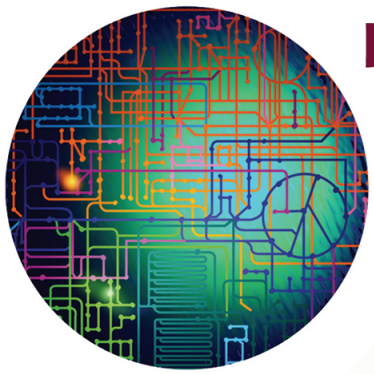
DOE/SC-0122



DOE GENOMIC SCIENCE

SYSTEMS BIOLOGY FOR ENERGY AND ENVIRONMENT

Genomic Science 2010

Awardee Workshop VIII and Knowledgebase Workshop

U.S. DEPARTMENT OF ENERGY

Office of Science

February 7–10, 2010

DOE Genomic Science Program
U.S. Department of Energy
Office of Science
Office of Biological and Environmental Research
genomic.science@science.doe.gov

Meeting Contact
Susan Gregurick
U.S. Department of Energy
Office of Biological and Environmental Research
301.903.7672
susan.gregurick@science.doe.gov

Sharlene Weatherwax
Director, Biological Systems Science Division
U.S. Department of Energy
Office of Biological and Environmental Research
301.903.3213
sharlene.weatherwax@science.doe.gov

An electronic version of this document is available at the Genomic Science web site:
http://GenomicScience.energy.gov/pubs/2010abstracts/

DOE/SC-0122

U.S. DEPARTMENT OF
ENERGY

Office of Science

Genomic Science 2010 Awardee Workshop VIII

and

Knowledgebase Workshop

Crystal City, Virginia
February 7–10, 2010

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

Welcome

Department of Energy
Washington, DC 20585
January 11, 2010

Dear Participant:

We wish to welcome you to the 2010 Genomic Science Contractor-Grantee and Knowledgebase Workshop. As the premier meeting for the Genomic Science program, this workshop brings together researchers supported by the program, representatives from the Department of Energy, and colleagues from other federal agencies. The Genomic Science program—within the Biological Systems Science Division of the Office of Biological and Environmental Research (BER)—supports fundamental research to achieve a predictive, systems-level understanding of plants, microbes, and biological communities through the integration of fundamental science and technology development. This program provides the foundation for biological solutions to DOE mission challenges, including energy, environment, and climate. The objectives of the program are:

- Determine the genomic properties, molecular and regulatory mechanisms, and resulting functional potential of microbes, plants, and biological communities central to DOE missions.
- Develop the experimental capabilities and enabling technologies needed to achieve a genome-based, dynamic systems-level understanding of organism and community functions.
- Develop the knowledgebase, computational infrastructure, and modeling capabilities to advance the understanding, prediction, and manipulation of complex biological systems.

This past year has been exciting and productive for the Genomic Science program. Our program continues to support groundbreaking research by individual investigators, interdisciplinary research teams, user facilities such as DOE's Joint Genome Institute and Environmental Molecular Sciences Laboratory (EMSL), and the Bioenergy Research Centers. Over the past year, researchers in our program have developed new methods to image, sequence, characterize, and engineer microbial systems to meet DOE missions in energy and the environment. This meeting highlights work in this area in the keynote address by Dr. Bernhard Palsson and in plenary sessions dedicated to Bioenergy Research and Systems Biology for Hydrogen Production.

The Genomic Science program depends on developments in fundamental research in technologies such as imaging, computing, and structural biology. In May 2009, DOE held the *New Frontiers in Characterizing Biological Systems* workshop, and this year our meeting will feature a breakout and talks highlighting the conclusions of the workshop as well as recent developments in imaging. Another strength of the Genomic Science program is our commitment to supporting research through outstanding DOE user facilities. Plenary talks at this meeting will highlight two BER-funded national user facilities: the Joint Genome Institute and EMSL capabilities in proteomics.

In addition to core DOE research programs, we have added new breakout sessions to discuss recent research advances in Genomic Structure and Epigenetics and in Synthetic Biology. Based on the success of last year, we will once again include lunchtime presentations by undergraduate and graduate research investigators. For this year's meeting, we are partnering with our sister BER division, the Climate and Environmental Sciences Division, to co-sponsor a breakout session highlighting approaches to integrate 'omics techniques to understand microbial activities influencing the fate and transport of metal and radionuclide contaminants.

The Biological Systems Science Division is holding a series of workshops focused on developing the DOE Systems Biology Knowledgebase. At this meeting, a plenary session and breakout workshop are dedicated to outlining the near-, mid-, and long-term objectives of the Knowledgebase. We hope you will join us as we map out the Knowledgebase as a cyber-infrastructure for systems biology information and data that not only includes data storage, retrieval, and management, but also enables new knowledge acquisition and management through free and open access to data, analysis tools, and information for the scientific research community.

We look forward to an exciting and productive meeting and encourage you to exchange ideas and share your expertise with other researchers. We thank you for lending your knowledge, creativity, and vision to the Genomic Science program and wish you continued success in the coming year.

Sincerely,

Sharlene C. Weatherwax, Ph.D.
Division Director
Biological Systems Science Division, SC-23.2
Office of Biological and Environmental Research
Office of Science

Contents

Agenda

Correct as of January 20, 2010

Even-numbered posters will be presented on Monday, odd-numbered posters on Tuesday. Please set up the posters no earlier than 5 p.m. on Sunday and leave them up until noon Wednesday.

Sunday, February 7

5:30pm	Early Registration and Poster Setup
6:00–8:30pm	Reception and Scientific Mixer

Monday, February 8

8:30–9:00am	**Welcome and State of Genomic Science Program**	
	8:30-8:45am	**Dr. Anna Palmisano**, Associate Director, Office of Biological and Environmental Research
	8:45-9:00am	**Dr. Sharlene Weatherwax**, Director, Biological Systems Science Division
9:00–11:00am	**Plenary Session: DOE Bioenergy Research Centers** **Moderator**: John Houghton	
	Speakers:	
	9:00–9:30am	**Dr. Robert Landick**, Dept. of Biochemistry, University of Wisconsin and Great Lakes Bioenergy Research Center (GLBRC) **Title:** Improving Bioenergy Microbes By Multidimensional Evolutionary and Functional Genomics
	9:30–10:00am	**Dr. Trent Northen**, Lawrence Berkeley National Laboratory and Joint BioEnergy Institute (JBEI) **Title:** High Throughput Multiplexed GlycoChip Enzymatic Assays for Biofuels Development
	10:00–10:30am	**Coffee Break**
	10:30–11:00am	**Dr. Lee Lynd**, Thayer School of Engineering, Dartmouth and BioEnergy Science Center (BESC) **Title**: Consolidated Bioprocessing: From Science to Impact
11:00am–12:00pm	**Keynote Presentation:** **Bernhard Palsson**, Dept. of Bioengineering, University of California, San Diego School of Engineering **Title:** Systems Biology of Metabolism	
12:30–2:00pm	**Lunchtime Student Oral Presentations** **Moderator:** Libby White	
2:00–5:00pm	**Breakout Sessions**	

Breakout Session A: Genomic Structure and Epigenetics
Moderator: Cathy Ronning

Description of Session: Epigenetic phenomena, heritable traits that occur in both prokaryotes and eukaryotes, result from chromatin modification without DNA sequence alteration and have been shown to contribute substantially to physiological processes such as tissue differentiation and alternative physiological states. The underlying mechanisms include DNA methylation, histone modifications, extracellular matrix composition, and regulatory RNA actions; however, others remain to be discovered. Epigenetic states are affected by the environmental history of a prior generation and may be passed on to subsequent generations, thus allowing for adaptation. Techniques such as ChIP-Seq, RNA-seq, and BS-seq allow measurements of epigenetic states on a genome-wide scale, and provide new insights in the roles and regulation of epigenetic modifications. This breakout session will focus on current developments in chromatin modification and structure.

Speakers:

2:00–2:25pm	**Dr. Zhiping Weng**, Dept. of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School **Title:** Developmental Regulation and Individual Differences of Neuronal Epigenomes in the Prefrontal Cortex
2:25–2:50pm	**Dr. Barbara Wold**, Division of Biology, California Institute of Technology
2:50–3:15pm	**Dr. Aleksandar Milosavljevic**, Dept. of Molecular and Human Genetics, Baylor College of Medicine **Title:** Genboree Open Hosting System for Collaborative Genomic and Epigenomic Research
3:15–3:40pm	**Dr. Matteo Pellegrini**, Dept. of Molecular, Cell and Developmental Biology, University of California, Los Angeles **Title:** Whole Genome DNA Methylation Profiles
3:40–4:00pm	**Coffee Break**
4:00–5:00pm	**Group Discussion**

Breakout Session B: Integrated Omics Approaches to Understand Environmental Processes
Moderator: Todd Anderson

Description of Session: Genome-enabled techniques are revolutionizing our understanding of microbial metabolism and are poised to enable an unprecedented understanding of microbial processes in the natural environment. This session will highlight genome-based techniques and approaches that are being employed to understand microbial processes of relevance to the DOE environmental mission. The session is co-sponsored by the Climate and Environmental Sciences Division within BER and is focused on, but not limited to, microbial activities influencing the fate and transport of metal and radionuclide contaminants. A central theme of this session is not only how genome-enabled studies provide in-depth knowledge of microbially mediated processes, but also how these approaches could lead to a more mechanistic integration of biology with the physical and chemical sciences in descriptions of environmental processes.

Speakers:

2:00–2:30pm	**Dr. Anne Summers**, Microbiology Dept., University of Georgia **Title:** The (Micro)biology of Mercury
2:35–3:05pm	**Dr. Eoin Brodie**, Lawrence Berkeley National Laboratory **Title:** Understanding and Predicting the Response of Soil Microbial Communities to Global Change
3:10–3:30pm	**Coffee Break**

3:30–4:00pm **Dr. Robert Hettich**, Oak Ridge National Laboratory
Title: Novel Experimental Approaches for Proteogenomic Characterizations of Microbial Soil Communities

4:05–4:35pm **Dr. Derek Lovley**, Dept. of Microbiology, University of Massachusetts
Title: BUGS Modeling: Predicting the Geochemical Consequences of Environmental Perturbations on Microbial Processes

4:40–5:10pm **Dr. Krishna Mahadevan,** Dept. of Chemical Engineering and Applied Chemistry, University of Toronto
Title: Integrated Modeling of Environmental Processes using Dynamic Genome-Based Models

Breakout Session C: Imaging and Technology for Systems Biology
Moderator: Arthur Katz

Description of Session: The Genomic Science program supports basic research aimed at achieving systems-level understanding of plants, microbes, and microbial communities relevant to DOE missions in bioenergy, carbon management, and environmental stewardship. While considerable progress has been made in developing imaging and analytical technologies over the last decade, we are just at the border of achieving the temporal and spatial resolution that will support comprehensive systems-level analyses. The biological challenge is to push our capabilities toward simultaneously measuring multiple chemical and biological species at multiple scales within complex, heterogonous cellular and environmental systems using the increasing temporal and spatial resolution of our technologies. This breakout session will identify and integrate the biological challenges and the technical capabilities needed to realize progress toward answering some of these pressing biological questions. The breakout will address the fundamental questions: Where do we need to go and How can we get there?

Speakers:

2:00–2:25pm **Dr. Alfred Spormann**, Dept. of Civil and Environmental Engineering, Sanford
Title: NanoSIMS Analysis of Carbon and Nitrogen Partitioning in the *Anabaena-Rhizobium* Microbial Consortium

2:25–2:50pm **Dr. Bruce Fouke**, Dept. of Geology, University of Illinois, Champaign-Urbana
Title: Yellowstone Hot-Spring Systems Geobiology: Quantification and Prediction of Mineral-Water-Microbe Feedback Interactions in the Subsurface

2:50–3:15pm **Dr. Jan Liphardt**, Lawrence Berkeley National Laboratory
Title: Super-Resolution Optical Methods for the Genomic Sciences: Possibilities and Problems

3:15–3:30pm **Coffee Break**

3:30–3:55pm **Dr. Jonathan Sweedler**, Dept. of Chemistry, University of Illinois, Urbana-Champaign
Title: Chemical Imaging using Mass Spectroscopy

3:55–4:20pm **Dr. Piero Pianetta**, Stanford Synchrotron Radiation Laboratory
Title: Applications of Full-Field Transmission X-ray Microscopy to Nano- and Thick Bio-Materials

4:20–4:45pm **Dr. Mitch Doktycz**, Oak Ridge National Laboratory
Title: Micro- and Nano-technologies for Studying the Plant-Microbe Interface

4:45–5:00pm **Closing Comments**

5:30–8:00pm **Poster Session**

Tuesday, February 9

8:30–10:00am **Plenary Session: Facilities**
Moderator: Dan Drell

- 8:30–9:00am **DOE Joint Genome Institute Update, Dr. Eddy Rubin,** Director, DOE Joint Genome Institute, Lawrence Berkeley National Laboratory
- 9:00–9:30am **EMSL Proteomics, Dr. Mary Lipton,** Pacific Northwest National Laboratory
  Title: Proteomic Insights into Microbes, Microbial Communities and Plants: An EMSL Perspective
- 9:30–10:00am **Break**

10:00am–12-:00pm **Plenary Session: Systems Biology Developments to Enable a Genomic Science Knowledgebase**
Moderator: Susan Gregurick

Speakers:

- 10:00–10:40am **Dr. Cathy Wu**, Dept. of Biological Sciences, University of Delaware
  Title: Systems Integration of Omics Data for Biological Network Discovery
- 10:40–11:20am **Dr. Nitin Baliga**, Institute for Systems Biology
  Title: A Knowledgebase for Rapid Inference and Re-engineering of Biological Circuits
- 11:20am-12:00pm **Dr. Harry J. Gilbert,** Complex Carbohydrate Research Center, University of Georgia
  Title: How Easy is it to Interpret Function from CAZy Family Location?

12:30–2:00pm **Lunchtime Student Oral Presentations**
Moderator: Libby White

2:00–5:00pm **Breakout Sessions**

Breakout Session D: Systems Biology Knowledgebase Workshop for Genomic Science Applications
Moderator: Susan Gregurick

Description of Session: This workshop will bring together researchers from the Genomic Science Community in microbial systems biology, computational biology and bioinformatics. The goal of this workshop is to outline the near-, mid-, and long-term trajectories of microbial sciences for energy and the environment. A second scientific objective is to map the associated workflows and data integration methods that can inform the specifications and requirements for the future development of the DOE Systems Biology Knowledgebase.

Speakers:

- 2:00–2:20pm **Dr. Robert Cottingham**, Oak Ridge National Laboratory
  Title: Microbial Systems Biology Knowledgebase: Scientific objectives and current prospects
- 2:20–2:50pm **Open Panel Discussion**
- 2:50–3:00pm **Coffee Break**
- 3:00–3:30pm **Dr. Robert Kelly**, Dept. of Chemical and Biomolecular Engineering, North Carolina State University
  Title: Near-Term Prospects for Functional Microbial Genomics: Moving Beyond the Monoculture Paradigm
- 3:30–4:00pm **Open Panel Discussion**
- 4:00–4:30pm **Dr. Adam Arkin**, Lawrence Berkeley National Laboratory
  Title: From Pathways to Populations and Back Again: Long Term Prospects for the Microbial Systems Biology Knowledgebase
- 4:30–5:00pm **Open Panel Discussions**

Breakout Session E: Synthetic Biology
Moderators: Marvin Stodolsky and Dan Drell

Description of Session: The Genomic Science program supports basic research aimed at achieving systems-level understanding of plants, microbes, and microbial communities relevant to DOE missions in bioenergy, carbon management, and environmental stewardship. The first result of genome sequencing efforts is the determination of the potential 'parts list' for an organism and can range from as few as 580 'parts' in a simple mycoplasm to many thousands in eukaryotes. The promise of synthetic biology is to treat these biological parts as building blocks and to assemble functional units from them. In addition to the opportunity to explore how nature has built functions from natural parts, synthetic biology offers a tool kit for the construction of new functions. This breakout session will explore current work using biological parts to build mission-relevant functions, to showcase student-initiated efforts to build novel biological activities, and to consider some of the social implications of this new technology.

Speakers:

2:00–2:30pm	**Dr. John Glass**, J. Craig Venter Institute **Title**: Synthetic Genomics: Progress on Construction of a Synthetic Bacterial Cell
2:30–3:00pm	**Dr. Nathan Hilson**, Lawrence Berkely National Laboratory **Title:** Integration of BioCAD tools, Parts Registries, and Automated-Assembly
3:00–3:30pm	**Dr. George M. Church**, Genetics Dept., Harvard Medical School **Title**: Genome Engineering, Multi-Virus Resistance, and Accelerated Evolution for Industrial Chemicals
3:30–3:50pm	**Coffee Break**
3:50–4:10pm	**iGEM, Cornell Team**, Alyssa Henning **Title:** Engineering the *Bacillus subtilis* Metal Ion Homeostasis System to Serve as a Cadmium-Responsive Biosensor
4:10–4:30pm	**iGEM, Utah Team**, Libbie Linton and faculty advisor Dr. Charles Miller, Biological and Irrigation, Utah State University **Title:** BioBricks without Borders: Investigating a Multi-Host BioBrick Vector and Secretion of Cellular Products
4:30–5:00pm	**Dr. Lori Knowles**, Health Law Institute, University of Alberta **Title:** Synthetic Genomics: Oxymorons and Communication Bloopers – Why Researchers' Enthusiasm for "Tinkering with Life" Might Not be Contagious

Breakout Session F: Structural Biology Applications in Genomic Science
Moderator: Roland Hirsch

Description of Session: This session will provide examples of how the Department of Energy's synchrotron light sources and neutron facilities are of value for Genomic Science research projects. Each talk will focus on a biological problem within the scope of the Genomic Science program. The speakers will define the problem, then explain the selection of the particular structural biology technique, and finally discuss how the technique has helped advance the research project. The emphasis will be on the biology rather than the technology. The techniques to be included in the talks in this session are macromolecular crystallography with X-rays and neutrons, small-angle X-ray and neutron scattering and infrared and X-ray imaging.

