the differentiation of heterocysts and supports the FRAP analysis above to define that role.

**Proteomics.** We have completed three replicates of three dimensional SDS-PAGE and LC fractionation of N<sub>2</sub>-grown, wild type *N. punctiforme* and protein identification by MudPit MS/MS. These analyses yielded 1,210 identified proteins in the membrane (40,000 x g pellet), molecular complex (150,000 x g pellet) and soluble (150,000 x g supernatant) fractions. Known heterocyst structural and catalytic proteins were present in all fractions. We have completed two replicates of symbiotic colonies isolated from the hornwort *Anthoceros punctatus*. Due to the small amount of biomass, the symbiotic cell extracts were not subjected to centrifugal fractionation. These analyses yielded only 306 identified proteins. However, the identified proteins were highly enriched in cell envelope, energy metabolism and transporters. Present were PsbB and PsbC, but not PsbA, which may account for the lack of photosynthetic oxygen evolution in heterocysts and the lower rate of symbiotic CO<sub>2</sub> fixation by *N. punctiforme*. We are currently exploring methods to fractionate the small-volume, symbiotic cell extracts for better resolution of less abundant proteins.

## 98

### Pathway of Fermentative Hydrogen Production by Sulfate-Reducing Bacteria

Kimberly L. Keller<sup>1\*</sup> (kellerkl@missouri.edu), Steve D. Brown,<sup>2,3</sup> Iris Porat,<sup>2</sup> Adam Deutschbauer,<sup>4,3</sup> Jennifer Kuhl,<sup>4,3</sup> Grant M. Zane,<sup>1,3</sup> and **Judy D. Wall**<sup>1,3</sup> (WallJ@missouri.edu)

<sup>1</sup>University of Missouri, Columbia; <sup>2</sup>Oak Ridge National Lab, Oak Ridge, Tenn.; <sup>3</sup>ENIGMA (Ecosystems and Networks Integrated with Genes and Molecular Assemblies), Berkeley, Calif.; and <sup>4</sup>Lawrence Berkeley National Laboratory, 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.**

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 [DvH]) may be informative to decipher the flow of electrons in those organisms chosen for industrial application for hydrogen production.

**Aim 1:** We proposed to tease apart the contribution of fermentation to the respiratory energy budget to determine the dependence of the bacterium on this process. We want to eliminate substrate-level phosphorylation by creating a deletion of the gene encoding acetate kinase to confirm that this enzyme is essential for substrate-level phosphorylation during pyruvate fermentation. Determine the effect of this deletion on the efficiency of pyruvate and lactate respiration.

The acetate kinase in both organisms is annotated to be in a 10 gene operon. Transposon libraries of G20 and DvH selected under conditions that would allow mutations in genes non-essential for respiration lacked mutations in the gene, *ack*. Proteomic data from *Desulfovibrio* G20 indicated that the abundance of the enzymes involved in substrate-level phosphorylation was actually decreased when cultures were fermenting pyruvate compared to growth on lactate/sulfate. Microarray data from *D. vulgaris* indicated expression of acetate kinase was decreased when growing with hydrogen as a electron donor where substrate-level phosphorylation is not possible. Efforts are being made to use these growth modes to obtain a deletion of *ack* in *D. vulgaris*. Once obtained, the mutant will be characterized to determine the contribution of this pathway to the overall energy budget of the SRB.

**Aim 2:** We proposed to identify the enzyme(s) responsible for oxidizing pyruvate during fermentation and the role of formate, if any, in pyruvate fermentation. This information will distinguish the potential pathways to hydrogen from substrate oxidation. Two different enzymes are annotated in the genome, pyruvate:ferredoxin oxidoreductase and pyruvate formate lyase, each of which has multiple possible orthologs/paralogs.

Transposon mutants of the two annotated pyruvate formate-lyase genes have been obtained in both *Desulfovibrio* strains. Because of a developed markerless deletion system, work has begun to delete the two candidates, singly and in combination, monitoring the products of the metabolism of the mutants. Transpo-

son mutants in three of the four operons encoding proteins annotated as pyruvate:ferredoxin oxidoreductases have also been obtained in both *Desulfovibrio* strains. Fitness experiments with a mini-pool of 1,163 transposon mutants of *Desulfovibrio* G20 indicated that one pyruvate ferredoxin oxidoreductase (Dde\_1639) mutant did have decreased fitness in all experiments performed to date and Dde\_3237 had no transposon insertions.

**Aim 3:** We proposed to channel electrons from alternative sinks to hydrogen during fermentation and determine the effects of removal of those sinks on the fermentation efficiency. We will attempt to eliminate alternative electron sinks that are likely to compete for protons, reducing the overall yield of hydrogen.