Speakers:

2:00–2:25pm	**Dr. Marianne Schiffer**, Argonne National Laboratory **Title:** Structural Studies of Heme Containing Proteins from *Geobacter sulfurreducens* using the Structural Biology Center Beam Lines at the Advanced Photon Source
2:25–2:50pm	**Dr. Paul Langan**, Los Alamos National Laboratory **Title:** Using Neutron Crystallography to Reveal the Mechanism of Enzymes for Renewable Energy and the Environment
2:50–3:15pm	**Dr. Greg Hura**, Lawrence Berkeley National Laboratory **Title:** Developed Proteomics Scale Solution X-ray Scattering (SAXS) Tools Applied to Metabolic Networks of Interest from the ENIGMA Program Project
3:15–3:30pm	**Coffee Break**
3:30–3:55pm	**Dr. Blake Simmons**, Lawrence Berkeley National Laboratory **Title:** Novel Insights into the Structural Mechanism of Lipid Accumulation within Algae using Soft X-ray Tomography
3:55–4:20pm	**Dr. Hoi-Ying Holman**, Lawrence Berkeley National Laboratory **Title:** Real-Time Monitoring of Chemical Environment in Cells during Stress-Adaptive Response
4:20–4:45pm	**Dr. Barbara Evans**, Oak Ridge National Laboratory **Title:** Changes in Switchgrass Structure during Acid Pretreatment Examined with Small-Angle Neutron Scattering and Wide-Angle X-ray Diffraction
4:45–5:00 pm	**Closing remarks and discussion**

5:30–8:00pm **Poster Session**

Wednesday, February 10

8:30–9:30am **Plenary Session**

8:30–9:00am	**Imaging Workshop Report:** Arthur Katz
9:00–9:30am	**Dr. Jennifer Lippincott-Schwartz**, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health **Title**: Insights into Organelle Biogenesis and Near Molecule Protein Distribution Using Photoactivatable Fluorescent Protein Technology

9:30–10:00am **Coffee Break**

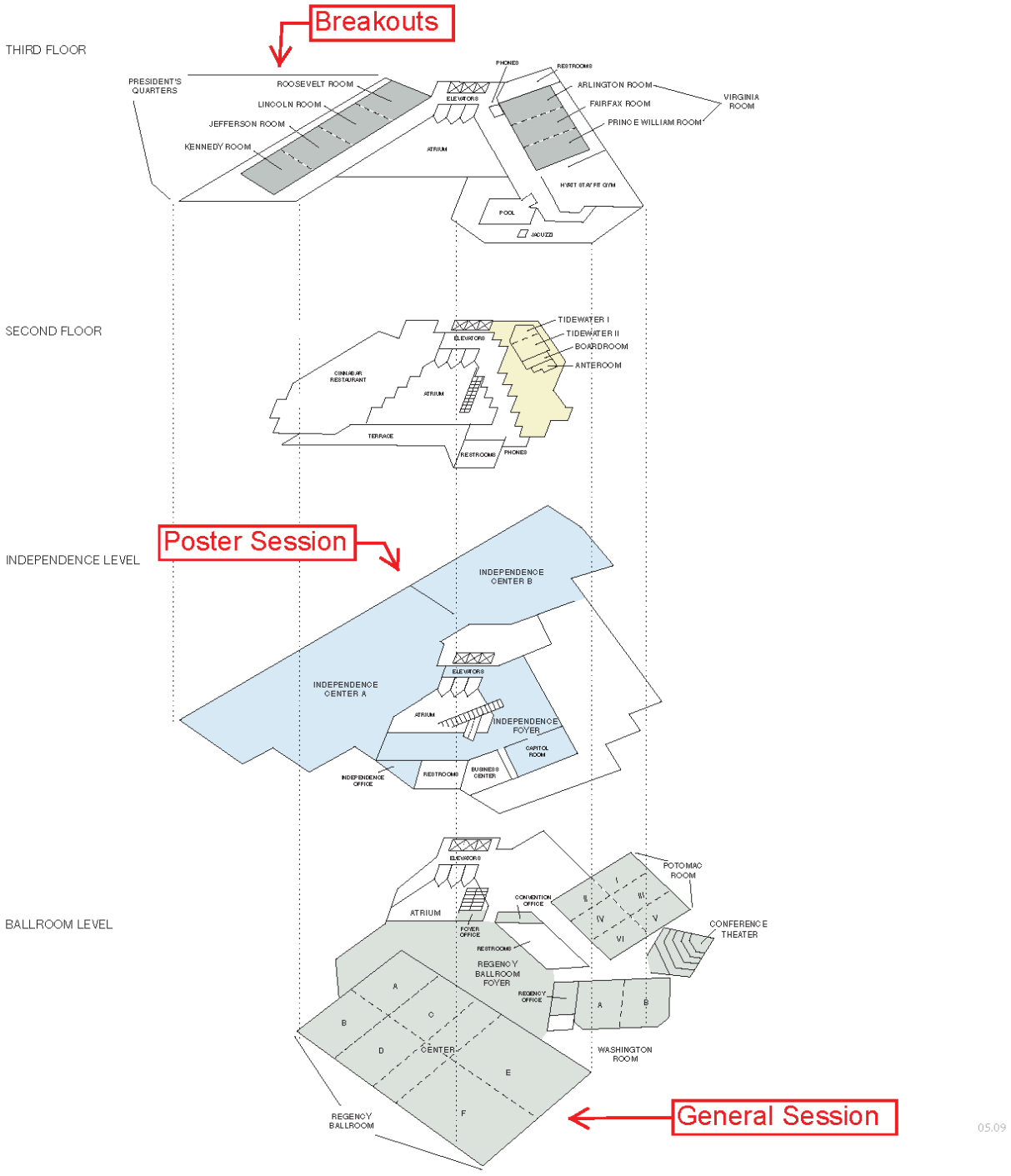
10:00–12:00pm **Plenary Session: Systems Biology for H_2 Production**
Moderator: Joe Graber

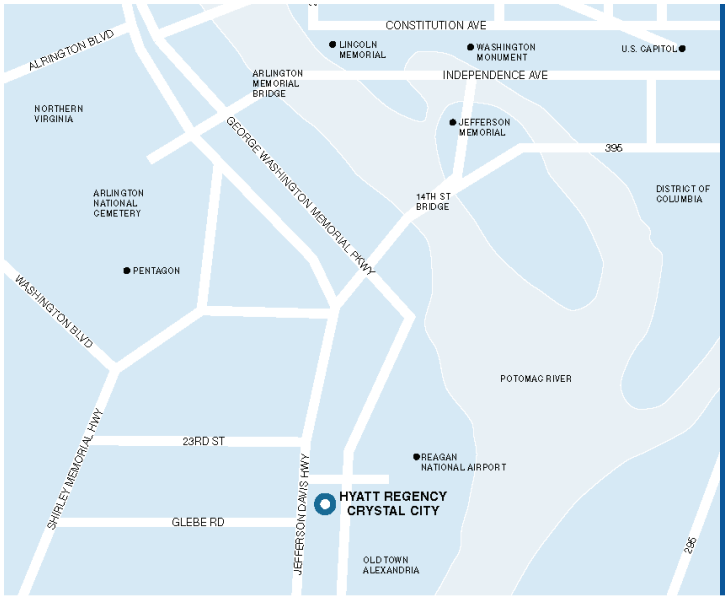
Speakers:

10:00–10:40am	**Dr. Josh Rabinowitz,** Lewis-Sigler Institute for Integrative Genomics, Princeton University **Title:** Metabolomic Data Reveals Pathways and Regulatory Principles
10:40–11:20am	**Dr. Caroline Harwood,** Dept. of Microbiology, University of Washington **Title:** Protons as Electron Acceptors: What Can This Teach Us About Bioenergy Generation?
11:20-12:00pm	**Dr. Judy Wall,** Dept. of Biochemistry, University of Missouri **Title:** Pathway of Fermentative Hydrogen Production by Sulfate-Reducing Bacteria

12:00pm **Closeout and Adjournment**

Hyatt Regency Crystal City Map





Hyatt Regency Crystal City

At Reagan National Airport

DIRECTIONS

From Reagan National Airport (1/2 mile): Follow signs to Crystal City. Take the Rt. 1 South exit and get in the left hand lane. Turn left at first light, 27th Street, the hotel is on the left. From Dulles Airport (32 miles): Take I-66 East to Exit 75 (Rt. 110 South/turns into Rt. 1 South). Proceed on Rt. 1 to 4th light (27th Street). Turn left onto 27th Street. Hotel is on left.

Abstract Titles

Biofuels: Analytical and Imaging Technologies for Studying Lignocellulosic Material Degradation

Systems Biology and Metabolic Engineering Approaches for Biological Hydrogen Production

Molecular Interactions, Protein Complexes, and Structural Biology

Validation of Genome Sequence Annotation

Computing for Systems Biology

Workshop Abstracts

Abstract Organization

Genomic Science program abstracts and posters are organized according to the following research areas important to achieving the program's ultimate scientific goal and objectives. Even-numbered posters will be presented on Monday, odd-numbered posters on Tuesday. Please set up the posters no earlier than 5 p.m. on Sunday and leave them up until noon Wednesday.

Systems Biology for DOE Energy Missions: Bioenergy

Bioenergy Research Centers

- Great Lakes Bioenergy Research Center (GLBRC)
- Joint BioEnergy Institute (JBEI)
- BioEnergy Science Center (BESC)

Biofuels: Analytical and Imaging Technologies for Studying Lignocellulosic Material Degradation

Systems Biology and Metabolic Engineering Approaches for Biological Hydrogen Production

Computing for Bioenergy

Small Business Innovation Research (SBIR)

Systems Biology for DOE Environmental Missions: Systems Environmental Microbiology

Systems Biology Strategies and Technologies for Understanding Microbes, Plants, and Communities

Analytical Strategies for the Study of Plants, Microbes, and Microbial Communities

- Biological Systems Interactions
- Plant-Microbe Interfaces
- The Predictive Microbial Biology Consortium

Molecular Interactions, Protein Complexes, and Structural Biology

Validation of Genome Sequence Annotation

Computing for Systems Biology

Communication and Ethical, Legal, and Societal Issues

Genomic Science Goal and Objectives

Ultimate Scientific Goal
Achieve a predictive, systems-level understanding of plants, microbes, and biological communities, via integration of fundamental science and technology development, to enable biological solutions to DOE mission challenges in energy, environment, and climate.

Objective 1: Determine the genomic properties, molecular and regulatory mechanisms, and resulting functional potential of microbes, plants, and biological communities central to DOE missions.

Objective 2: Develop the experimental capabilities and enabling technologies needed to achieve a genome-based, dynamic systems-level understanding of organism and community functions.

Objective 3: Develop the knowledgebase, computational infrastructure, and modeling capabilities to advance the understanding, prediction, and manipulation of complex biological systems.

The following table is a summation of how the Genome Science program and DOE missions align (*DOE Genomics:GTL Roadmap: Systems Biology for Energy and Environment*; October 2005, p. 40). (GenomicScience.energy.gov)

Summary Table. GTL Science Roadmap for DOE Missions

	DOE Mission Goals		GTL Science Roadmaps	
Selected Processes	Biofuels	**Processes to convert cellulose to fuels** • Understanding and improving cellulase activity • Improving sugar transportation and fermentation to alcohols • Integrated processing **Microbial processes to convert sunlight to hydrogen fuels** • Understanding photolytic fuel production • Designing photosynthetic biofuel systems	Science Objectives	▸ **Characterize genes, proteins, machines, pathways, and systems** • Conducting genomic surveys and comparisons • Mining natural systems for new functions • Producing and characterizing proteins • Analyzing interactions, complexes, and machines ▸ **Understand functions and regulation** • Measuring molecular responses: Inventories • Performing functional assays ▸ **Develop predictive mechanistic models** • Conducting experimental design • Designing and manipulating molecules • Using cellular and cell-free systems
	Environmental Remediation	**Microbial processes to reduce toxic metals** • Understanding microbe-mineral interactions • Devising restoration processes	Mission Outputs	**Systems engineering** • System-design strategies for deployment • Living and extracellular systems • Validation and verification analyses
Natural Systems' Behavior	Environmental Remediation	**Subsurface microbial communities' role in transport and fate of contaminants** • Understanding fate and effects • Supporting remediation decisions	Science Objectives	▸ **Analyze communities and their genomic potential** • Sequencing and comparing genomes • Screening natural systems for processes • Producing and characterizing proteins ▸ **Understand community responses, regulation** • Comparing CO_2, nutrients, biogeochemistry cycles • Producing cellular and community molecular inventories • Performing community functional assays ▸ **Predict responses and impacts** • Building interactive and predictive models • Applying natural and manipulated scenarios
	Carbon Cycling and Sequestration	**Ocean microbial communities' role in the biological CO_2 pump** • Understanding C, N, P, O, and S cycles • Predicting climate responses • Assessing impacts of sequestration **Terrestrial microbial communities' role in global carbon cycle** • Understanding C, N, P, O, and S cycles • Predicting carbon inventories and climate responses • Assessing sequestration concepts	Mission Outputs	**Robust science base for policy and engineering** • Model ecosystem response to natural events • Efficacy and impacts of intervention strategies **Sensor development** • Community dynamics • Environmental and functional assays

A capsule summary of systems being studied, mission goals that drive the analysis, generalized science roadmaps, and outputs to DOE missions. To elucidate design principles, each of these goals entails the examination of thousands of natural primary and ancillary pathways, variants, and functions, as well as large numbers of experimental mutations.

Systems Biology for DOE Energy Missions: Bioenergy

Bioenergy Research Centers

Great Lakes Bioenergy Research Center (GLBRC)

1

Discovery of Novel Genes Regulating Polysaccharide Biosynthesis and Secretion in Plants

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

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

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

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

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

2

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

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

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

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

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

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

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

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

3

Generation and Analysis of Transgenic Poplars with Altered Wall Compositions

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

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

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

Wood is gaining popularity as a source of fermentable sugars for liquid fuel production. However, our current knowledge on the genetic control of woody biomass formation is limited. Secondary wall of wood consists of a complex mixture of cellulose, hemicellulose, and lignin. Proportional variability within the mixture of the three major components varies depending on the species of feedstock used, growing site, climate, age and the part of the plant harvested. The essentially uncontrolled variability of biomass properties presents process design and operating challenges for the production of bioenergy from woody feedstocks. Better understanding of the molecular mechanisms underlying its biosynthesis will help us develop biotechnological means to genetically control key pathways that determine the quantity and quality of the biomass.

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

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

* Presenting author

4

Transgenic Poplars with Altered Lignins for Improved Biomass Pretreatment and Saccharification

Hoon Kim[1]* (hoonkim@wisc.edu), Fachuang Lu,[1] Ali Azarpira,[1] Jorge Rencoret,[1] **John Ralph**,[1] Bruce E. Dale,[2] Curtis G. Wilkerson,[2] Nick Santoro,[2] Shane Cantu,[2] Cliff Foster,[2] Rebecca Van Acker,[3] Bart Ivens,[3] Wout Boerjan,[3] and Shawn D. Mansfield[4]

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

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

Introduction

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

F5H-upregulated Poplar

A lignin compositional change particularly targeted for improved alkaline pulping also results in a significant improvement in CW digestibility. F5H upregulation, driven by an appropriate promoter (C4H), produces lignins derived primarily from sinapyl alcohol; the Mansfield group has generated plants strikingly rich in syringyl units.[5] As a result, the lignin chains are rather homogeneous, being composed essentially only from one resinol unit (from initial dimerization) and β-ether-linked units. The lignins also have only a low degree of polymerization (and are therefore low-molecular-weight). These and other factors result in cell walls that saccharify more efficiently following acidic or basic pretreatment methods[6] – see Figure 1. We are currently evaluating the effects of the Ammonia Fiber Expansion (AFEX) pretreatment.

CCR-deficient Poplar

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

Conclusions

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

References

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

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

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

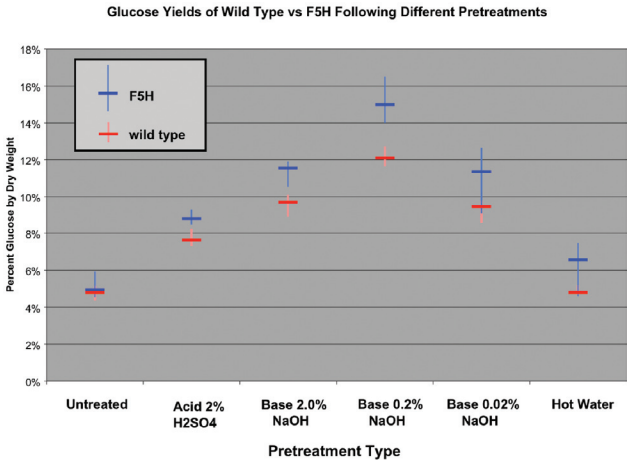


Figure 1. Comparison of saponification glucose yields following various pretreatments; except for untreated material, the F5H transgenic produced significantly higher yields.

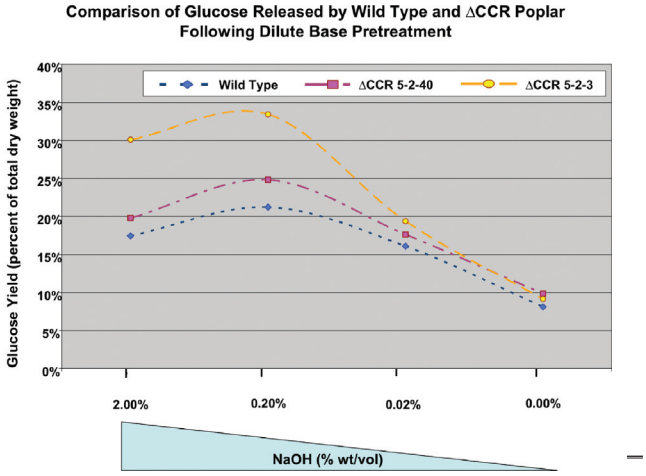


Figure 2. Glucose release for two CCR-deficient transgenics and their controls, following base pretreatment. The most CCR-deficient line has markedly elevated glucose release.
Note: in both cases, non-exhaustive saccharification conditions were used allow differences to be meaningfully represented.

5

Biomass Trait Screening in *Brachypodium* Accessions and Mutant Populations

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

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

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

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

* Presenting author

6
EST-SSR Markers Discriminate Switchgrass Ecotypes

M.D. Casler[1]* (michael.casler@ars.usda.gov), J.E. Zalapa,[2] S.M. Kaeppler,[2] N. de Leon,[2] and C.M. Tobias[3]

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

http://www.greatlakesbioenergy.org

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

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

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

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

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

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

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

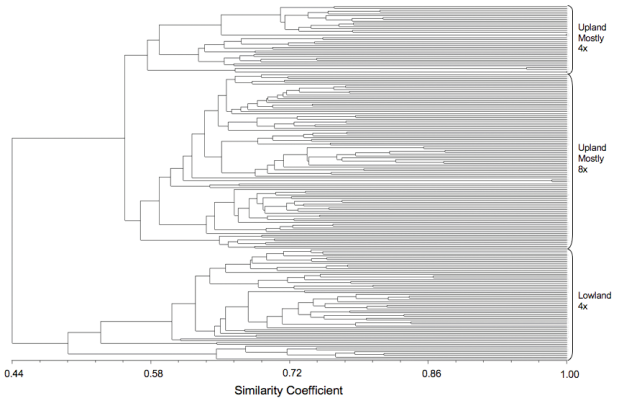


Figure 1. Cluster dendrogram of 192 switchgrass plants of known origin (upland 4x, upland 8x, and lowland 4x) grouped by similarity coefficients based on 381 EST-SSR markers.

7

Characterizing the Microbiome of Leaf-Cutter Ant Fungus Gardens

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

[1]U.S. Department of Energy (DOE) Great Lakes Bioenergy Research Center and [2]Dept. of Bacteriology, University of Wisconsin, Madison; [3]Smithsonian Tropical Research Institute, Balboa, Ancon, Republic of Panama; [4]DOE Joint Genome Institute, Walnut Creek, Calif.; [5]Departamento de Bioquímica, Facultad de Medicina and [6]Centro de Investigaciones en Estructuras Microscópicas, Universidad de Costa Rica, Ciudad Universitaria Rodrigo Facio, San José; [7]Dept. of Biochemistry and Molecular Biology and [8]DOE Plant Research Laboratory, Michigan State University, East Lansing; [9]Dairy Forage Research Center, U.S. Department of Agriculture–Agricultural Research Services, Madison, Wis.; [10]DOE Joint Genome Institute, Los Alamos National Laboratory, Los Alamos, New Mexico; [11]454 Life Sciences, a Roche Company, Branford, Conn.; and [12]Roche Diagnostics, Roche Applied Science, Indianapolis, Ind.

Project Goals: Understanding efficient plant biomass degradation is a critical step toward the technological goal of ethanol production. Microbes are the primary organisms on earth capable of deconstructing lignocellulose, and herbivores can gain access to organic carbon stored in plant cell walls by forming symbiotic relationships with lignocellulolytic microbes. Our project aims to characterize an insect herbivore, leaf-cutter ants, which have been farming a fungus for food for ~10 million years. These ants, one of the most widespread and dominant herbivores in the Neotropics, are capable of harvesting a tremendous amount of leaf-material to grow their fungus. They form massive colonies consisting of hundreds of fungus garden chambers supporting millions of workers. We are utilizing a combination of culture-independent and culture-dependent approaches to characterize the microbiome of leaf-cutter ant fungus gardens, and thereby understand how this microbial community synergistically degrades plant biomass.

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

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

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

* Presenting author

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

8

Optimization of Enzymes for Alkaline-Pretreated Biomass Conversion

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

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

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

Enzymes that release fermentable sugars from biomass feedstocks are one of the major costs in converting lignocellulose to ethanol. The major goal of this project is to build optimal enzyme mixtures for alkaline-pretreated biomass. This project has two stages of development. The first is to define an optimized "minimal enzyme set" composed of those enzymes that are almost certainly essential for lignocellulose degradation. This set includes exo-glucanases (cellobiohydrolases; CBHs), endo-glucanase (EG), endo-xylanase (EX), β-xylosidase (BX), and β-glucosidase (BG). The second part is to develop an "optimized enzyme mixture" in which additional enzymes, called here "accessory" enzymes, are added to the minimal set in order to create a superior mixture. By optimum we mean having the highest specific activity (lowest protein loading) to achieve a benchmark release (e.g., 85%) of glucose and xylose in a specified time (48 hr). We used Ammonia Fiber Expansion (AFEX) pretreated corn stover for this work. The enzymes for these experiments came from several sources. These include enzymes purified from commercial *Trichoderma reesei* preparations, *Trichoderma* genes expressed in *Pichia pastoris*, *Trichoderma* proteins over expressed in *Trichoderma* itself and *A. nidulans* orthologs of the *Trichoderma* enzymes expressed in *Pichia pastoris*. Another important source of enzymes for our experiments are proteins, mainly bacterial, from other projects in the GLBRC. We have successfully completed the first stage of the project by defining the core enzymes needed to achieve the benchmark glucan and xylan conversions in 48 hr of hydrolysis. The second phase of the project using accessory enzymes along with the core enzymes is in progress. Preliminary results indicate that there is synergy operating between bacterial hemicellulases and fungal cellulases. When completed, the optimized mixture is predicted to have higher specific activity than current commercial mixtures, enabling a reduced enzyme load during hydrolysis.