Proteomic and transcriptomic data from *Desulfovibrio* G20 indicate the importance of the fumarate reductase in growing cultures, especially in the later stages of growth. Transposon mutants lacking fumarate reduction to succinate are being studied since fumarate may be a possible electron sink in competition with protons and loss of this activity may alter the energy budget of the cells. Growth on fumarate is inhibited by the presence of formate, hydrogen, or carbon dioxide; whereas, formate is only accumulated during growth of G20 on fumarate inhibited with hydrogen. The hydrogen inhibition may be due to an inability to reoxidize ferredoxin or a possible blockage of proton pumping. In addition, the histidine kinase (Dde\_1261) and the response regulator (Dde\_1260) upstream of the fumarate reductase operon appear to play a critical role in the utilization of fumarate as transposon mutants in either of these two genes renders the cells unable to grow on fumarate.

## 99

### Defining the Metabolic Landscape of *Chlamydomonas reinhardtii* by Large-Scale Annotation of its Encoded Open Reading Frames and Metabolic Network Modeling

Lila Ghamsari,<sup>1</sup> Balaji Santhanam,<sup>1</sup> Roger L. Chang,<sup>2</sup> Paul Jensen,<sup>3</sup> Patrick Ng,<sup>4</sup> Ani Manichaikul,<sup>3</sup> Erik F.Y. Hom,<sup>5</sup> Dawit Balcha,<sup>1</sup> Yun Shen,<sup>1</sup> Changyu Fan,<sup>1</sup> Tong Hao,<sup>1</sup> Bernhard Ø. Palsson,<sup>2</sup> Haiyuan Yu,<sup>4</sup> **Jason A. Papin,<sup>3</sup>** and **Kourosh Salehi-Ashtiani<sup>1,6\*</sup>** (ksa3@nyu.edu)

<sup>1</sup>Center for Cancer Systems Biology and Department of Cancer Biology, Dana-Farber Cancer Institute, and Department of Genetics, Harvard Medical School, Boston, Mass.; <sup>2</sup>Department of Bioengineering, University of California, San Diego; <sup>3</sup>Department of Biomedical Engineering, University of Virginia, Charlottesville; <sup>4</sup>Department of Biological Statistics and Computational Biology and Weill Institute for Cell and Molecular Biology, Cornell University, Ithaca, N.Y.; <sup>5</sup>Department

of Molecular and Cellular Biology, Harvard University, Cambridge, Mass.; and <sup>6</sup>New York University Abu Dhabi, Abu Dhabi, UAE, and Center for Genomics and Systems Biology, Department of Biology, New York University, New York

**Goals:** The release of the complete genome sequence of *Chlamydomonas 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.

**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 (and AraCyc) 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 over 900 enzyme annotations to 1,427 *JGI* and to 1,877 *Augustus* models. Over 93% of the EC terms were assigned to both *JGI* and *Augustus* models. 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, indicating that the structural annotation of many of the two sets differ from one another.

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 tested the structure of the metabolic-related ORF models by reverse transcription-PCR of the functionally annotated ORFs. Following cloning, we carried out 454FLX sequencing of the ORFs. Based on alignment of the 454FLX reads to the ORF predicted sequences, we obtained more than 90% coverage for 80% of the metabolic ORFs. Only 99 ORFs were not verified using this experimental pipeline.

We obtained expression evidence for 93% of the metabolic ORFs in the algal cells grown under constant light and in the presence of acetate.

Using our in-house generated functional annotation (described above), combined with literature and publicly available database resources, we have reconstructed the first genome-scale reconstruction of *C. reinhardtii* metabolic network, accounting for all pathways and metabolic functions indicated. The reconstruction accounts for 1,080 genes, associated with 2,190 reactions and 1,068 unique metabolites. Our reconstruction accounts for multiple wavelengths of light and includes considerable expansion of fatty acid metabolism over previous reconstructions. Further, the metabolic network reconstruction 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.

**Conclusion:** Our 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. The verified metabolic ORF clones will provide the experimental resource needed for downstream experiments and will be made available to the research community.

The DOE support for this research was generously made through a grant from the Office of Science (Biological and Environmental Research), U.S. Department of Energy, grant No. DE-FG02-07ER64496 to J.P. and K.S.-A.

## Computing for Bioenergy

# 100

### Multiscale Coarse-Grain Simulation Studies of Cellulosic Biomass

Goundla Srinivas<sup>1\*</sup> (srinivasg@ornl.gov), Dennis Glass,<sup>1</sup> Xiaolin Cheng,<sup>1</sup> Jeremy Smith,<sup>1,2</sup> Mark S. Gordon,<sup>3</sup> Yana Kholod,<sup>3</sup> Monica H. Lamm,<sup>3</sup> Sergiy Markutsya,<sup>3</sup> John Baluyut,<sup>3</sup> and Theresa L. Windus

<sup>1</sup>Oak Ridge National Laboratory, Oak Ridge, Tenn.; and <sup>2</sup>Biochemistry and Cellular and Molecular Biology, University of Tennessee, Knoxville; and <sup>3</sup>Applied Mathematics and Computational Science, Ames Laboratory, Department of Chemistry, Iowa State University, Ames

#### Project Goals:

1. To develop computationally affordable large scale coarse-grain force field for cellulosic biomass.
2. Study structure and dynamics of cellulosic biomass from a micro and macroscopic viewpoint in order to understand biomass recalcitrance.