9

Physicochemical Characterization of Alkali-Pretreated Lignocellulosic Biomass

Shishir Chundawat,[1,3] Leonardo Sousa,[1,3] James Humpula,[1,3] Venkatesh Balan,[1,3] **Bruce Dale**[1,3]* (bdale@egr.msu.edu), Ramin Vismeh,[2,3] A. Daniel Jones,[2,3] Fachuang Lu,[4,5] Lily Summer,[4,5] Ali Azarpira,[4,5] John Ralph,[4,5] Chris Bianchetti,[4,5] George Phillips,[4,5] Bryon Donohoe,[6] Michael Himmel,[6] Umesh Agarwal,[7] Paul Langan,[8] S. Gnanakaran,[9] and Albert Che[10]

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

Project Goals: See goals for abstract 8.

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

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

10

Combinatorial Discovery of Enzymes for Biomass Deconstruction

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

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

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

The goal of our work is to provide a new combinatorial paradigm for evaluation of enzymes from new environmental sources as well as synthesized genes and engineered enzymes. The GLBRC bioenergy platform derives from work with cell-free translation at the NIH Protein Structure Initiative-funded Center for Eukaroytic Structural Genomics, where over 10,000 genes from various eukaryotic organisms have been cloned, tested for expression, and in the best performing cases, purified and subjected to structure determinations. An adaption of the modular vector design of this platform provides the basis for this new effort on genes and proteins contributing to cellulose destruction. Genes identified by bioinformatic evaluation of new genomes or by microarray evaluation of gene expression in cellulose utilizing organisms, and proteins identified by mass spectral proteomic studies can be targeted for high-throughput cloning and cell-free translation. Methods to prepare combinations of genes or to supplement existing mixtures of enzymes with new translation products are demonstrated to provide biomass deconstruction without need for purification of translation products. Iterative substitution of gene variants can be used to identify protein homologs with better behavior in defined assays that can include diagnostic small molecules, model purified celluloses, or actual biomass substrates. Results of the application of these methods to characterization of the reactivity of various treated and untreated biomass materials will be presented.

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

11

Construction of Gram-Negative Consolidated Bioprocessors

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

Great Lakes Bioenergy Research Center, University of Wisconsin, Madison

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

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

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

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

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

12

Modeling and Summarizing Growth Curves from High-Throughput Screening Data

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

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

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

We have developed R scripts to process and summarize our screening data. We chose to fit theoretical functions, instead of using numerical differentiation or local regression methods, in order to be able to work with limited number of data points, which maintains maximal robotic screening throughput. We have tested several known growth curve models and found that the Gompertz function[1,2] gives the best results with our data. We have also developed methods for automatic generation of initial guesses of curve parameters and for dealing with growth curves that have anomalous shapes. Upon processing the screening data, our software generates a table of biologically meaningful growth curve characteristics, such as the fastest intrinsic growth rate, lag time, and total growth. These characteristics were used to select yeast strains with optimal growth properties in the examined media. Although the development of this methodology was motivated by the needs of the yeast screening program at GLBRC, it has general utility for high-throughput screening studies of microbial strain collections.

References

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

13

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

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

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

Project Goals: see below

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

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

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

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

14

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

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

Great Lakes Bioenergy Research Center, University of Wisconsin, Madison

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

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

15

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

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

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

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

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

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

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

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

[a]Enterohemorrhagic *E. coli* (EHEC)
[b]Uropathogenic *E. coli* (UPEC)
[c]Enteroinvasive *E. coli* (EIEC)
[d]Avian pathogenic *E. coli* (APEC)
[e]Enteropathogenic *E. coli* (EPEC)
[f]Enterotoxigenic *E. coli* (ETEC)

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

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

[a]Reactions with no orthologous ORF (s), but are essential to the *in silico* model

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

16

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

Joonhoon Kim[1,2]* (jkim92@wisc.edu), Dipak Barua,[1,2] and **Jennifer L. Reed**[1,2]

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

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

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

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

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

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

17

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

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

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

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

Although commonly used for the industrial production of grain ethanol, *S. cerevisiae* has a number of biological hurdles that currently prevent it from being widely utilized in the generation of fuel ethanol from lignocellulosic biomass. Cellular and physiological stresses imposed by degradation products from feedstock pretreatment are known to limit the yield and efficiency of fermentation. Well-known degradation products include acetic acid, furfurals and lignin-derived phenolic compounds, all of which can vary in structure and concentration with different pretreatment processes.

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

18

Efforts to Enhance Solar Hydrogen Production by Heterocyst-Forming Cyanobacteria

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

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

http://www.greatlakesbioenergy.org/

Project Goals: The overall goal of our project is to engineer *Anabaena* to enhance photobiological H_2 production to a commercially practical level. To accomplish this task we are using a number of different strategies including mutating nitrogenase to produce more H_2, expressing a native bidirectional [NiFe]-hydrogenase in heterocysts, and heterologously expressing an [FeFe]-hydrogenase in heterocysts. In addition, to guide genetic engineering and to enhance further H_2 production, we are elucidating the

* Presenting author

pathways by which electrons are channeled to H_2 and the major pathways that compete for those electrons.

H_2 has the potential to become an important clean and renewable energy commodity, especially if it is generated by organisms such as cyanobacteria that use sunlight as the sole energy source and water as the ultimate electron donor. Hydrogenases (H_2ases) and nitrogenases (N_2ases) are the enzymes that produce H_2. Because these enzymes are also O_2-sensitive, oxygenic photosynthesis and H_2 production are normally separated temporally or spatially. *Anabaena* spp. and related cyanobacteria form specialized cells known as heterocysts in which N_2ases and H_2ases are protected from O_2 by inactivation of O_2-producing PSII, accelerated respiration, and synthesis of a thick envelope of glycolipids and polysaccharides that impedes O_2 penetration. Reductant required for N_2 fixation and H_2 production is generated by photosynthesis in vegetative cells and is transported to heterocysts as sugar. Our ultimate goal is to engineer a Hup^- strain of *Anabaena* (in which the uptake hydrogenase is inactive) to increase H_2 production to a commercially practicable level.

To increase H_2 production by N_2ase in *Anabaena*, we mutated residues near the active site, generating a total of 49 single variants. Several mutants exhibited significantly higher H_2 production rates in the presence of N_2 than did the parental Hup^- strain. Nineteen additional single and double site-directed mutants designed to obstruct a putative channel connecting the active site to the protein surface did not exhibit significantly higher H_2 production rates than did the parental Hup^- strain.

Two different strategies are being pursued to express H_2ases in *Anabaena* heterocysts. In the first strategy, efforts are under way to overexpress the native, bidirectional [NiFe]-H_2ase (Hox) genes in heterocysts using the strong *nif* promoter. Genes required for the maturation of Hox are being expressed on a replicating plasmid from the heterocyst-specific promoters of *coxBII* and *patB*. In the second strategy, [FeFe]-H_2ases and the proteins needed for their maturation are being heterologously expressed in *Anabaena* heterocysts using both integrating and replicating plasmids driven by heterocyst-specific promoters. To obtain optimal expression and activity, we are testing different combinations of [FeFe]-H_2ases and maturation proteins from a variety of organisms including *Chlamydomonas reinhardtii, Clostridium acetobutylicum, and Shewanella oneidensis*.

To elucidate how electrons are channeled to H_2 production, we compared gene expression in phototrophic, photoheterotrophic, and heterotrophic *Anabaena* cultures using RNA extracted separately from vegetative cells, heterocysts, and whole filaments. Principal component analysis of gene expression data confirmed that gene expression patterns in the vegetative cells differ from those in the heterocysts. 19%, 17%, and 16% of the genes are differentially expressed (at least 2X difference, $p<0.01$) between vegetative cells and heterocysts in phototrophic, photoheterotrophic, and heterotrophic growth conditions, respectively. In addition, gene expression patterns in vegetative cells vary in the different growth conditions, consistent with the fact that vegetative cells are solely responsible for carbon uptake. Surprisingly, Hup genes and several N_2ase maturation genes are expressed in the vegetative cells in both phototrophic and photoheterotrophic cultures. These and other results will be discussed.

19

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

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

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

Project Goals: We seek to understand and improve light- and feedstock-powered production of renewable fuels by the photosynthetic bacterium *Rhodobacter sphaeroides*. We have initially chosen to investigate hydrogen (H_2) production, both to optimize the production of H_2 itself and to serve as a model for partitioning of intracellular reducing potential (high-energy electron equivalents) utilized in microbial biofuel production in general. We are characterizing the intracellular pathways involved in distributing electrons throughout the cell via genome-enabled assays, such as microarray and proteomics analyses, coupled with biomass, polyhydroxybutyrate (PHB) and chemical oxygen demand (COD) analyses. We are using mutant strains to determine the effects of deleting genes predicted by the global gene expression assays to impact the intracellular reducing potential available for fuel production.

Rhodobacter sphaeroides is a photosynthetic purple non-sulfur bacterium that can accumulate a large pool of intracellular reducing potential (high energy electrons) during photoheterotrophic growth on various carbon sources, such as organic acids and biomass-derived sugars. We are studying the distribution of this reducing potential in the interest of increasing the fraction that goes toward fuel production. We are initially focusing on H_2 production, for its own utility as a fuel and in industrial purposes, and as a model to understand how electrons needed for production of other fuels will be portioned to other pathways.

We have obtained global transcript levels in cultures with various levels of H_2-production and compared them to those in non-H_2 producing control cultures to determine candidate gene products that contribute electrons to or siphon reductant from H_2 production. We also compared transcript levels in H_2-producing cultures grown on various carbon sources to characterize the relative expression levels of the genes involved in cellular pathways that act as electron sinks. In addition, we are assaying these same cultures for the

distribution of electrons between the cellular end-products, PHB, other biomass, H_2, and soluble microbial products.

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

- Nitrogenase is the primary (or sole) source of H_2, produced as a byproduct of nitrogen fixation. As expected, expression of the nitrogenase structural genes is increased in all cultures that produce detectable H_2. We are testing whether expression level correlates to total H_2 production amounts; preliminary evidence suggests that cellular nitrogenase enzymatic activity does correlate with the amount of H_2 produced.
- Hydrogenase is expected to oxidize H_2 and recycle reducing potential back into cellular metabolism. We find that expression of the hydrogenase structural genes varies inversely with net H_2 production; cultures that produce relatively low amounts of H_2 (those using xylose, glucose or glycerol) have relatively high expression of hydrogenase genes, while cultures that produce relatively high amounts of H_2 (those using lactate or succinate) have relatively low expression of hydrogenase genes, which suggests that the presence of functional hydrogenase may contribute to the disparities in the amounts of H_2 produced from different carbon sources. Experiments are in progress to determine the level of functional hydrogenase enzyme made in these cultures and how the reducing potential is distributed in defined hydrogenase mutants.
- Carbon dioxide fixation consumes reducing power, so it has the potential to siphon electrons from fuel production. We find that expression of carbon fixation genes is also inversely proportional to H_2 production (as is the case for hydrogenase genes), which suggests that carbon dioxide fixation may contribute to the disparities in H_2 amounts from cultures grown on different carbon sources. *R. sphaeroides* has two sets of carbon fixation genes, and we are investigating mutants which only contain one gene from each set (*cfxA*- and *cfxB*-) to determine the relative importance of the individual pathways on H_2 production.
- *R. sphaeroides* can accumulate the polymer PHB as an energy and carbon storage compound. We find little difference in expression of PHB synthase genes in H_2-producing cultures grown on different carbon sources. However, we find that the amounts of PHB produced differ between such cultures, and we are testing the effect of deleting the PHB synthase gene on H_2 production.

20

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

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

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

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

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

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

* Presenting author

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

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

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

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

21

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

John Ohlrogge[1]* (Ohlrogge@msu.edu), Timothy Durrett,[1] Adrian Troncosco,[1] Jilian Fan,[1] Xia Cao,[1] Nick Thrower,[1] Erika Lindquist,[2] Christa Pennacchio,[2] Curtis Wilkerson,[1] and Mike Pollard[1]

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

http://www.greatlakesbioenergy.org/

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

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

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

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

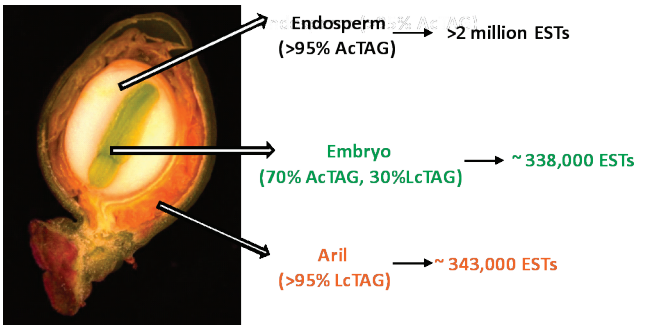


Fig 1. Euonymus fruit produce ac-TAGs in a tissue specific manner. The endosperm and embryo tissues of the Euonymus seed accumulate high levels of ac-TAGs whereas the aril tissue surrounding the seed synthesizes only lc-TAGs. Transcript profiles were obtained of these different tissues to isolate the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

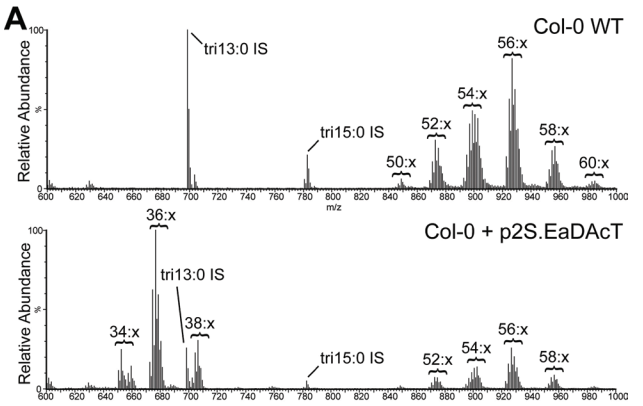


Fig 2. *Arabidopsis* seeds expressing EaDAcT accumulate ac-TAGs. Positive-ion ESI mass spectra of neutral lipid extracts from Col-0 wildtype seed or T_3 seed from a representative Col-0 plant expressing EaDAcT. Peaks correspond to *m/z* values of the $[M + NH_4]^+$ adduct. Tritridecanoin (tri13:0) and tripentadecanoin (tri15:0) were added as internal standards. The number of acyl carbons in each series of TAG molecules is indicated.

22

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

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

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

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

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

* Presenting author

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

23

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

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

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

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

Biofuels are attractive for economic, environmental, and strategic reasons. Reducing our dependency on foreign oil is a key national security issue. Although corn and soybeans are excellent sources of biofuels, cellulosic feedstocks clearly are the wave of the future. Poplar, switchgrass, *Miscanthus*, and even managed prairie ecosystems are potential sources of cellulosic biofuels. If we can solve the problems of effectively transforming this biomass into cellulosic ethanol or other fuels, there will be significant environmental positive gains in reducing emissions of greenhouse gases and other pollutants. However, concerns have been raised about the sustainability of cellulosic cropping systems. Biofuel production systems based on annual grains are the most straightforward for growers but may not be the most productive or sustainable in the long term. Understanding the basis for sustainable biofuel production systems is crucial for the long-term success of these systems. The rapid growth, low mineral content and high biomass yield of cellulosic crops make them a favored feedstock choice. But the important question remains: can we produce enough biomass to sustain the biofuel industry without compromising environmental security. The biomass production potential of these crops is directly linked to concerns related to land-use change, environmental degradation, and food security. To answer these questions, field experiments are underway at several scales in Michigan and Wisconsin. Eight model cropping systems are studied for productivity, carbon balance, and biodiversity impacts. A major goal of this initiative is to test and develop biofuel crops that provide both high yields and environmental benefits such as greenhouse gas mitigation, clean water, and pest protection. Our aim is to provide a comprehensive portrait of the production potential of different cellulosic biofuel crops as well as their environmental impacts and benefits. We use field results to parameterize and test quantitative models that can then be used to simulate crop growth across larger regions.

24

Metagenomics of Bacterial Communities from the Rhizosphere of Switchgrass

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

Michigan State University, East Lansing

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

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

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

25

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

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

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

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

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

There is currently a significant research effort to develop sustainable biofuels from cellulosic plant materials. To be broadly sustainable, this research utilizes a diverse array of modern scientific methods, and is very interdisciplinary and collaborative in nature. Given that this work is in the public eye, there are significant opportunities to engage learners in both the details of bioenergy; and importantly, into the underlying scientific principles of biogeochemical and energy systems. Assessments of student understanding, however, reveal many consistent misconceptions that hinder students' ability to comprehend these systems.

A significant focus of our work is to develop 1) a further categorization of the range of understandings related to carbon cycles and energy flow, and 2) K-16 educational materials that will use biofuels as an entrée to engage learners in a more accurate comprehension of these basic scientific concepts. Specific content at the base of these materials includes life-cycle assessments and systems thinking. The effects of biofuel production on biodiversity, ecosystem services, climate change, and global energy dynamics are considered as well. A number of persistent difficulties are seen in assessments of student understanding at all levels. Many students, for instance, do not connect the decomposition of plants and animals to CO_2 in the atmosphere. In response to this and related patterns, we are designing a series of activities to make connections to fundamental biological concepts with which they are more familiar. Associated with the development of these educational materials are summer research experiences for undergraduates and teachers.

submitted post-press

Functional Annotation of *Fibrobacter succinogenes* Carbohydrate Active Enzymes

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

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

Project Goals: Eliminate bottlenecks in plant cell wall deconstruction.

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

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

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

Joint BioEnergy Institute (JBEI)

26

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

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

Joint BioEnergy Institute, Emeryville, Calif.

Project Goals (Abstracts 26-54): In the San Francisco Bay Area, three national laboratories, major public and private universities, industry, and federal agencies have joined together to create the Joint BioEnergy Institute (JBEI). This institute will develop the basic science and technology to create an array of environmentally friendly biofuels using plant biomass and microbes. JBEI will focus its scientific effort in three key areas: feedstock production, deconstruction, and fuels synthesis. JBEI will employ an opportunistic "start-up company" approach, partnering with industry, to develop new science and technologies that address the most challenging steps in industrial bioenergy processing. Crosscutting technologies in computational tools, systems and synthetic biology tools, and advanced imaging will be applied in a multi-pronged approach for biomass-to-biofuel solutions in addition to discovery-driven benefits for biohydrogen research, solar-to-fuel initiatives, and broader DOE programs.

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

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

In the Fuels Synthesis, Division, a challenge is to produce advanced biofuels needed for diesel and jet engines. Many fuel-producing organisms incompletely metabolize C_5 (xylose, arabinose) sugars, while inhibitors (e.g., acetate) released from biomass pretreatment can limit fuels production. Introducing new metabolic routes to advanced biofuels relies on Synthetic Biology approaches. Few tools are available for biomass conversion research. The Technologies Division is developing high-throughput 'omics and microfluidic approaches for many aspects of bioenergy research, as well as high-throughput biochemical, spectroscopic, and imaging methods for the rapid characterization of biomass.

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

27

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

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

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

Project Goals: See goals for Abstract 26.

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

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

28

Towards the Plant Golgi Proteome

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

[1]Joint BioEnergy Institute, Emeryville, Calif.

Project Goals: See goals for Abstract 26.

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

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

29

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

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

Feedstocks Division, Joint BioEnergy Institute, Emeryville, Calif.

Project Goals: See goals for Abstract 26.

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

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

30

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

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

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

Project Goals: See goals for Abstract 26.

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

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

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

31

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

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

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

Project Goals: See goals for Abstract 26.

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

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

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

32

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

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

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

Project Goals: See goals for Abstract 26.

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

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

33

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

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

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

Project Goals: See goals for Abstract 26.

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

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

34

Screening Glycosyltransferases for Enzymatic Activity

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

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

Project Goals: See goals for Abstract 26.

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

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

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

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

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

35

Microfluidic Technology for Biofuels Applications

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

Joint BioEnergy Institute, Emeryville, Calif.

Project Goals: See goals for Abstract 26.

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

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

36

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

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

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

Project Goals: See goals for Abstract 26.

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

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

37

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

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

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

Project Goals: See goals for Abstract 26.

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

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

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

38

Understanding Ionic Liquid Pretreatment of Lignocellulosic Biomass by Hyperspectral Raman Imaging

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

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

Project Goals: See goals for Abstract 26.

Pretreatment of ligonocellulosic biomass is essential for breaking down the highly interwoven matrix of cellulose, lignin and hemicellulose, which facilitates enzyme accessibility and adsorption to cellulose for efficient saccharification. Ionic liquids are a new class of non-volatile solvents exhibiting excellent solvating properties and have shown great promise for lignocellulosic biomass pretreatment with easy recovery of cellulose by rapid precipitation with anti-solvents. Ionic liquids have been demonstrated to be very effective in cellulose solubilization in bulk, and have shown to swell cell walls perhaps by breaking inter and intra chain hydrogen bonding. However, to date, molecular level understanding of ionic liquid pretreatment on lignin and its impact on different tissue and cell types of biomasses is lacking. The aim of this research is to develop a fundamental understanding of ionic liquid pretreatment by monitoring the compositional changes during the pretreatment process. Raman microscopy based on molecular vibrational spectroscopy is a label-free imaging technique capable of real-time and noninvasive examination of plant cell walls with chemical selectivity. In this research, we employed hyperspectral Raman imaging to study tissue and cell type specific distribution of cell wall components and the impact of ionic liquid pretreatment on various cell types of corn stover to identify signatures for predicting deconstructionability and understand pretreatment dynamics. The Raman mapping results have shown that the distribution of lignin and cellulose varies significantly across different tissue and cell types in the following order: sclerenchyma cells and tracheids > epidermal cells > bundle sheath cells > parenchyma cells. Lignin content decreases rapidly in tracheids and sclerenchyma cells and slowly in parenchyma cells during ionic liquid pre treatment. Significant cell wall swelling of various cell types during ionic liquid pretreatment was revealed by confocal fluorescence microscopy.

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

39

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

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

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

Project Goals: See goals for Abstract 26.

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

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

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

40

Targeted Enzyme Discovery in Feedstock-Adapted Microbial Communities

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

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

Project Goals: See goals for Abstract 26.

Efficient saccharification of pre-treated feedstocks is essential to produce inexpensive biofuels derived from biomass. The enzyme cocktails used for this step need to be robust and able to withstand inhibitors produced during pretreatment. To identify enzymes suitable for saccharification on an industrial scale, we have focused on adapting microbial communities to specific feedstocks and pretreatment conditions (i.e. high temperatures, ionic liquid contamination, etc.), followed by functional characterization of secreted glycosyl hydrolases. To select for thermophilic enzymes, microbial communities derived from compost were adapted to biomass feedstocks at 60°C. Phylogenetic profiling of these communities show that each consists of a simple mixed consortia with just two or three species in high abundance. The secretomes obtained from these consortia have biomass-deconstructing enzymatic activity that is both thermostable and active in high concentrations of ionic liquid, two highly desirable characteristics for industrial enzymes. Zymography was used to investigate the complement of glycosyl hydrolase enzymes expressed by the consortia, detecting at least a dozen active enzymes for multiple polysaccharide substrates. Currently, proteins within individual zymogram bands are being studied by MS-based proteomics. Candidate glycosyl hydrolases will be identified by comparing measured peptides masses to predicted protein sequences from Carbohydrate Active enZYme (CAZy) database and genome sequences of reference organisms related to consortial members. Future work will utilize metagenomic and single-cell genomic sequencing to document comprehensively the glycosyl hydrolases secreted by these highly active microbial communities. Cataloguing the glycosyl hydrolases in these secretomes will enable us to design thermophilic enzyme cocktails for biomass deconstruction that function under the conditions required for industrial conversion of biomass to biofuels.

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

41

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

Anthe George[1]* (AGeorge@lbl.gov), Peter Benke,[2] Kim Tran,[1] **Blake Simmons**,[1] and Bradley Holmes[1]

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

Project Goals: See goals for Abstract 26.

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

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

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

42

Trapping Lignin Degrading Microbes in Tropical Forest Soil

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

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

Project Goals: See goals for Abstract 26.

Lignin in feedstock plant material represents a barrier to more efficient plant-to-biomass conversion and can also hinder enzymatic access to cellulose. For this reason, it is critical to develop a way to degrade recovered lignin for next generation feedstock-derived biofuels. While the best-known ligninases are fungal, bacteria are more amenable to emerging cellulosic biofuels technologies. Tropical rain forest soils in Puerto Rico are likely dominated by bacterial decomposers because of the frequent anoxic conditions and fluctuating redox characteristic of these soils, so we focused here to search for novel bacterial lignase producers. To do this, we buried bug traps containing lignin-amended and unamended biosep beads in the soil and incubated them for 1, 4, 13 and 30 weeks. At each time point, phenol oxidase and peroxidase enzyme activity was found to be elevated in the lignin-amended versus the unamended beads, while cellulolytic enzyme activities were significantly depressed in lignin-amended beads. Quantitative PCR of bacterial communities showed more colonization in the lignin-amended compared to the unamended beads after one and four weeks, which attenuated over the course of the incubation. The microbial community was analyzed by microarray (PhyloChip) and by pyrotag sequencing of the community16S ribosomal RNA genes. Community trends were strongly driven by time but also lignin-amendment to the beads. These techniques also allow us to identify which taxa were increased in lignin-amended compared to unamended beads, which included representatives from the phyla Actinobacteria, Firmicutes, Acidobacterial and Proteobacteria.

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

43

High-Resolution Electron Microscopy Imaging of Plants and Pretreated Biomass

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

Joint BioEnergy Institute, Emeryville, Calif.

Project Goals: See goals for Abstract 26.

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

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

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

44

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

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

Joint BioEnergy Institute, Emeryville, Calif.

Project Goals: See goals for Abstract 26.

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

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

45

Flux Analysis of Biodiesel-Producing *E. coli*

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

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

Project Goals: See goals for Abstract 26.

Biofuel yields can be limited below their theoretical maximum because of pathway characteristics (e.g. lack of enzyme activity or unbalanced gene expression) or a lack of flux directed towards the synthesizing pathway. The goal of the Host Engineering directorate is to prevent the latter. A way to achieve this is to knock genes out in such a way that carbon flux gets channeled towards the desired pathway. In order to do that in a rational manner it is desirable to know the internal metabolic fluxes and have a way to predict the outcome of knockout experiments. Here we present a internal metabolic flux profiles for a biodiesel-producing *E. coli* measured through ^{13}C metabolic flux analysis (^{13}C MFA) and knockout suggestions obtained through Flux Balance Analysis (FBA) constrained by ^{13}C MFA flux measurements.

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

46

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

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

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

Project Goals: See goals for Abstract 26.

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

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

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

47

Towards Automated Assembly of Biological Parts

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

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

Project Goals: See goals for Abstract 26.

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

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

48

Improving Biofuel Production by Using Efflux Pumps to Limit Solvent Toxicity

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

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

Project Goals: See goals for Abstract 26.

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

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

49

Quantitative Proteomics for Metabolically Engineered Biofuel Pathway Optimization

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

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

Project Goals: See goals for Abstract 26.

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

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

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

50

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

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

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

Project Goals: See goals for Abstract 26.

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

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

51

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

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

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

Project Goals: See goals for Abstract 26.

Increasing energy costs and environmental concerns have emphasized the need to sustainably produce renewable fuels and chemicals. Major efforts to this end are focused on the microbial production of high-energy fuels through cost-effective "consolidated bioprocesses". Fatty acids are composed of long alkyl chains and represent nature's "petroleum," being a primary metabolite class used by cells for both chemical and energy storage functions. These energy rich molecules are today isolated from plant and animal oils for a diverse set of products ranging from fuels to oleochemicals. A more scalable, controllable, and economic route to this important class of chemicals would be through the microbial conversion of renewable feedstocks, such as biomass-derived carbohydrates. Here we demonstrate the engineering of *E. coli* to produce structurally tailored fatty acid ethyl esters (biodiesel), fatty alcohols, and waxes directly from simple sugars and the further engineering of the biodiesel-producing cells to secrete hemicellulases, a step toward producing these compounds directly from hemicellulose, a major component of plant-derived biomass[1]. Although this complete production scheme has been demonstrated, increases in titer, productivity, and yield are necessary for industrial transition. Strategies employed for increasing yields of biodiesel include balancing the enzymes in the pathway, condensing the pathway onto a triple-operon, single-plasmid system, and subsequent chromosomal integration. These efforts guided towards understanding fermentation scalability, pathway stability, and balancing pathway enzymes for biodiesel production have resulted in higher yields.

Reference

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

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

52
JBEI Electronic Laboratory Notebook System

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

Joint BioEnergy Institute, Emeryville, Calif.

Project Goals: See goals for Abstract 26.

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

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

53
JBEI Computational Biology Core

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

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

Project Goals: See goals for Abstract 26.

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

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

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

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

54

Techno-Economic Modeling of Cellulosic Biorefineries

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

Deconstruction Division, Joint BioEnergy Institute, Emeryville, Calif.

Project Goals: See goals for Abstract 26.

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

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

BioEnergy Science Center (BESC)

55

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

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

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

http://bioenergycenter.org

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

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

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

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

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

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

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

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

56

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

Mustafa Syed[1,2]* (syedmh@ornl.gov), Byung-Hoon Park,[1,2] Tatiana Karpinets,[1,2] Michael Leuze,[1,2] **Ed Uberbacher**,[1,2] and Martin Keller[1,2]

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

http://bioenergycenter.org

Project Goals: See goals for abstract 55.

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

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

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

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

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

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

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

57

Technology and Transgenics for Unparalleled Improvements in Switchgrass Biomass Quality

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

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

http://bioenergycenter.org

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

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

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

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

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

58

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

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

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

http://bioenergycenter.org

Project Goals: See goals for abstract 57.

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

* Presenting author

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

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

59

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

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

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

http://bioenergycenter.org

Project Goals: See goals for abstract 57.

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

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

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

60

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

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

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

http://bioenergycenter.org

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

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

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

A high throughput pretreatment and enzymatic hydrolysis (HTPPH) method has been developed as part of the BioEnergy Science Center (BESC) to screen natural and genetically modified biomass types for those with lower recalcitrance to sugar release, define pretreatment conditions, and screen enzyme formulations. The high throughput system has been shown to mimic conventional pretreatment and enzymatic hydrolysis laboratory methods but has the advantage of being able to quickly screen hundreds of samples. Thus far, the HTPPH system has been successfully applied to a set of 47 natural *Populus trichocarpa* samples in BESC's Poplar Association Study to define trends in sugar release behavior and identify outliers that warrant further study. In addition, the HTPPH system has enabled screening of individual annual rings from a cross section of *Populus tremuloides* to investigate radial variation of sugar release and the importance of sampling technique. Results will be shown from these studies to demonstrate the power of the HTPPH system for screening biomass samples for sugar release and identifying those with reduced recalcitrance. In support of the system, a scaled-down method was also developed to determine biomass composition that speeds compositional analysis, increases accuracy, and reduces labor demands.

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

61

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

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

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

http://bioenergycenter.org

Project Goals: See goals for abstract 60.

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

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

62

Structural Characterization of the Xylan Oligosaccharides by Mass Spectrometry

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

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

http://bioenergycenter.org

Project Goals: See goals for abstract 60.

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

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

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

63

Understanding Cellulase Activity Using Single Molecule Spectroscopy

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

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

http://bioenergycenter.org

Project Goals: See goals for abstract 60.

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

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

64

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

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

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

http://bioenergycenter.org

Project Goals: See goals for abstract 60.

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

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

65

Elucidating the Mechanism of Xylan Biosynthesis: A Biochemical Approach

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

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

http://bioenergycenter.org

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

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

Biosynthesis research provides detailed studies in these and other plant models on key biosynthetic pathways for lignin, pectin, xylan and cellulose. The coordinated transformation of target genes for validation is in progress.

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

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

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

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

* Presenting author

66

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

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

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

http://bioenergycenter.org

Project Goals: See goals for abstract 65.

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

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

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

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

67

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

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

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

http://bioenergycenter.org

Project Goals: See goals for abstract 65.

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

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

68

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

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

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

http://bioenergycenter.org

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

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

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

Quantitative trait loci (QTL) studies are an integral part of plant research and are used to characterize the genetic basis of phenotypic variation observed in structured populations and inform marker-assisted breeding efforts. These QTL intervals can span large physical regions on a chromosome comprising hundreds of genes, thereby hampering candidate gene identification. Genome history, evolution, and expression evidence can be used to narrow the genes in the interval to a smaller list that is manageable for detailed downstream functional genomics characterization. As a first step in this study, we analyzed the lignin content of 29 cell wall constituents both in stem and root in an inter-specific three generation hybrid poplar pedigree. By establishing a high density genetic map for this pedigree, QTL analyses were conducted to explore the underlying genetic loci. The major findings are as follows; a) Lignin content in stem is significantly higher than that in root , 2) Cell wall constituents can be classified into four groups, with strong correlation only observed within organs (stem or root), 3) Pleiotropic QTLs are common and 4) QTLs are differentially expressed in stem and root. Our second motivation for the present study was to address the need for a research methodology that identifies candidate genes within a broad QTL interval. Towards this end, a bioinformatics-based approach for subdividing candidate genes within QTL intervals into alternate candidate groups of decreasing probability. Application of this approach in the context of studying cell wall traits, specifically, lignin content and S/G ratios of stem and root in *Populus* plants, resulted in the identification of manageable sets of genes of both known and putative cell wall biosynthetic function. These results provide a roadmap for future experimental work leading to identification of new genes controlling cell wall recalcitrance, and ultimately, in the utility of plant biomass as an energy feedstock.

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

69

Transcriptome and Metabolome Profiling of *Populus* Tension Stress Response

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

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

http://bioenergycenter.org

Project Goals: See goals for abstract 68.

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

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

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

70

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

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

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

http://bioenergycenter.org

Project Goals: See goals for abstract 68.

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

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

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

71

Visualizing Supramolecular Cell Wall Degrading Enzyme Complexes and Aggregates

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

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

http://bioenergycenter.org

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

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

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

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

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

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

72

High Throughput Pretreatment and Enzyme Hydrolysis: A Massively Parallel Approach

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

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

http://bioenergycenter.org

Project Goals:See goals for abstract 71.

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

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

73

Critical Enzymes for Lignin Degradation

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

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

http://bioenergycenter.org

Project Goals: See goals for abstract 71.

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

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

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

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

74

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

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

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

http://bioenergycenter.org

Project Goals: See goals for abstract 71.

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

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

75

Application of Phylogenomic Techniques in Studying Glycosyltransferase and Glycoside Hydrolase Families

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

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

http://bioenergycenter.org

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

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

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

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

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

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

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

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

76

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

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

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

Project Goals: See goals for abstract 75.

Mass-spectrometry (MS) based proteomics provides the most direct evidence of gene product presence in specific samples, but MS data interpretation requires sophisticated computational algorithms to uncover the wealth of the encoded information. Our research aims to develop more robust and reliable identification algorithms for shotgun tandem MS (MS/MS) experiments. More comprehensive proteome characterizations will lead to more complete assessments of functional potential for collected plant samples, as well as, to more confident reconstruction of protein co-expression subnetworks, identification of regulatory signals, etc.

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

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

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

77

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

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

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

http://bioenergycenter.org

Project Goals: See goals for abstract 75.

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

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

78

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

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

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

http://bioenergycenter.org

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

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

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

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

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

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

79

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

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

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

http://bioenergycenter.org

Project Goals: See goals for abstract 78.

Biofuel production from renewable lignocellulosic biomass is dependent on the efficient enzymatic hydrolysis of cellulose in order to release fermentable sugars. Cellulolytic microorganisms inhabiting geothermally heated terrestrial hot springs are known to possess multidomain/multifunctional cellulases and hemicellulases that display increased heat-stability. *C. obsidiansis* is a newly characterized extremely thermophilic anaerobe capable of hydrolyzing cellulose, xylan, and pretreated lignocellulosic biomass (switchgrass and *Populus*) while fermenting the resulting sugars into acetate, lactate, CO_2, H_2, and ethanol. In this study, we used confocal laser scanning microscopy and 3-D image reconstruction to monitor the spatial and temporal dynamics of colonization and degradation of cellulose by *C. obsidiansis*. These data indicate that *C. obsidiansis* forms colonies that spread horizontally and vertically after attachment on the cellulose membrane, forming depressions that punctured the substrate within 72 hours. The distance between the cells and the substrate surface played a crucial role in the conversion rate of the cellulose, and thus overall fermentation efficiency. In addition to these modeling efforts, we show that end-product profiles including ethanol can be influenced by applying different growth conditions. By understanding the mechanism of microbial colonization of recalcitrant polymerized sugars, it may be possible to manipulate the fermentation conditions and target genetic modifications to improve the utilization efficiency of substrate carbon to produce desired end-products such as ethanol.

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

80

Advances in Organism Development for Consolidated Bioprocessing

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

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

http://bioenergycenter.org

Project Goals: See goals for abstract 78.

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

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

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

81

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

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

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

http://bioenergycenter.org

Project Goals: See goals for abstract 78.

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

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

82

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

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

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

http://bioenergycenter.org

Project Goals: See goals for abstract 78.

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

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

83

BioEnergy Science Center Education and Outreach

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

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

http://bioenergycenter.org

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

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

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

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

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

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

Biofuels: Analytical and Imaging Technologies for Studying Lignocellulosic Material Degradation

84

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

Li-Qiang Chu[1]* (lchu1@nd.edu), Zhen Li,[2] Rachel Masyuko,[1] **Jonathan V. Sweedler**,[2] and **Paul W. Bohn**[1]

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

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

The ability to efficiently use lignocellulosic materials (LCMs) to feed the biorefinery of the future depends on high-efficiency pre-enzymatic processing to render lignin separable from cellulose/hemicelluloses. The complex 3-D network structure and chemical characteristics of LCMs pose daunting challenges for imaging and molecular characterization: (1) they are opaque and highly scattering; (2) their chemical composition is a spatially variegated mixture of heteropolymers; (3) the nature of the matrix evolves in time during processing. Presently, there are few *in situ* characterization tools that can be applied to materials with these characteristics, especially during processing—yet acquiring this information is of paramount importance. Here we present a combination of Raman microscopy and mass spectrometric imaging (secondary ion mass spectrometry, SIMS, and laser desorption ionization mass spectrometry, LDI MS) to visualize the structural and chemical changes of LCMs upon various treatments, such as H_2SO_4, $NaClO_2$, NaOH, etc. *Miscanthus x giganteus* as a model LCM was sectioned into 50 μm thick, and was investigated in this work. Raman and SIMS imaging results indicated that lignin and cellulose are collocated in the cell wall of raw *miscanthus*. A globular structure, composed predominantly of hemicellulose and lignin, is associated with the interior cell wall. Pretreatment of *Miscanthus* using NaOH or $NaClO_2$ solutions results in the removal of lignin at long processing time. Interestingly, Raman experiments reveal that the H_2SO_4 treated *Miscanthus* exhibits a higher autofluorescence, which might be due to the formation of highly conjugated hydrocarbon species during H_2SO_4 treatment. We are currently correlating these Raman results with mass spectrometric imaging to identify the unknown

* Presenting author

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

85

Dynamic Visualization of Lignocellulose: A Biofuels Scientific Focus Area

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

[1]Oak Ridge National Laboratory, Oak Ridge, Tenn.; [2]Institute of Paper Science and Technology, Georgia Institute of Technology, Atlanta; and [3]University of Tennessee, Knoxville

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

The Oak Ridge National Laboratory (ORNL) Scientific Focus Area (SFA) Biofuels program will provide fundamental information about the structure and deconstruction of plant cell walls that is needed to drive improvements in the conversion of renewable lignocellulosic biomass to biofuels. This program integrates neutron scattering with computational simulation and molecular dynamics to understand the physicochemical processes taking place across multiple length scales during deconstruction of lignocellulosic biomass. These molecular-level methods are assisted and complemented by technical expertise in lignocellulose characterization at the Institute of Paper Science and Technology (Georgia Tech), and chemical force microscopy methods for surface characterization. A multipurpose neutron imaging chamber will be designed and used for in situ, dynamic observation of biomass processing. Deuteration of the biomass crop switch grass and other cellulose sources is being carried out to enable higher contrast between the components of lignocellulosic biomass for neutron scattering that will enable the examination of surface accessibility. These novel technological capabilities are being applied to specific problems in the pretreatment and enzymatic hydrolysis of biomass to produce the fundamental understanding of plant cell architecture that is needed to develop the next generation of cost-effective cellulosic ethanol production. The interrelated research is organized as three principal tasks: (1) sample preparation and characterization; (2) neutron scattering and diffraction; and (3) computer simulation and modeling.

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

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

86

SANS Study of Dilute Acid Pretreatment of Switchgrass

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

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

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

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

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

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

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

87

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

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

[1]Oak Ridge National Laboratory, Oak Ridge, Tenn.; [2]Institute of Paper Science and Technology, Georgia Institute of Technology, Atlanta; and [3]University of Tennessee, Knoxville

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

Computer simulation can provide the integration of structural information obtained from multiple imaging and analysis methods that is needed to visualize the molecular structure of lignocellulose and its changes during pretreatment and hydrolysis. The objective is the development and demonstration of computer simulation technology that will provide multi-length scale, real-time imaging of biomass during pretreatment and enzymatic hydrolysis based on experimental data obtained through neutron science, surface force microscopy, and advanced NMR methods. The integration and application of these combined capabilities will provide new information on lignocellulosic degradation at an unprecedented level of detail.

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

References

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

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

Student Presentation

88

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

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

Cornell University, Ithaca, New York

Project Goals: Extracellular secretion of cellulases by engineered bacteria

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

89

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

Nicolai S. Panikov[1]* (n.panikov@neu.edu), Maria V. Sizova,[1] **Slava S. Epstein**,[1] Anthony V. Palumbo,[2] and Martin Keller[2]

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

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

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

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

Results: Stable cellulose-degrading consortia were obtained in most tested soils and biocomposts. The typical fermentation products in enrichments and stabilized consortia were CO_2, H_2, CH_4, acetate, ethanol and lactate. In some cases, acetate was replaced with formate. At high sulfate concentration (some soils and sediments), cellulose degradation was suppressed, probably because of toxic by-products formed by sulfate-reducing bacteria. The methanogenic communities grew slow (specific growth rate 0.05-0.1 h^{-1} at 55°C) with sustained oscillations of fermentation rate (the effect of reversible product inhibition).

More than 15 pure cultures of novel organisms related to *C. clariflavum, C. straminisolvens* and *C. thermocellum* have been isolated. *C. clariflavum* was able to ferment cellulose, xylan and their mixture as well as pretreated wood into ethanol, formate, CO_2 and H_2. Other isolates degraded only cellulose with acetate as end product. New strains varied in respect to growth and maintenance rates, enzymes localization (free and cell-bound), sensitivity to product inhibition as well as to starvation and O_2-stress (see illustration below).

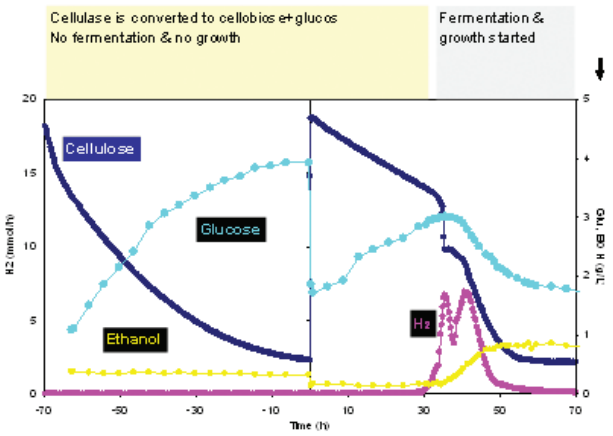


Fig 1. Example of starvation stress. Culture of cellulolytic bacteria *C. thermocellum* enters non-growing state after short-term starvation. First addition of cellulose-mineral medium (time -70 h) does not result in a normal cell growth: extracellular cellulases remaining from the past growing period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

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

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

References

1. Kaeberlein, T., K. Lewis and S.S. Epstein (2002). Isolating "uncultivable" microorganisms in pure culture in a simulated natural environment. *Science* 296 (5570): 1127-1129.
2. Lynd, L.R., P.J. Weimer, W.H. van Zyl and I.S. Pretorius (2002). Microbial cellulose utilization: fundamentals and biotechnology. Microbiol Mol Biol Rev 66 (3): 506-577, table of contents.
3. Nichols, D., K. Lewis, J. Orjala, S. Mo, R. Ortenberg, P. O'Connor, C. Zhao, P. Vouros, T. Kaeberlein and S.S. Epstein (2008). Short Peptide Induces an "Uncultivable" Microorganism To Grow In Vitro. Appl. Environ. Microbiol. 74 (15): 4889-4897.
4. Panikov, N.S. (1995). *Microbial Growth Kinetics.* Chapman and Hall. 378 p.
5. Panikov, N.S. (2008). Kinetics, Microbial Growth. In: Michael C. Flickinger and Stephen W. Drew (Eds.) *Encyclopedia of Bioprocess Technology: Fermentation, Biocatalysts and Bioseparation*. New York: John Wiley & Sons, Inc.: 1513-1543.
6. Panikov, N.S. and L.R. Lynd (2010). Physiological and Methodological Aspects of Celluloytic Microbial Cultures. In: *Manual of Industrial Microbiology and Biotechnology*, 3rd Ed. (Baltz, Davies, and Demain, eds.)

90

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

Kenneth E. Hammel[1]* (kehammel@wisc.edu), Daniel J. Yelle,[1] John Ralph,[2] Dongsheng Wei,[1] Alexander N. Kapich,[1] and Daniel Cullen[1]

[1]USDA Forest Products Laboratory, Madison, Wis. and [2]Dept. of Biochemistry, University of Wisconsin, Madison

http://www.fpl.fs.fed.us/

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

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

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

91

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

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

Lawrence Berkeley National Laboratory, Berkeley, Calif.

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

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

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

92

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

Jamie Cate[1,2]*, **Hoi-Ying N. Holman**[1] (hyholman@lbl.gov) and Haw Yang[3]

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

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

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

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

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

93

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

Christopher G. Hunt[1]* (cghunt@fs.fed.us), **Kenneth E. Hammel**[1,2] (kehammel@wisc.edu), and Don Carl Jones[1]

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

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

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

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

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

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

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

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

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

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

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

* Presenting author

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

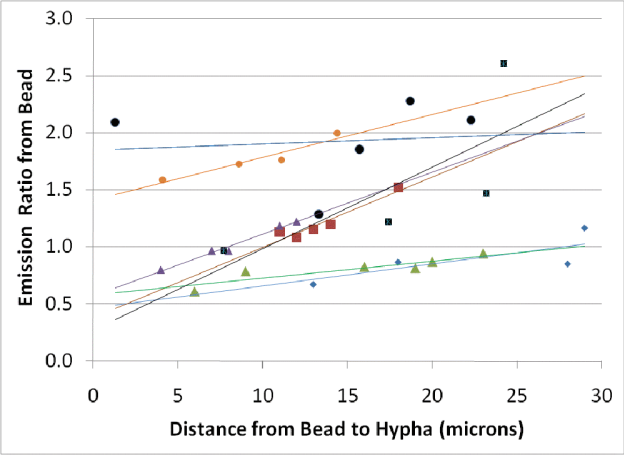


Figure 1: Fluorescent emission from individual beads as a function of distance from hypha. Lower Y axis values indicate more oxidation.

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

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

94

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

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

University of Florida, Gainesville, Florida

Project Goals: Our *long-term goal* is to develop a quantitative structural model for changes that occur in the organization and chemical composition of plant biomass during pretreatment, enzymatic degradation and bioconversion to ethanol or other products. The *objectives* of this proposed work are to 1) use advanced high resolution magnetic resonance microscopy (MRM), micro and nano x-ray computed tomography (x-ray CT), electron microscopy and imaging mass spectrometry (IMS) to quantify changes in the architecture, porosity, permeability, surface area, pore size, interconnectivity, chemical organization and composition of bagasse and particularly *Populus* and pine wood chips during pretreatment and enzymatic degradation, and 2) integrate the results from these different quantitative imaging methods into a model for disassembly of the plant cell wall during pretreatment and bioconversion. We are developing methods for imaging biomass with MRM, x-ray CT and IMS.

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

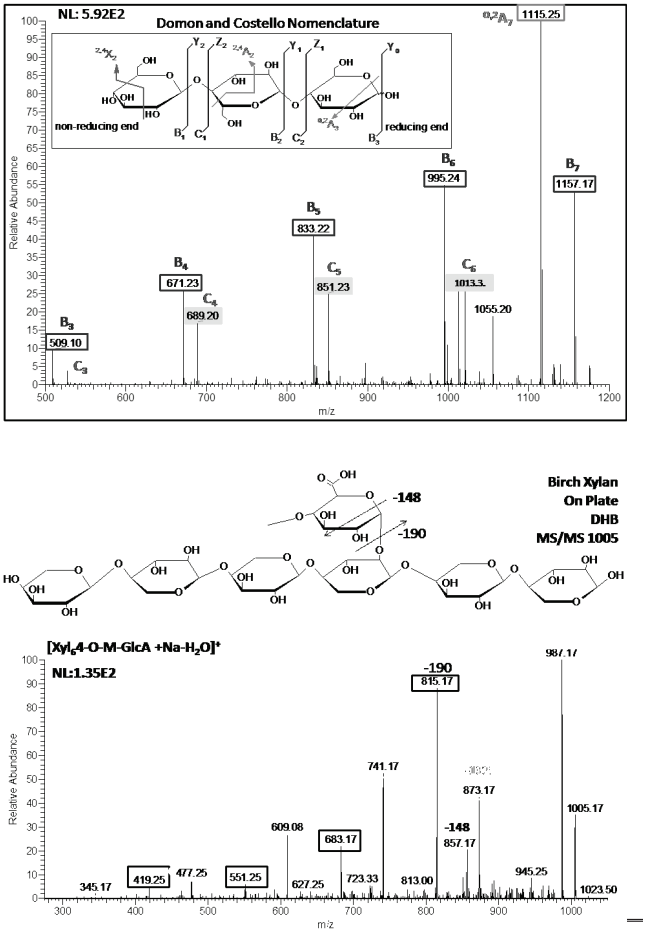


Figure 1

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

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

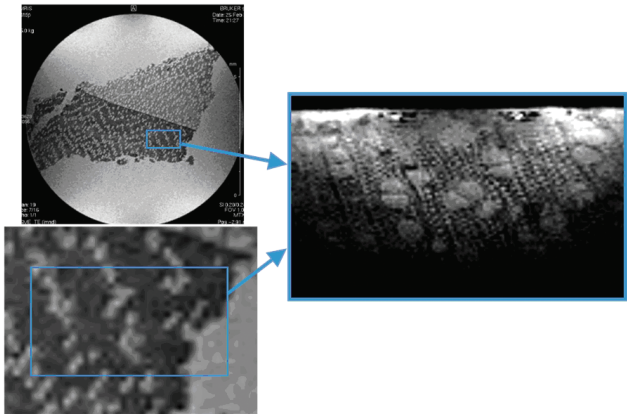
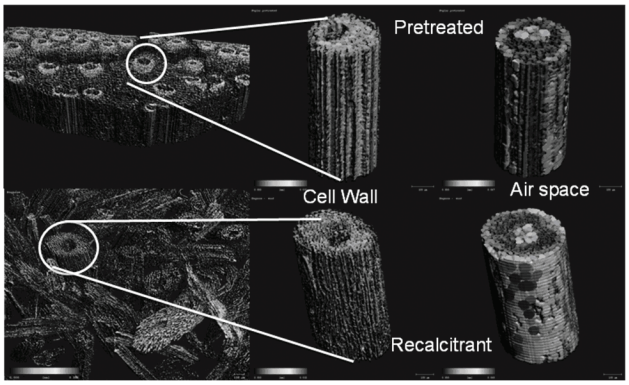


Figure 2. Illustration of the microcoil SNR improvement. On the left is a conventional 39 micron resolution image (top) with an expanded view (bottom). Using a microcoil an 8 micron image (right) with an equivalent field of view (indicated by the boxes) is show – the clear improvement in the ability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

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



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

95

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

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

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

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

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

* Presenting author

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

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

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

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

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

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

96

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

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

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

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

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

submitted post-press

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

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

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

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

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

In this poster we will present the recent results from our work developing instruments and methods for carrying out correlated soft x-ray tomography and high numerical aperture immersion light microscopy on cryogenic specimens[1,2]. These new imaging modalities have enormous potential as precision structural phenotyping tools for bioenergy research. The novel use of a cryogenic immersion fluid in the cryolight microscope minimizes the refractive index mismatch between the specimen and lens, leading to a more efficient coupling of the light from the sample to the image forming system[3]. The instrument can be used for correlating detailed spectral imaging with a high fidelity x-ray tomographic map of any microorganism. We will show results of correlated imaging on yeast, and also show results of using soft x-ray tomography to phenotype algae for biofuel production.

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

References:

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

Systems Biology and Metabolic Engineering Approaches for Biological Hydrogen Production

97

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

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

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

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

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

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

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

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

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

98

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

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

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

Project Goals: See goals for abstract 97.

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

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

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

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

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

Student Presentation

99

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

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

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

Project Goals: See goals for abstract 97.

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

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

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

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

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

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

100

Photobiological H_2 Production in *Cyanothece* ATCC 51142

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

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

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

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

Biological H_2 production by bacteria and microalgae has been known for more than a century, and research directed at practical application of such microbial processes has been carried out for more than three decades. Although many biohydrogen production concepts have been described, fundamental technological challenges remain in making any such process a practical reality. Advances in microbial genome sequencing and functional genomics are greatly improving the ability to conduct system-level studies of microbial metabolism and to use the obtained knowledge to identify fundamental questions that must be resolved to advance biofuel applications. Genomics and metabolic engineering hold great promise for the rational design and manipulation of biological systems to make such systems efficient and economically attractive.

The research conducted as part of the PNNL Biofuels Scientific Focus Area (BSFA) focuses on elucidating the mechanisms of light-driven metabolism in a unicellular diazotrophic cyanobacterium *Cyanothece* sp. strain ATCC51142. Conditions promoting H_2 production by *Cyanothece* 51142 are being studied in order to develop a strategy for maximizing the output of H_2 using metabolic modeling approach. Initially, two-phase experiments have been employed to promote photosynthetically driven accumulation of glycogen that is subsequently converted to H_2. Specifically, during the first phase, strain 51142 was grown in continuous cultures under N-limitation in a photobioreactor sparged with CO_2-enriched Ar (0.3% v/v) and continuously illuminated at 150 μmol/m^2•s. Upon reaching steady-state (biomass concentration 80 mg/l of ash-free dry weight), the cultures were incubated in N-free medium in the absence of CO_2 using 100% Ar as sparging gas while measuring the off-gas composition by in-line mass-spectrometry. Phase two was initiated by placing cultures either under dark or the light (200 μmol/ m^2•s) conditions. Appropriate controls consisting of light and dark cultures amended with ammonium or 3- (3,4-dichlorophenyl)-1,1-dimethylurea (DCMU), an inhibitor of photosystem II, were also included in the experiment. The data obtained revealed that, unlike non-nitrogen fixing cyanobacteria such as *Synechocystis* and *Synechococcus* spp., *Cyanothece* 51142, is not capable of generating significant amounts of H_2 from stored glycogen under dark fermentative conditions, nor does it produce H_2 under light conditions in the presence of ammonia or N_2. However, illuminated cultures exposed to an Ar atmosphere and deprived of N_2 and CO_2 produced significant amounts of H_2. O_2 was also produced along with H_2 at 1:2 ratio, whereas DCMU significantly (4.5-fold) decreased H_2 generation. It should be noted that analysis of cell-free culture supernatants did not revealed any accumulation of organic acids. These results suggest that PSII and therefore water photolysis played a significant role in H_2 evolution by strain 51142. The inhibition of H_2 production by ammonia or N_2 strongly *suggest that nitrogenase was the* enzyme primarily responsible for light-driven H_2 production, and whole cell assays revealed high nitrogenase activity in H_2-producing cells. Preliminary analysis of the material balance suggests the nitrogenase activity was supported by light-driven electron transfer. Within the scope of the proposed BSFA research, we will further elucidate and validate the pathways of light-driven two-step H_2 production by cyanobacteria and incorporate the experimental data into the metabolic model of *Cyanothece* 51142 to identify the means for maximization of H_2 production by this organism.

101

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

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

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

Project Goals: See goals for abstract 100.

Photosynthetic microorganisms possess the unique ability to convert sunlight into chemical energy using water as the electron donor. Despite the wealth of information on the mechanistic aspects of bacterial photosynthesis and supramolecular complexes catalyzing the process of light conversion and CO_2 fixation, a system-level understanding of photosynthetic metabolism is yet to be achieved. Different phototrophic microorganisms display varying levels of light conversion efficiencies, which ultimately translate in different rates of electron transfer, ATP/NAD (P)H production, and growth. Understanding the origin of these properties will provide fundamental new insights that could be widely applied to the development of photosynthetic systems for biofuels development. Integral to that is the question, is the process of reductant partitioning in photoautotrophs which links energy-generating reactions with biosynthesis of biomass precursors and storage compounds. Within the scope of the PNNL Biofuels Scientific Focus Area (BSFA), we are exploring the mechanisms of energy conservation and carbon partitioning in cyanobacteria. One important outcome of the project will be development of a predictive tool, *i.e.* a genome-scale model, which provides a platform for integrating all knowledge and experimental data generated within the project. It will also have the ability to serve as an *in silico* tool for manipulating photosynthetic microorganisms to act as catalysts for solar energy conversion and will potentially allow development of a highly efficient biofuel production process.

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

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

102

Genome-Enabled Studies of Photosynthetic Microorganisms for Bioenergy Applications

Alex S. Beliaev[1]* (alex.beliaev@pnl.gov), **Allan E. Konopka**,[1] Harvey Bolton Jr.,[1] Mary S. Lipton,[1] Thomas O. Metz,[1] Grigoriy E. Pinchuk,[1] Aaron T. Wright,[1] Thomas C. Squier,[1] Sergey M. Stolyar,[1] Donald A. Bryant,[2] Matthew C. Posewitz,[3] and Jennifer L. Reed[4]

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

Project Goals: See goals for abstract 100.

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

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

* Presenting author

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

103

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

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

[1]Pacific Northwest National Laboratory, Richland, Wash.; [2]University of Wisconsin, Madison; and [3]Pennsylvania State University, University Park

Project Goals: See goals for abstract 100.

Much of terrestrial and aquatic photosynthesis occurs in complex microbial consortia, but little is presently known about interactions among microorganisms that contribute to their efficient solar energy capture and conversion. Under the PNNL Foundational Scientific Focus Area (FSFA), we are utilizing a combination of complementary laboratory-based research and exploration of naturally-occurring associations with combinations of defined photoautotroph-heterotroph cultures to understand the collective energy, carbon, and nutrient processing in microbial systems. The complex natural systems we have selected for study including the phototrophic mats of Yellowstone National Park (YNP) and surface waters of central Washington as of yet have not yielded cultivated members that can be used as model systems for detailed laboratory-based research. Due to the paucity of information on the heterotrophic population (s) in these natural systems, initial research and development of methodology for studying microorganism interactions is utilizing co-cultures of representative cyanobacteria and well-studied *Shewanella* species. While recognizing the opportunistic nature of this system, there is ample evidence that certain species of *Shewanella* live in association with autotrophic prokaryotes and examples of these associations are well documented. More importantly, we believe that a *Synechococcus-Shewanella* co-culture can be instrumental in gaining basic understanding of opportunistic interactions between photoautotrophic and heterotrophic bacteria. Our preliminary results using cyanobacteria-*Shewanella* co-cultures demonstrated that metabolic coupling and interactions between photoautotrophic and heterotrophic microorganisms may serve as a mechanism for controlling dissolved O_2 concentration, increasing Fe and Mn availability, and recycling nutrients in natural communities.

Complementary to the FSFA work, we are also exploring the potential of photoautotroph-heterotroph associations for bioenergy applications as part of the PNNL Biofuels Scientific Focus Area. Although phototroph-heterotroph associations are abundant in nature, the co-culture approach has been seriously under-appreciated. To date, engineering of microbes for biofuel production is being carried out using single strains by enhancing or deleting specific steps of a pathway or modulating activities of specific enzymes. However, synthesis of biofuel precursor molecules requires precise coordination and interactions of many proteins within various pathways, where any adjustments or increases in expression and/or activity levels can lead to substantial metabolic burden and suboptimal yields. In that regard, engineering of photosynthetic organisms, which carry out simultaneous light- and dark-phase reactions, is inherently challenging. Photosynthetic production of biofuels often requires optimization of two or more metabolic functions which can be mutually exclusive in a single microbial cell and therefore require either spatial and/or temporal separation (e.g. O_2 evolution and H_2 production; O_2 evolution and N_2 fixation, sensitivity of RuBisCo to O_2). To that end, co-culturing of photosynthetic and heterotrophic microorganisms offers efficient ways to optimally engineer the photosynthetic production of biofuels. By engineering photosynthetic strains which excrete organic carbon compounds (organic acids or sugars) and co-culturing them with a heterotrophic organism capable of utilizing the excreted compounds, one can physically separate the processes of photosynthesis and photosynthate conversion while allowing for net CO_2 consumption. The co-cultivation of phototrophs and aerobic heterotrophs also eliminates technical problems associated with oxygen-sensitivity and substrate delivery by creating favorable microaerobic CO_2-enriched environments for the photorophic microorganisms. Moreover, utilization of exogenously-added organic carbon by the heterotroph, decreases dissolved O_2 concentrations and induces the expression of O_2-sensitive enzymes (*e.g.* hydrogenase and nitrogenase) in the phototroph which in turn will generate reducing equivalents by light-driven water photolysis. Overall, we believe that the implementation of the co-culture approach will open new perspectives for designing efficient and cost-effective processes and will provide a novel platform for the development of consolidated bioprocessing methods leading to production of carbon-neutral energy at reduced economic and energetic costs.

104

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

Stephen P. Cramer[1,2*] (spjcramer@ucdavis.edu)

[1]Dept. of Applied Science, University of California, Davis and [2]Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, Calif.

Project Goals: We are trying to develop a resource for biological and environmental x-ray spectroscopy at ALS beamline 4.0.2.

Nitrogenase is the enzyme responsible for the 'fixation' of nearly inert atmospheric dinitrogen to ammonia. It is ultimately responsible for half of the world's protein, while the other half depends on industrial fertilizer produced with hydrogen derived from fossil fuels. Nitrogenase uses a complex $MoFe_7S_9X$-homocitrate 'FeMo-cofactor' to convert N_2 to NH_3, but the detailed mechanism remains poorly understood.[1-3] Another type of enzyme, hydrogenase, catalyzes the interconversion of dihydrogen with protons and electrons. These enzymes use unusual forms of Fe-S clusters or Fe carbonyls, and their catalytic mechanisms are not understood.

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

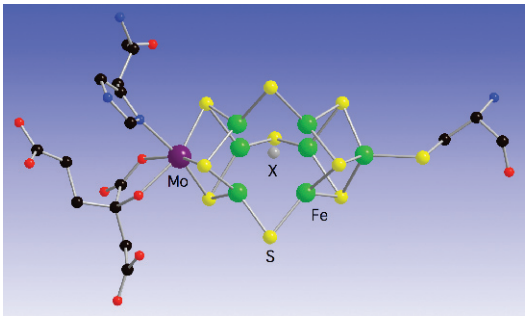


Figure. Structure of the nitrogenase FeMo-cofactor.

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

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

References

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

Student Presentation

105

Stopped-Flow IR Spectroscopy of Hydrogenases

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

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

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

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

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

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

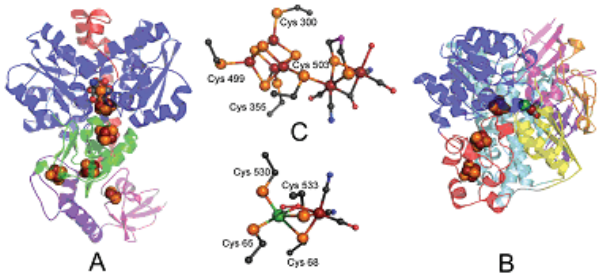


Figure. Structures of (A) CpI H_2ase, (B) [NiFe] H_2ase, (C) H-cluster, and (D) NiFe active site.

106

Novel Hydrogen Production Systems Operative at Thermodynamic Extremes

Michael J. McInerney,[1] Ralph S. Tanner,[1] Gregory Hurst,[2] Lars Rohlin,[3] and **Robert P. Gunsalus**[3]* (robg@microbio.ucla.edu)

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

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

Project Goals: The goals of this collaborative project are to develop new research strategies to address the Genomics:GTL program needs in the area of bio-hydrogen production. This includes the delineation of the molecular machinery involved in hydrogen production from thermodynamically difficult substrates, as well as the characterization of new microbial model systems that generate high H_2 concentrations approaching 17% of the gas phase. We are performing systems-based studies of bio-hydrogen production in model anaerobic consortia as well as with pure culture model strains to identify key regulated steps. The results of these studies will greatly expand our ability to predict and model systems for H_2 production in novel anaerobes that are currently very poorly understood.

The goals of this collaborative project are to develop new research strategies to address the Genomics:GTL program needs in the area of bio-hydrogen production. This includes the delineation of the molecular machinery involved in hydrogen production from thermodynamically difficult substrates, as well as the characterization of new microbial model systems that generate high H_2 concentrations approaching 17% of the gas phase.

To identify the strategies used for hydrogen production in one model butyrate-degrading organism, we are performing genomic, proteomic, and transcript analysis on *Syntrophomonas wolfei*. This microbe is representative of an important but poorly understood class of hydrogen-producing organisms that are capable of syntrophic fatty and aromatic acid metabolism when co-cultured with suitable microbial partner(s). Their ability to produce H_2 requires reverse electron transport with energy input. Analysis of the *S. wolfei* genome reveals many genes with potential to accomplish this task. It possesses genes for three cytoplasmic and two externally located formate dehydrogenases plus two cytoplasmic and one externally located hydrogenase. By implication, either hydrogen or formate could be produced by *S. wolfei* during syntrophic interactions. Interestingly, the three cytoplasmic formate dehydrogenases plus one of the soluble-type hydrogenases appear to be NADH-linked since the respective gene clusters contain genes for NADH:quinone oxidoreductases chains E and F. This suggests that *S. wolfei*, like several other sequenced syntrophic metabolizers and anaerobes known to produce high molar ratios of hydrogen from glucose, may produce H_2 and/or formate from NADH by an electron bifurcation mechanism. To determine which of the above enzymes are utilized in pure culture, proteomic studies were performed initially using crotonate-grown cells. This was accomplished by analyzing whole cell-derived peptide mixtures with two-dimensional liquid chromatography/tandem mass spectrometry (2D LC-MS-MS) via the MudPIT approach. Two highly expressed hydrogenase enzymes were detected plus a novel, electron transfer flavoprotein-linked FeS-type reductase complex that is probably used to process electrons generated by the oxidation acyl-CoA intermediates. *S. wolfei* metabolizes fatty acids by the β-oxidation pathway and surprisingly, the genome reveals multiple homologues for many of the enzymatic steps even though it has an extremely restricted range of fatty acid substrates. Provisional protein assignments were also made by LC-MS-MS for each of the eight reactions leading to acetate formation. Energy is harvested by substrate level phosphorylation via acetate kinase to yield one ATP. Additional proteomic and transcript studies are in progress to further characterize the expression of genes/proteins involved in the reversed electron transfer process for hydrogen production. These studies will yield improved understanding of how *S. wolfei* and other syntrophic metabolizers thrive at low thermodynamic driving forces not possible for many other anaerobes.

In a companion project we are characterizing the genetic, biochemical, and physiological properties of a newly isolated anaerobic bacterium called *Anaerobaculum hydrogenoformans* strain OS1 that can generate H_2 at concentrations up to 17%. In one approach, genomic sequencing is being performed on this representative member of the Synergistetes group. The genome of approximately 2.4 MB in size has a GC content of 46.5%. The current assembly consists of 403 contigs with about 1.9 MB contained in the top six contigs. Machine annotation and manual curation is currently in progress to support a metabolic reconstruction of the cellular metabolism leading to hydrogen formation in strain OS1 when grown on hexoses and pentoses. Since strain OS1can also grow syntrophically in the presence of a H_2-consuming methanogen, it suggests an ability for a more complex alternative lifestyle.

107

Hydrogen Production Comes at the Expense of Calvin Cycle CO_2 Fixation During Photoheterotrophic Growth by *Rhodopseudomonas palustris*

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

Dept. of Microbiology, University of Washington, Seattle

Project Goals: The goals of this project are (i) to use ^{13}C-metabolic flux analysis and other approaches to understand the central metabolic changes involved in H_2 production by *R. palustris* and (ii) to use the resulting information to guide the metabolic engineering for improving H_2 production characteristics.

There is currently a pressing need for renewable fuels to negate the adverse social, economic, and environmental impacts of burning fossil fuels. H_2 is a promising biofuel, having about three-times the energy content of gasoline. Although most manufactured H_2 comes from fossil fuels H_2 can also be produced biologically. *Rhodopseudomonas palustris* uses energy from sunlight and electrons from organic waste to produce H_2 via nitrogenase. In order to understand and improve this process we used ^{13}C-acetate to track and compare central metabolic fluxes in non-H_2 producing wild-type *R. palustris* and an H_2-producing mutant. Wild-type cells metabolized 22% of the acetate to CO_2 and then fixed 68% of this CO_2 into cell material using the Calvin cycle. This Calvin cycle flux enabled *R. palustris* to re-oxidize nearly half of the reduced cofactors generated during acetate oxidation. The H_2-producing mutant produced a similar amount of CO_2 but the Calvin cycle flux was much lower, re-assimilating only 12% of the CO_2. In this mutant, H_2 production assumed much of the redox balance burden as about 90% of the electrons for H_2 production were diverted away from the Calvin cycle. Microarray and Q-PCR analyses showed that the shift of electrons from the Calvin cycle to H_2 production involved transcriptional control of Calvin cycle operons. However, this transcriptional control did not require the redox-sensing two-component regulatory system RegSR. When Calvin cycle flux was disrupted completely by deleting the genes encoding ribulose 1,5-bisphosphate carboxylase, *R. palustris* was forced to use H_2 production alone to maintain redox balance. This mutant exhibited a 1.5-fold increase in H_2 yield but at a cost to the growth rate. These results demonstrate how systems level approaches, such as ^{13}C-metabolic flux analysis, can lead to effective strategies to improve H_2 yield. Furthermore, our results underscore that the Calvin cycle and nitrogenase have important electron-accepting roles separate from their better known roles in ammonia production and biomass generation.

108

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

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

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

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

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

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

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

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

Reference

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

109

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

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

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

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

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

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

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

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

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

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

Reference

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

110

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

D. White,[1]* R. Singh,[1] K. Noll,[2] R. Kelly,[3] and **P. Blum**[1]

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

Project Goals: The anaerobic hyperthermophile, *Thermotoga maritima* (Tma), ferments carbohydrates to from molecular hydrogen (H_2) as one of its by-products. A high overall H_2 yield makes *T. maritima* a preferable model to evaluate hydrogen production. Our focus is to develop and use genetic methods to manipulate the *Tma* genome to investigate the metabolic flux of carbon and hydrogen synthesis.

The anaerobic hyperthermophile, *Thermotoga maritima (Tma),* ferments carbohydrates to from molecular hydrogen (H_2) as one of its by-products. A high overall H_2 yield makes *T. maritima* a preferable model to evaluate hydrogen production. Our focus is to develop and use genetic methods to manipulate the *Tma* genome to investigate the metabolic flux of carbon and hydrogen synthesis. A new *Tma* genetic marker has been developed by screening for uracil auxotrophs among spontaneous mutants resistant to the pyrmidine analog, 5-fluroroorotic acid (5-FOA). The *pyrE-64* mutant (strain PBL3001) arose by a two nt deletion (-TG) at chromosomal positions 351,539 (-T) and 351,538 (-G), 155 nt from the end of *pyrE*. This mutation results in a premature stop codon (TGA) 64 nt before the natural stop and therefore reduces protein length by 21 AA (from an original 187 AA). Auxotrophy was confirmed by demonstrating growth in a defined medium was dependent upon uracil supplementation. Stability of *pyrE-64* was evaluated by enrichment and characterization of gain-of-function prototrophic suppressors. The *pyrE-100* mutant (strain PBL3021) restored the *pyrE* reading frame by deletion of an additional one nt flanking the primary lesion. Additional genetic markers may also arise from studies on spontaneous novobiocin resistant isolates that target gyrase. Current efforts are underway to repair *pyrE-64* by directed recombination using a suicide vector and, by complementation using a *groESp*::*pyr*E promoter fusion fragment carried on a replicating shuttle vector based on a synthetic copy of pRQ7 fused to pUC19. Recombination strategies will target carbon catabolic pathways and components of the multiple hydrogenases while complementation strategies will be used to import new traits into *Tma*. Both genetic approaches will support collaborative efforts on transcriptomics, studies of the *Tma* toga and metabolic modeling.

* Presenting author

Student Presentation

111

Carbohydrate Fermentation to Hydrogen by Hyperthermophilic *Thermotoga* Communities

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

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

Project Goals (Abstracts 111-112): Objective 1—Examine the regulation of substrate catabolic proteins and pathways as this relates to carbon partitioning, disposition of reducing power, and H_2 generation in *Thermotoga maritima*. Objective 2—Dissect catabolic and regulatory pathways using genetic approaches based on past success with other hyperthermophiles. Objective 3—Thermotogales biodiversity arises from adaptive specialization that expands on a conserved minimal genome; physiological characterization of selected novel traits will be done to expand understanding of biohydrogenesis.

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

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

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

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

112

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

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

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

Project Goals: See goals for abstract 111.

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

References

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

113

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

DaeHee Lee[1]* (dlee@bioeng.ucsd.edu), Dmitry Rodionov,[2] Vasiliy Portnoy,[1] **Andrei Osterman**,[2] **Bernhard O. Palsson**,[1] and **Karsten Zengler**[1]

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

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

This project is focused on a systems-level understanding of biological hydrogen production using *Thermotoga maritima* as a model organism. We therefore have integrated a metabolic reconstruction of *T. maritima* that contains 479 metabolic genes, 565 metabolites (non-unique) and 646 internal and external metabolic reactions with 478 protein structures to generate the first three-dimensional reconstruction of the central metabolic network of a bacterium. To understand how the flow of information from the genome to the different states of the metabolic network is archived we studied the transcriptome architecture of *T. maritima.* A protocol for chromatin immunoprecipitation (ChIP) has been adapted to work for *T. maritima*. Furthermore, we developed a method that allows for the genome-wide determination of transcription start sites (TSSs) with a single base-pair resolution. Genome-wide transcription profiles using high-density tilled arrays were integrated with binding regions of RNA polymerase and TSS information to generate an experimentally verified map of the transcriptional landscape, laying the foundation of the experimentally elucidation of the operon structure in *T. maritima.* In addition we used an integrated approach to systematic annotation and reconstruction of transcriptional regulons in the available genomes of the Thermotogales. Two major components of this analysis are (i) annotation and propagation of previously known regulons from model organisms to others (e.g., arabinose regulon AraR, arginine regulon ArgR), and (ii) *ab initio* prediction of novel regulons (e.g., inositol regulon InoR, mannose regulon ManQ). In addition to playing a key role in regulon reconstruction, regulons provide an additional layer of genome context, helping to significantly improve the accuracy of functional annotations and metabolic reconstruction. Comparative analysis of the InoR regulon led to a discovery of a new pathway of inositol catabolism, which is currently under experimental investigation by in vitro enzymatic assays.

114

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

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

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

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

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

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

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

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

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

References

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

115

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

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

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

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

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

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

116

Using Hydrogen Isotopes to Assess Proton Flux during Biological Hydrogen Production 1. Determining Fractionation Factors and the Proton Transfer Pathway in Hydrogenases

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

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

Project Goals: To improve our understanding of biological H_2 production using H/D isotope ratios.

Biological H_2 production by hydrogenase enzymes has enormous potential as an environmentally sustainable source of energy. Hydrogenases, found throughout nature in many diverse organisms, are among the most efficient H_2-producing catalysts known. Although considerable progress has been made in elucidating the metabolic pathways involved in H_2 metabolism, many uncertainties remain. One major impediment to improving our understanding of H_2 metabolism is our inability to adequately define the regulation of and the flux through key pathways involved in H_2 production. Thus far, few attempts have been made to utilize hydrogen isotopes to improve our knowledge of the H_2 metabolic pathways, perhaps because the source of protons for hydrogenase enzymes is intracellular water. Until recently, intracellular water was generally assumed to be isotopically equivalent to extracellular water, and therefore it was perhaps thought that hydrogen isotopes would not be informative. We are exploiting our recent discoveries that intracellular water can be isotopically distinct from extracellular water, and that the contribution of protons from metabolic substrates to intracellular water can be quantified, to develop the use of hydrogen isotopes for studying intracellular proton trafficking.

We predicted that the isotope ratio of H_2 produced by various hydrogenases would differ because of slight differences in the active sites and proton transfer pathways. We further predicted that we can measure this difference via isotope-ratio mass spectrometry, and that the H/D isotope ratios would allow us to address fundamental questions concerning biological H_2 production including the source of the H_2. To test this predictions, we purified five different hydrogenases (three [FeFe]-H_2ases and two [NiFe]-H_2ases) and established conditions that allowed us to quantify the specific activity of the purified H_2ases and the amount of H_2. In addition, we built a custom chromatographic system for the analysis of H_2 and interfaced this system with an isotope ratio mass spectrometer (IRMS). We obtained a reproducibility of better than 3‰ for δD, and this precision is maintained down to a lower sensitivity limit of 0.2 μmol H_2 in 1 mL of headspace volume. In addition, because H_2 diffuses so readily through most materials (resulting in isotopic fractionation), it was necessary to develop and validate a robust protocol for capturing biologically-produced H_2. Subsequently, reaction conditions were developed that allow for the reproducible formation and capture of the optimal concentration of H_2.

Using the enzymes and optimized protocols established above, we determined the isotope ratio of the H_2 produced by three different [FeFe]-H_2ases (*Clostridium pasteurianum, Shewanella oneidensis,* and *Chlamydomonas reinhardtii*) and two [NiFe]-H_2ases (*Shewanella oneidensis* and *Desulfovibrio fructosovorans*). Significantly, the data indicate that all 5 hydrogenases produce H_2 with a unique isotopic signature. This proves our initial hypothesis, that different H_2-producing enzymes have different fractionation factors, and that these differences are reflected in the isotope ratio of the H_2.

Building on these results, we are using this data to help elucidate the proton transfer pathway in [FeFe]-H_2ases. To accomplish this task, we mutated a number of residues proposed to be critical for proton transport in the *C. pasteurianum* enzyme. Mutations that affect the proton pathway will change the fractionation factor by changing the ener-

* Presenting author

getics of proton vs. deuteron migration, while mutations that affect H_2ase activity by other mechanisms will not change the fractionation factor. We have generated nine variants and all have been shown to alter H_2ase activity. Future studies will ascertain if these mutations also alter the H/D isotope ratio of the H_2.

In the second phase of the project, we are also utilizing the experimentally-determined H_2ase fractionation factors for in vivo studies to identify the major H_2-forming pathway in *S. oneidensis* under a variety of growth conditions (see accompanying poster).

117

Using Hydrogen Isotopes to Assess Proton Flux during Biological Hydrogen Production 2. In Vivo Studies with *Shewanella oneidensis*

Eric A. Hill,[1] James J. Moran[1]* (James.Moran@pnl.gov), **Eric L. Hegg**,[2] and **Helen W. Kreuzer**[1]

[1]Pacific Northwest National Laboratory, Richland, Wash. and [2]Dept. of Biochemistry, Michigan State University, East Lansing

Project Goals: To improve our understanding of biological H_2 production using H/D isotope ratios.

Biological hydrogen production by hydrogenase enzymes represents a potentially sustainable, non-polluting source of energy. In order to fully exploit the capabilities of biological systems for hydrogen production, a deeper understanding of regulation of and fluxes through key pathways involved in hydrogen production is needed. We have developed methods using stable isotope measurements that allow us to trace protons from extracellular water and metabolic substrates into intracellular water and cellular metabolites, including hydrogen gas. In this part of our project, we are applying those methods to elucidate hydrogen production pathways in vivo, using *S. oneidensis* as a model organism.

Shewanella oneidensis MR-1 is a facultative anaerobe capable of transferring electrons to a variety of terminal acceptors including iron, manganese, and other metals. *S. oneidensis* encodes two hydrogenase enzymes, the [FeFe]-hydrogenase HydA and the [NiFe]-hydrogenase HyaB. Hydrogenases catalyze the reversible reaction of protons plus electrons to form hydrogen gas. If electron acceptors in its growth environment are limited, *S. oneidensis* reduces protons, producing hydrogen gas.

The objectives of this portion of our project are to determine:

- **What is the contribution of each hydrogenase enzyme to hydrogen production?** When the organism is perturbed, does the manipulation differentially affect the flux through each hydrogenase?
- **Is there channeling of protons between organic substrates and hydrogenases?** Although it is well established that the addition of organic substrates increases H_2 production under certain conditions (e.g. acetate to the growth media of green algae or glucose to the growth media of cyanobacteria), the precise mechanism by which this occurs is not entirely clear. Some have suggested that protons are directly channeled from specific substrates into hydrogen production. We will test this hypothesis.

We have characterized hydrogen production by wild-type and electron transfer-deficient *S. oneidensis* strains. When the wild-type organism is cultured in sealed headspace vials with limited electron acceptors, the headspace hydrogen concentration initially increases, then decreases, then steadily increases. In strains deficient in metal reduction, we do not observe the decrease in headspace H_2 concentration, demonstrating that the electron transport deficiency impacts hydrogenase activity.

We hypothesize that we will be able to use stable isotopes to dissect proton fluxes through the two hydrogenase enzymes under these and other culture conditions. We have determined that HydA and HyaB evolve isotopically distinct hydrogen from the same substrate water (presented in a companion poster). In this, the second phase of our project, we are measuring the stable isotope content of the intracellular water that presumably is the substrate for hydrogenase activity, as well as that of the hydrogen gas produced in vivo, in the wild-type and mutant strains. Similar measurements will enable us to dissect proton trafficking in the presence of different organic substrates.

118

Systems Biology of Hydrogen Regulation in *Methanococcus maripaludis*

Qiangwei Xia,[1] Erik L. Hendrickson,[1] Sung Ho Yoon,[2] Thomas J. Lie,[3] **William B. Whitman**,[4] **Murray Hackett**,[1] **Nitin Baliga**,[2] and **John A. Leigh**[3]* (leighj@u.washington.edu)

[1]Dept. of Chemical Engineering, University of Washington, Seattle; [2]Institute for Systems Biology, Seattle, Wash.; [3]Dept. of Microbiology, University of Washington, Seattle; and [4]Dept. of Microbiology, University of Georgia, Athens

http://maggie.systemsbiology.net
http://faculty.washington.edu/leighj/

Project Goals: 1. Use transcriptomics, proteomics, and metabolomics to study the systems biology of H_2 metabolism, formate metabolism, nitrogen fixation, and carbon assimilation in *Methanococcus maripaludis*. 2. Determine the mechanism of H_2 sensing and transcriptional regulation by H_2.

Background
We are engaged in a long-term effort to understand regulatory networks in hydrogenotrophic methanogens, members

of the Archaea whose energy metabolism specializes in the use of H_2 to reduce CO_2 to methane. (Many hydrogenotrophic methanogens can use formate as an alternative to H_2 and CO_2). Our studies focus on *Methanococcus maripaludis*, a model species with good laboratory growth characteristics, facile genetic tools, and a tractable genome of 1722 annotated ORFs. Much of our work to date has focused on the response that occurs when supplies of essential nutrients are decreased to growth-limiting levels. Thus, we have studied the responses to H_2 limitation, nitrogen limitation, phosphate limitation, and leucine limitation (using a leucine auxotroph) (1-3). A key aspect of our approach is the use of continuous culture for maintaining defined nutrient conditions (4).

Proteomics
We used high-coverage quantitative proteomics to determine the response of *M. maripaludis* to growth-limiting levels of H_2, nitrogen, and phosphate (1). Six to ten percent of the proteome changed significantly with each nutrient limitation. H_2 limitation increased the abundance of a wide variety of proteins involved in methanogenesis. However, one protein involved in methanogenesis decreased: a low-affinity [Fe] hydrogenase, which may dominate over a higher-affinity mechanism when H_2 is abundant. Nitrogen limitation increased known nitrogen assimilation proteins. In addition, the increased abundance of molybdate transport proteins suggested they function for nitrogen fixation. An apparent regulon governed by the euryarchaeal nitrogen regulator NrpR was identified. Phosphate limitation increased the abundance of three different sets of proteins, suggesting that all three function in phosphate transport. The global proteomic response of *M. maripaludis* to each nutrient limitation suggests a wider response than previously appreciated. The results give new insight into the function of several proteins, as well as providing information that should contribute to the formulation of a regulatory network model.

Five different approaches were compared for measuring protein abundance ratios (5). The results suggest that at the limit of deep sampling, frequency based measurements are competitive with metabolic stable isotope labeling in terms of power to detect abundance change. In addition, false discovery rates and local false discovery rates were compared as complementary approaches to multiple hypothesis testing for quantitative significance. These findings will be discussed in detail in a poster by M. Hackett.

Transcriptomics and metabolomics
A tiling array was designed and used to measure gene expression changes along a growth curve. An initial transcriptome map has been constructed and is currently being hand-annotated. The results will be discussed in detail in a poster by S.H. Yoon and N. Baliga. In addition, methods are being worked out for the measurement of key metabolites.

Hydrogen regulation and metabolism
mRNA levels for key enzymes of methanogenesis are regulated by H_2 availability (3). Results to be presented suggest that this regulation relies on sensing of some intracellular redox indicator, rather than on the external H_2 concentration.

We are also investigating whether H_2 is a necessary intermediate during growth on formate (6). This is of interest in the context of the potential for H_2 production by nitrogenase during growth on formate. If H_2 is not a necessary intermediate, then we should be able to eliminate enzymes that would deplete H_2 that is produced by nitrogenase. Experiments are underway to test the essentiality of genes whose products could produce H_2 during growth on formate.

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

References

1. Xia Q, *et al.* (2009) Quantitative proteomics of nutrient limitation in the hydrogenotrophic methanogen *Methanococcus maripaludis*. *BMC Microbiol* 9:149.
2. Hendrickson EL, *et al.* (2008) Global responses of *Methanococcus maripaludis* to specific nutrient limitations and growth rate. *J Bacteriol* 190:2198-2205.
3. Hendrickson EL, Haydock AK, Moore BC, Whitman WB, and Leigh JA (2007) Functionally distinct genes regulated by hydrogen limitation and growth rate in methanogenic Archaea. *Proc Natl Acad Sci U S A* 104:8930-8934.
4. Haydock AK, Porat I, Whitman WB, and Leigh JA (2004) Continuous culture of *Methanococcus maripaludis* under defined nutrient conditions. *FEMS Microbiol Lett* 238:85-91.
5. Xia Q, Wang T, Beck DAC, Taub F, Leigh JA, and Hackett M (2010) Quantitative local false discovery rates, deep sampling and detecting protein abundance change for the model organism *Methanococcus maripaludis*. *Proteomics* (in review).
6. Lupa B, Hendrickson EL, Leigh JA, and Whitman WB (2008) Formate-dependent H_2 production by the mesophilic methanogen *Methanococcus maripaludis*. *Appl Environ Microbiol* 74:6584-6590.

119

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

Qiangwei Xia,[1,2,4] Tiansong Wang,[2] David A.C. Beck,[3] Fred Taub,[2] **John A. Leigh**,[2] and **Murray Hackett**[1]* (mhackett@u.washington.edu)

[1]Dept. of Chemical Engineering, [2]Dept. of Microbiology, and [3]eScience Institute, University of Washington, Seattle; and [4]Dept. of Chemistry, University of Wisconsin, Madison

Project Goals: 1. To find the best balance between statistical power to detect protein abundance change and the need to decrease the time required for a complete proteome analysis for Methanococcus maripaludis. 2. To establish just how much sampling is required for spectral counting to become as efficient or perhaps even more efficient for generating protein abundance ratios relative to traditional metabolic stable isotope labeling. The answer to this question is heavily dependent on the specific mass

spectrometry instrumentation and methods used in the investigation, both of which have experienced several upgrades and improvements since we last addressed this question in the peer reviewed literature in 2006. 3. To establish the most efficient data reduction and transformation procedures for the use of local false discovery rates (lfdr) with spectral counting data. Lfdr is one approach among several to multiple hypothesis testing, based on the more general concept of false discovery rate. Although proposed initially in the context of microarray-based transcriptome data, lfdr is equally applicable to large-scale quantitative proteomic experiments and RNA-Seq.

Protein abundance ratios were measured using five different approaches for the Archaeon *Methanococcus maripaludis*, a model organism of interest for analytical studies because of the tractable nature of its proteome in terms of size, proteome extraction efficiency and other positive features. Multidimensional capillary HPLC coupled with tandem mass spectrometry was used for analysis of heavy (^{15}N) and natural abundance (^{14}N) tryptic digests of *M. maripaludis* grown in chemostats. Here we report our comparison of abundance ratios based on heavy and light proteomes mixed prior to mass spectrometry; spectral counting of heavy and light proteomes mixed; spectral counting of heavy and light proteomes analyzed separately; summed signal intensities for mixed heavy and light proteomes; and summed signal intensities for heavy and light proteomes analyzed separately. Protein identifications were saturated and proteome penetration maximized at ~ 91% of the predicted protein-encoding open reading frames. False discovery rates and local false discovery rates were compared as complementary approaches to multiple hypothesis testing for quantitative significance. Power calculations, dynamic range and other observations reported suggest that at the limit of deep sampling frequency based measurements are competitive with metabolic stable isotope labeling in terms of power to detect abundance change.

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

120

Development of Metabolic Network Models of *Rhodobacter sphaeroides* for the Prediction of Quantitative Contributors to H_2 Production

L. Safak Yilmaz[1]* (yilmaz@cae.wisc.edu), Timothy J. Donohue[2] (tdonohue@bact.wisc.edu), and **Daniel R. Noguera**[1] (noguera@engr.wisc.edu)

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

Project Goals: 1) Develop a metabolic network model of *Rhodobacter sphaeroides*. 2) Use this model to (a) understand the bioenergetics of growth conditions that produce hydrogen from different carbon sources, so that optimal conditions of hydrogen production can be predicted and (b) determine reactions and pathways that detract electrons from hydrogen production, so that hydrogen production can be potentially maximized by genetic manipulations.

Introduction and Objective

The metabolically versatile organism *Rhodobacter sphaeroides* produces H_2 while using light as an energy source and organic substrates as an electron donor. The overall goal of this project is to elucidate gene networks that contribute to and detract from H_2 production and genetically modify *R. sphaeroides* to divert as high a fraction of substrate electrons as possible to H_2 generation. As a tool to achieve this goal, we are developing metabolic network models of *R. sphaeroides* which can integrate genomic knowledge and experimental data, thereby allowing the prediction of optimization strategies.

Methodology

Stoichiometric reaction network models (Palsson, 2006) were reconstructed using the KEGG database (http://www.genome.jp/kegg/pathway.html), the genome of *R. sphaeroides* strain 2.4.1 (http://www.rhodobacter.org/), and other metabolic network models available in the literature (http://gcrg.ucsd.edu/In_Silico_Organisms). Model development is divided into three stages based on the scale of metabolic pathways represented, as explained below.

Stage 1 activities built a core network model that represented central carbon metabolism. Carbon pathways including glycolysis, citric acid cycle, and the pentose-phosphate pathway are connected to an electron transport chain previously proposed for purple non-sulfur bacteria (Klamt et al., 2008). The energetic contribution of light is part of the electron transport chain, H_2 production is added as a nitrogenase reaction in the absence of N_2 based on available data (see poster by Kontur et al.). The synthesis of the known electron sink polyhydroxybutyrate (PHB) is part of the network as a pathway starting from acetyl coenzyme A. To simulate cell synthesis, precursor metabolites are assembled in a biomass reaction taken from *Escherichia coli* core network models (Palsson, 2006).

Ongoing activities in *Stage 2* expands the Stage 1 model to include amino acid and CO_2 fixation pathways. The existing biomass equation (Stage 1) will also be modified based on the utilization of precursor metabolites in the new pathways and on the average protein composition of purple non-sulfur bacteria (Kobayashi and Kobayashi, 1995) to include amino acids as part of biomass.

A planned *Stage 3* model will add other major metabolic pathways (lipid metabolism, nucleic acid metabolism, glycan biosynthesis, and metabolism of cofactors) to the Stage 2 model. Biomass assembly will be based on building blocks using average cell composition of purple non-sulfur bacteria (Kobayashi and Kobayashi, 1995).

Once reaction networks are constructed, the models are established in MATLAB (MathWorks, Natick, MA) and GAMS (GAMS Development Corporation, Washington, DC) to carry out in silico flux balance analysis (FBA) with

linear programming. The models are trained and tested with experimental results from a systematic analysis of electron flow in *R. sphaeroides* growing photosynthetically on a variety of carbon sources (Yilmaz et al., in review). The data include H_2 generation, PHB synthesis, biomass formation, and production of soluble microbial products (SMP) during exponential and stationary phases of batch cultures, and similar data is being generated from chemostats in ongoing work. All cultures are fed with glutamate as the sole nitrogen source to maximize H_2 production, while single organic acids (succinate, lactate, pyruvate, and fumarate) or sugars (mostly glucose) are used as the carbon source.

Results

The Stage 1 model qualitatively captures aerobic growth without H_2 production and anoxygenic photosynthetic growth with H_2 production. The model predicts that H_2 production is a necessary electron accepting pathway in photosynthetic growth, without which excess electrons cannot be recycled. Another surprising prediction is that, SMP production, which was a significant electron sink in experimental cultures (Yilmaz et al., in review), occurs during dimmer light conditions (i.e. with a limited flux of the light reaction), possibly representing shadowing in dense cultures. The FBA results always predict a larger CO_2/H_2 ratio (≥1.0 in partial pressure) than what is found in the headspace of batch cultures (~0.2). This difference is attributed to the lack of CO_2 fixation pathways in our Stage 1 model. Although unconstrained simulations do not quantitatively match experimental data, constrained FBA fits to experimental results for the production of key electron sinks.

The Stage 1 model lacks pathways to simulate glutamate consumption as a carbon, nitrogen and electron source, but the Stage 2 can account for this due to addition of amino acid pathways. This addition and the introduction of CO_2 fixation pathways to the Stage 2 model are expected to provide quantitative and qualitative predictions for the mechanisms of growth and H_2 production as a function of carbon sources. The databases used were checked for the absence of missing reactions in amino acid and carbon fixation pathways before we began reconstruction of the Stage 2 model. A major improvement in the Stage 3 model will be the ability to include a global analysis of metabolism with FBA and existing microarray data. To this end, we developed a prototype visualization tool that overlays flux and gene expression profiles to KEGG pathway maps. The poster will present results from the Stage 1 model, the current Stage 2 model, and illustrations from the visualization tool for integrated microarray and flux data.

References

1. Klamt, S, Grammel, H, Straube, R, Ghosh, R, and Gilles, E D (2008) Modeling the electron transport chain of purple non-sulfur bacteria. Molecular Systems Biology. 4:156.
2. Kobayashi, M, and Kobayashi, M (1995) Waste Remediation and Treatment Using Anoxygenic Phototrophic Bacteria. In: Blankenship, R E, Madigan, M T, and Bauer, C E (ed), Anoxygenic Photosynthetic Bacteria. Kluwer Academic Publishers Dordrecht, Netherlands, pp 1269-1282
3. Palsson, B O. 2006. Systems Biology. Cambridge University Press, New York, NY.
4. Yilmaz, L S, Kontur, W, Sanders, A P, Sohmen, U, and Donohue, T J. Electron Partitioning During Light- and Nutrient-Powered Hydrogen Production by *Rhodobacter sphaeroides*. Bioenergy Research. In review.

121

Cellular Redox Balance and the Integrative Control of Carbon Assimilation and Hydrogen Production in *Rhodobacter sphaeroides* and *Rhodopseudomonas palustris*

Rick A. Laguna* (laguna.2@osu.edu) and **F. Robert Tabita**

Dept. of Microbiology, Ohio State University, Columbus

Project Goals: Elucidate molecular mechanisms by which evolved strains of NSP bacteria derepress the synthesis of nitrogenase, allowing this enzyme system to catalyze the production of high levels of hydrogen gas. Perform computational and molecular modeling studies to identify further control circuits that might be manipulated to improve hydrogen production. Determine how these different regulatory networks are integrated in the cell. Determine how various regulatory protein complexes we have implicated contribute towards regulating hydrogen production. Examine the intracellular organization of these complexes and determine how small effector metabolites influence their function. Integrate these studies to engineer the most efficient strain for maximal hydrogen production.

Nonsulfur purple (NSP) photosynthetic bacteria are characterized by their metabolic versatility. These organisms are capable of growth during photosynthetic and non-photosynthetic conditions, in the absence or presence of oxygen, respectively, and they can synthesize cell mass via the assimilation of either organic or inorganic carbon sources, with needed energy obtained via photochemical or dark chemical processes. During aerobic chemolithoautotrophic and anaerobic photolithoautotrophic growth conditions, CO_2 serves as the sole carbon source and is reduced into cellular carbon by the Calvin-Benson-Bassham (CBB) reductive pentose phosphate pathway. Under photoheterotrophic growth conditions, organic carbon compounds, usually organic acids, are oxidized into cell mass. In NSP bacteria such as *Rhodobacter sphaeroides* and *Rhodopseudomonas palustris*, when malate is used as the carbon source during photoheterotrophic growth, the reduction of metabolically produced CO_2 through the CBB cycle is the preferred means by which these organisms consume excess reductant produced from the assimilation of the organic electron donor (1-3). However, when the CBB cycle is inactivated, these organisms face the problem of either not growing or adapting by using other redox balancing mechanisms. In one interesting adaptive scenario, excess reductant is consumed by the derepression of the nitrogenase complex, resulting in the production of hydrogen gas in both *Rb. sphaeroides*

and *Rps. palustris*, as well as *Rhodospirillum rubrum* and *Rb. capsulatus* (4,5). This is of interest considering that malate is more oxidized than cell mass. Therefore the need to consume excess reductant through the CBB cycle should not be required. However upon examining the assimilation of malate in *Rb. sphaeroides* and *Rps. palustris* (both assimilate malate through the tricarboxylic acid pathway) and by using various CBB mutant strains, it has been shown that excess reductant is produced and must be consumed by the CBB cycle in order to maintain cellular redox balance. The reason for this is that as malate is assimilated into cell mass in both organisms, excess reductant is generated. Acetate can also be used as a carbon source for photoheterotrophic growth in *Rb. sphaeroides* and *Rps. palustris*. Unlike malate, acetate is at the same oxidation state as cell mass. Therefore excess reductant should not be generated by the assimilation of acetate and the CBB cycle should not be required. In *Rb. sphaeroides* this is the case; however in *Rps. palustris* there is a need for the CBB cycle. The reason for this is that the two organisms use different metabolic pathways to assimilate acetate. *Rb. sphaeroides* uses the ethylmalonyl-CoA pathway to assimilate acetate (6), which does not generate excess reductant from this process. Therefore CBB mutant strains of *Rb. sphaeroides* are capable of photoheterotrophic growth using acetate as the carbon source. However *Rps. palustris* uses the glyoxylate pathway to assimilate acetate (7), which does produce excess reductant. Therefore CBB mutant strains of *Rps. palustris* cannot grow photoheterotrophically with acetate as the carbon source. We are currently examining how the assimilation of different carbon sources affects redox balance and hydrogen production in *Rb. sphaeroides* and *Rps. palustris*, as well as *R. rubrum* and *Rb. capsulatus*, during photoheterotrophic growth conditions. Indications are that there are differences that might be eventually exploited to maximize hydrogen production.

References

1. Falcone, D. L. and F. R. Tabita. 1991. Expression of endogenous and foreign ribulose 1,5-bisphosphate carboxylase-oxygenase (RubisCO) genes in a RubisCO deletion mutant of Rhodobacter sphaeroides. J. Bacteriol. 173:2099-2108.
2. Romagnoli, S. and F. R. Tabita. 2006. A novel three-protein two-component system provides a regulatory twist on an established circuit to modulate expression of the cbbI region of Rhodopseudomonas palustris CGA010. J. Bacteriol. 188:2780-2791.
3. Joshi, G. S., Romagnoli, S., VerBerkmoes, N. C., Hettich, R. L., Pelletier, D., and Tabita, F. R. 2009. Differential accumulation of form I RubisCO in Rhodopseudomonas palustris CGA010 under photoheterotrophic growth conditions with reduced carbon sources. J. Bacteriol. 191:4243-4250.
4. Joshi, H. M. and F. R. Tabita. 1996. A global two component signal transduction system that integrates the control of photosynthesis, carbon dioxide assimilation, and nitrogen fixation. Proc. Natl. Acad. Sci. U. S. A. 93:14515-14520.
5. Tichi, M. A., and Tabita, F.R. 2000. Maintenance and control of redox poise in Rhodobacter capsulatus strains deficient in the Calvin-Benson-Bassham pathway. Arch. Microbiol. 174:322-333.
6. Alber, B.E., Spanheimer, R., Ebenau-Jehle, C., and Fuchs, G. 2006. Study of an alternate glyoxylate cycle for acetate assimilation by Rhodobacter sphaeroides. Mol. Microbiol. 61:297-309.
7. Larimer, F.W., Chain, P., Hauser, L., Lamerdin, J., Malfatti, S., Do, L., Land, M.L., Pelletier, D.A., Beatty, J.T., Lang, A.S., Tabita, F.R., Gibson, J.L., Hanson, T.E., Bobst, C., Torres, J.L., Peres, C., Harrison, F.H., Gibson, J., Harwood, C.S. 2004. Complete genome sequence of the metabolically versatile photosynthetic bacterium Rhodopseudomonas palustris. Nature Biotechnol. 22:55-61.

122

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

Anindita Banerjee,[1] Jana Stöckel,[1] Lawrence Page,[1] Xueyang Feng,[2] Bing Wu,[2] Yinjie Tang,[2] Hongtao Min,[3] Sujata Mishra,[3] Xiaohui Zhang,[3] Louis A. Sherman[3]* (lsherman@purdue.edu), Stephen J. Callister,[4] Richard D. Smith,[4] and **Himadri B. Pakrasi**[1,2]

[1]Depts. of Biology and [2]Energy, Environmental and Chemical Engineering, Washington University, St. Louis, Mo.; [3]Dept. of Biological Sciences, Purdue University, West Lafayette, Ind.; and [4]Biological Sciences Division, Pacific Northwest National Laboratory, Richland, Wash.

sysbio.wustl.edu/Pakrasi/projects/hydrogen.php

Project Goals: The objective of this proposal is to develop the cyanobacterium *Cyanothece* as a model organism for photobiological hydrogen production. Members of the genus *Cyanothece* 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 *Cyanothece*: *Cyanothece* are unicellular, nitrogen-fixing cyanobacteria. They possess both the nitrogenase and hydrogenase enzyme systems that can catalyze biological hydrogen production. Hydrogen production by different *Cyanothece* strains was assessed by employing appropriate physiological perturbations, based on knowledge from prior transcriptomic and proteomic studies. *Cyanothece* sp. ATCC 51142 exhibited exceptionally high rates of hydrogen production under aerobic conditions. This is in striking contrast to other known unicellular photosynthetic hydrogen producing strains, which produce hydrogen under largely anaerobic conditions. The rates of hydrogen production were significantly enhanced by growing *Cyanothece* cells in the presence of glycerol or high levels of CO_2. Both of these carbon sources enhanced the level of glycogen, an intracellular energy reserve in the cells. Our studies revealed that the high rate of hydrogen production by this strain is

largely mediated by an efficient nitrogenase enzyme system. Programmable photobioreactors equipped with sensors to monitor several critical parameters are now being used to further characterize these strains to achieve higher rates of hydrogen production.

Comparative genomics: Complete genome sequences of five *Cyanothece* strains (ATCC 51142, PCC 7424, PCC 7425, PCC 8801, PCC 8802) are currently available and two more (PCC 7822, ATCC 51472) are in the process of completion at the DOE Joint Genome Institute. The sequences reveal significant metabolic diversity within this group of cyanobacteria. The genome sequence information is being used to generate a *Cyanothece* pan-genome (in collaboration with JGI), comprising of the "core genome" (containing all of the genes common to each genus member) and the "dispensable genome" (containing unique genes or genes shared between two or more strains). The unique genes are likely to confer strain-specific attributes and will be analyzed for their role in hydrogen production.

Genetics: *Cyanothece* 7822 was successfully transformed by electroporation using a modification of the asymmetric PCR technique originally developed for eukaryotic algae. The first target was *nifK*; the gene encoding one of the two subunits of the nitrogenase MoFe protein and a spectinomycin-resistance cassette was inserted in the middle of the *nifK* gene. This strain has been stable for more than 9 months and is resistant to spectinomycin and is incapable of growing on plates or in liquid medium that is lacking combined nitrogen (*i.e.*, it cannot fix atmospheric N_2). The strain is also completely unable to produce hydrogen under any growth condition, although it can still reduce acetylene at rates that are particularly high when the cultures are incubated under argon prior to measurement. This indicates that the strain can still assemble a MoFe complex, although it cannot produce the main products of N_2 fixation, H_2 and NH_3.

Metabolomic Studies: Over the past year, we have developed new metabolite extraction methods and mass spectrometry techniques. Using such new approaches, we characterized the central metabolic pathways in *Cyanothece* 51142, and discovered a novel isoleucine biosynthesis pathway that involves citramalate synthase. We have studied impacts of carbon and nitrogen sources on the central metabolism and hydrogen production by *Cyanothece* 51142, using both biochemical methods and ^{13}C isotopomer approaches. We quantitatively determined CO_2 fixation and carbon substrate utilization under mixotrophic growth conditions in *Cyanothece* 51142. We are currently developing ^{13}C-assisted dynamic flux models to study autotrophic metabolism in cyanobacteria.

Proteomic Studies: In preparation for conducting relative quantitative proteomics analyses, using the AMT tag proteomics approach, construction of reference peptide databases for 6 of 7 *Cyanothece* strains has been completed. Over 460 LC-MS/MS datasets have been generated and analyzed using the high-throughput proteomics capabilities at PNNL. Complete databases correspond to strains *Cyanothece* sp. PCC 8801, PCC 8802, PCC 7424, PCC 7425, PCC 7822, and ATCC 51142. LC-MS/MS datasets for *Cyanothece* sp. ATCC 51472 have also been generated with identification of peptide sequences waiting upon the draft completion of the genome sequence (at the Joint Genome Institute) of this strain. Percent observed coverage of predicted proteins from unique peptides (10% false discovery rate) ranges from roughly 40% to 70%, which inversely correlates to the size range of genome sequences; *Cyanothece* 7424 has the largest genome sequence (~6.5 Mb) and smallest percent observed coverage among the strains.

Project number - DE-FC02-07ER64694

123

Genome-Wide Network Analysis of Metabolism in *Chlamydomonas reinhardtii*

Roger L. Chang[1]* (rlchang@ucsd.edu), Ani Manichaikul,[2] Balaji Santhanam,[3] Lila Ghamsari,[3] Erik F.Y. Hom,[4] **Kourosh Salehi-Ashtiani**,[3] and **Jason A. Papin**[2]

[1]Dept. of Bioengineering, University of California, San Diego; [2]Dept. of Biomedical Engineering, University of Virginia, Charlottesville; [3]Center for Cancer Systems Biology and Dept. of Cancer Biology, Dana-Farber Cancer Institute, and Dept. of Genetics, Harvard Medical School, Boston, Mass.; and [4]Dept. of Molecular and Cellular Biology, Harvard University, Cambridge, Mass.

http://bme.virginia.edu/csbl
http://ccsb.dfci.harvard.edu

Project Goals: Our goal in this study is to carry out genome-wide network reconstruction of *C. reinhardtii* metabolism integrated with structural and functional annotation, to provide a framework to integrate these metabolic pathways for investigation of global properties of metabolism.

Algae have garnered significant interest in recent years, both for their potential as a source of biofuel, as well as for their use in the production of nutritional supplements. Among eukaryotic microalgae, *Chlamydomonas reinhardtii* has established its position as an ideal model organism, popular for its relatively fast doubling time and its ability to grow under standardized conditions in the laboratory. Characterization of metabolic functions in *C. reinhardtii* provides a framework for developing engineering strategies towards generation of strains with improved production of commercial targets, as well as studying diverse cellular processes such as photosynthesis and cell motility. Extensive literature on *C. reinhardtii* metabolic function and many mutants with metabolic phenotypes provide a solid foundation toward detailed characterization of individual metabolic pathways in this organism.

Our goal in this study is to carry out genome-wide network reconstruction of *C. reinhardtii* metabolism integrated with structural and functional annotation, to provide a framework to integrate these metabolic pathways for investigation of global properties of metabolism. The availability of complete

genome sequence data (*JGI* v4.0), recently released for *C. reinhardtii*, allowed us to perform in-house annotation of the metabolic genes encoded within this genome (*please see our accompanying abstract by Ghamsari et al. for details*). Briefly, functional annotations enable the identification of the presence of enzymes encoded within the genome of the organism to define the scaffolding of the reconstruction. Comprehensive literature searches are used as the primary form of evidence to establish the structure of all metabolic pathways of interest. Finally, this information is supplemented with more general knowledge of metabolic pathways, as provided in classical biochemistry textbooks and also available in online databases.

We report the first genome-scale reconstruction of *C. reinhardtii*, accounting for all pathways and metabolic functions indicated by the latest release of the genome (JGI v4.0) combined with our in-house generated functional annotation. The reconstruction accounts for 978 genes, associated with 1671 reactions, and includes 1029 unique metabolites. As the most comprehensive metabolic network reconstruction of *C. reinhardtii* to date, ours is the first to account for multiple wavelengths of light involved in metabolism and includes considerable expansion of fatty acid metabolism over previous reconstructions, with pathway details accounting for metabolism of individual R-groups. Further, the metabolic network reconstruction presented here provides a greater level of compartmentalization than existing reconstructions of *C. reinhardtii*, with the inclusion of the lumen as a distinct component of the chloroplast for photosynthetic functionality, and the eyespot used to guide the flagella in phototaxis.

We present simulations under a variety of growth conditions (e.g. acetate/no acetate, light/no light, aerobic/anaerobic), and physiological validation of *in silico* gene knockout against known mutant data for a variety of phenotypes (e.g. increased use of acetate; light; CO_2; nitrogen; and other media components, amino acid requiring, altered color). We further present detailed simulations demonstrating how photon absorption and different wavelengths of light affect downstream metabolic processes, elucidating the benefits of sunlight versus artificial light conditions. Our well-validated and comprehensive genome-scale reconstruction of *C. reinhardtii* metabolism provides a valuable quantitative and predictive resource for metabolic engineering toward improved production of biofuels and other commercial targets.

124

Identifying the Metabolic Potential of *Chlamydomonas reinhardtii* by Large-Scale Annotation of its Encoded Open Reading Frames

Lila Ghamsari[1]* (Lila_Ghamsari@dfci.harvard.edu), Balaji Santhanam,[1] Roger L. Chang,[2] Ani Manichaikul,[3] Erik F.Y. Hom,[4] Marc Vidal,[1] **Jason A. Papin**,[3] and **Kourosh Salehi-Ashtiani**[1]

[1]Center for Cancer Systems Biology and Dept. of Cancer Biology, Dana-Farber Cancer Institute, and Dept. of Genetics, Harvard Medical School, Boston, Mass.; [2]Dept. of Bioengineering, University of California, San Diego; [3]Dept. of Biomedical Engineering, University of Virginia, Charlottesville; and [4]Dept. of Molecular and Cellular Biology, Harvard University, Cambridge, Mass.

http://ccsb.dfci.harvard.edu
http://bme.virginia.edu/csbl

Project Goals: The release of the complete genome sequence of *C. reinhardtii* has made this unicellular alga an ideal model for metabolic engineering; however, the annotation of the relevant genes has not been verified yet and the much-needed metabolic network model is currently unavailable. Using the integrated annotation and metabolic network modeling that we recently established (*Manichaikul et al., Nature Methods 2009*), we are engaged in efforts to: 1) assign enzymatic functions to the annotated proteome of *C. reinhardtii*, 2) experimentally verify or refine the structure of the annotated open reading frames (ORFs), and 3) build a genome-wide metabolic network model for the organism based on the assigned metabolic functions (*please also see our accompanying abstract by R.L. Chang et al.*).

Results: We used the new *JGI* "filtered transcript models" (Chlre4_best_transcripts and Chlre4_best_proteins), and the *Augustus 5* models released through the *JGI* portal (http://genome.jgi-psf.org/Chlre4/Chlre4.home.html) for both functional assignments and structural annotation verifications. Enzymatic functional assignments were made by associating Enzyme Classification (EC) numbers through reciprocal blast searches against Uniprot enzyme database (with over 100,000 protein entries). The best match for each translated ORF was identified (with an e-value threshold of 10^{-3}) and the EC number from the Uniprot best match was transferred on to the ORF. We extended the EC assignments to the respective paralogs of the ORFs by clustering ORFs using BLASTCLUST (sequence identity cut-off of 35% and sequence length cut-off of 70%) within each annotation group (i.e., *Augustus* 5 and *JGI filtered models*). Altogether, we were able to assign 970 EC numbers to 1,448 *JGI* and to 1,877 *Augustus* models. Over 93% of the EC terms were assigned to both *JGI* and *Augustus* models (**Fig. 1A**). We then carried out all possible pairwise alignments between the *JGI* and *Augustus* transcripts that had been assigned the same EC numbers by the above-mentioned

procedure. In contrast to the high overlap between the two models in terms of EC assignments, less than half of each set were found to be 100% identical in sequence (**Fig. 1B**), indicating that the structural annotation of many of the two sets differ from one another.

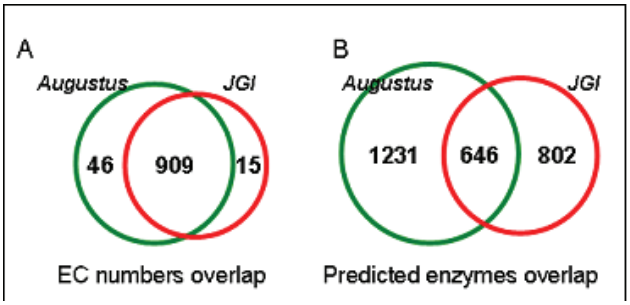


Figure 1. Annotation and overlap of *JGI* and *Augustus* ORF models.

To experimentally verify the structure of both *JGI* and *Augustus* ORF models, we carried out open reading frame (ORF) verification by RT-PCR on all ORFs that we had assigned EC numbers to (as well as a set of positive control ORFs). Following optimization of the RT-PCR procedure for high GC content of the *C. reinhardtii* transcriptome, we were able to observe positive RT-PCR products for approximately 70% of the transcripts. Following cloning, we carried out 454FLX sequencing of the ORFs and aligned the 454FLX reads to the ORF reference sequences (**Fig. 2**). We obtained 95-100% coverage of the ORF length for 940 of the ORF models; 215 ORFs with 50 to 95% coverage, 207 with 10 to 50%, and less than 10% coverage for the remainder. These results indicate that at least half of the transcript models are accurately annotated. Cloning and parallel sequencing of the remaining *JGI* metabolic ORFs are in progress.

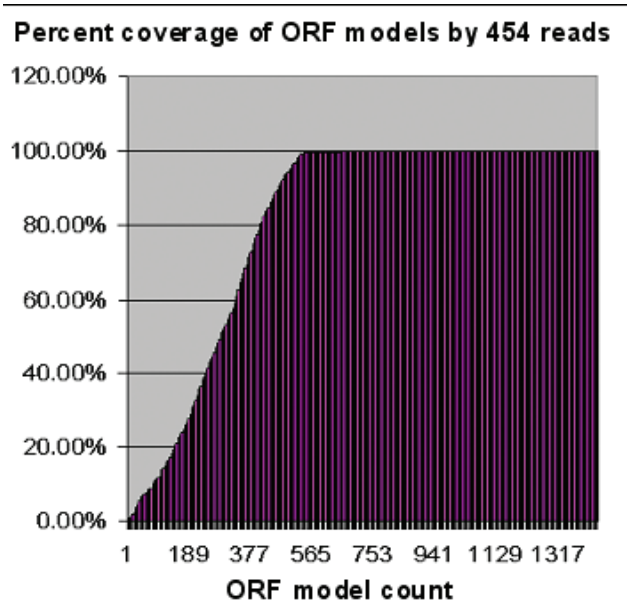


Figure 2. Sequence verification of metabolic ORF models. 1,877 ORF models were tested by RT-PCR, recombinational cloning and 454 FLX sequencing.

The verified metabolic ORF clone resource that we have generated will be made available without restrictions to the research community.

125

Pathway of Fermentative Hydrogen Production by Sulfate-Reducing Bacteria

Kimberly L. Keller[1,2]* (kellerkl@missouri.edu), Steve D. Brown,[3,2] Iris Porat,[3] and **Judy D. Wall**[1,2] (WallJ@missouri.edu)

[1]University of Missouri, Columbia; [2]VIMSS (Virtual Institute of Microbial Stress and Survival) http://vimss.lbl.gov/, Berkeley, Calif.; and [3]Oak Ridge National Laboratory, Oak Ridge, Tenn.

Project Goals: The production of hydrogen by fermentative pathways of the anaerobic sulfate-educing 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.

* Presenting author

We are exploring the production of hydrogen by fermentative pathways of the anaerobic sulfate-reducing bacteria (SRB) of the genus *Desulfovibrio*. The SRB have not been considered model organisms for hydrogen production, yet they have a multifaceted hydrogen metabolism. Strains of *Desulfovibrio* can ferment organic acids in the absence of terminal electron acceptors and produce rather large amounts of hydrogen. We believe a study of the limitations to hydrogen production in these model organisms (in particular, *Desulfovibrio* G20 and *Desulfovibrio vulgaris* Hildenborough) may be informative to decipher the flow of electrons in those organisms chosen for industrial application for hydrogen production. We have proposed to:

- Determine the contribution of fermentation to the respiratory energy budget as well as the contribution of respiration processes to the fermentation of pyruvate. Ultimately, we seek to separate these pathways and deduce the dependence of the bacterium on these combined processes.
- Identify the enzyme (s) responsible for the oxidation of pyruvate to elucidate the reduced substrate initially available for hydrogen generation during fermentation.
- Eliminate alternative electron sinks that are likely to compete for protons, reducing the overall yield of hydrogen.

An analysis of metabolic end products of *Desulfovibrio* G20 cultures grown on pyruvate as the sole electron donor and acceptor has indicated that respiration and fermentation occur simultaneously. Conversely, cells respiring sulfate with lactate also apparently gain energy by substrate level phosphorylation. We are attempting to delete the gene encoding acetate kinase to confirm that this enzyme is essential for substrate-level phosphorylation during pyruvate fermentation and determine the effect of this deletion on the efficiency of pyruvate and lactate respiration.

To determine what pathways are present for energy generation during respiration or fermentation, we have initiated an analysis of the proteomics in each condition. Cultures from wild-type *Desulfovibrio* G20 grown by pyruvate fermentation and by sulfate respiration with lactate are being used to identify enzymes which appear differentially expressed between growth modes, as well as those are found in high abundance in both growth modes. These results provide candidates for further exploration

The construction of a Tn5 transposon mutant library in *Desulfovibrio* G20 is allowing us to establish the importance of the candidate enzymes identified by the proteomics. The mutants are grown on different substrates and their growth and metabolites are compared to those of wild-type *Desulfovibrio* G20. We have already begun an analysis of hydrogen production and metabolite changes of a number of G20 Tn*5* mutants, including those lacking a fumarate reductase, molybdopterin oxidoreductase, formate dehydrogenases, and malic enzyme. No growth differences where found between wild-type and mutants lacking formate dehydrogenases or malic enzyme mutants, probably because there are apparently multiple isozymes encoding these enzymes that could compensate for the single enzyme loss. As expected, there was no succinate measured for the mutant with the Tn5 inserted in the fumarate reductase. In addition, the fumarate reductase mutant was unable to grow when fumarate only was provided as the sole electron donor and acceptor. This lack of fumarate dismutation has also been observed with a *Desulfovibrio* G20 plasmid insertion mutation in the cycA gene which encodes the type-1 tetraheme cytochrome c3.

Advances continue to be made in the genetic manipulation *D. vulgaris*, including a markerless deletion system allowing the sequential deletion of multiple genes. The system, which uses the *upp*-encoded uracil phosphoribosyltransferase as an element for counterselection, is being refined to increase efficiency of mutant selection. Preliminary experiments show the wild-type G20 strain is quite sensitive to 5-fluorouracil and attempts to make *Desulfovibrio* G20 more amenable to gene specific mutations continues. We are currently in the process of generating a barcoded transposon library in *Desulfovibrio vulgaris* Hildenborough.

126

Transcriptomic Analyses of the Sulfate-Reducing Bacterium *Desulfovibrio* G20 and a Type-1 Tetraheme Cytochrome c3 Mutant during Their Transitions into Stationary Phase

Iris Porat[1]* (porati@ornl.gov), Kimberly L. Keller,[2,3] Barbara J. Giles,[2] Timothy J. Tschaplinski,[1] Steve D. Brown,[1] and **Judy D. Wall**[2,3] (WallJ@missouri.edu)

[1]Oak Ridge National Laboratory, Oak Ridge, Tenn.; [2]University of Missouri, Columbia; and [3]VIMSS (Virtual Institute of Microbial Stress and Survival) http://vimss.lbl.gov/, Berkeley, Calif.

Project Goals: The production of hydrogen by fermentative pathways of the anaerobic sulfate-reducing bacteria (SRB) of the genus *Desulfovibrio*. is the focus of the project. The limitations to hydrogen production identified in these model organisms may be informative for those microbes chosen for industrial hydrogen generation. We propose to determine the contribution of substrate-level phosphorylation to respiratory growth on sulfate and the contribution of respiration to fermentation of pyruvate. The enzymes for pyruvate oxidation will be established in two strains of *Desulfovibrio* to identify the reduced product available for hydrogen generation. Electron sinks potentially competing with protons will be eliminated individually and together to determine the plasticity of electron flow to hydrogen.

Fermentative hydrogen generation provides a mechanism for anaerobic microbes to release electrons in a neutral fashion during oxidation of vast quantities of organic matter. We are examining the capacity of a soil anaerobe, the sulfate-reducing bacterium *Desulfovibrio* G20, to generate hydrogen from organic acids. These apparently simple pathways have yet to be clearly established. Previously, an insertion muta-

tion in *cycA,* encoding the type-1 tetraheme cytochrome c_3, was constructed in *Desulfovibrio* G20 (Rapp-Giles et al. 2000). The growth rate of the CycA$^-$ mutant was similar to that of the wild type during lactate-supported sulfate respiration or pyruvate fermentation. However, CycA$^-$ appeared to be unable to respire sulfate with pyruvate as the electron donor and it was unable to dismutate fumarate. Interestingly, CycA$^-$ growing with lactate/sulfate generated more hydrogen gas than G20, suggesting that electron flow was rerouted.

To determine any changes in gene expression that might account for these differences, both *Desulfovibrio* G20 and the CycA mutant were grown in pH-controlled fermentors in define medium by lactate-supported sulfate respiration or by pyruvate fermentation. RNA was prepared and microarray analysis performed to determine differential gene expression between exponential and stationary phase G20 cells grown by respiration or by fermentation and then determine any differences in gene expression in the CycA mutant grown under comparable conditions. The transcriptomic results demonstrated gene expression profile changes of G20 in response to carbon and energy depletion upon entering stationary phase as well as those occurring in response to pyruvate fermentation versus lactate respiration with sulfate. Hierarchical cluster analysis of transcriptomic profiles for G20 and the CycA mutant revealed 12 distinct clusters. Also, the arrays were clustered by growth phases. As would be predicted, in both strains genes involved in translation and transcription were expressed at high levels during the log phase. In addition, some genes for biosynthesis of amino acids, co-enzymes, lipid and carbohydrates were up-regulated during logarithmic growth. During the stationary phase, large numbers of genes encoding proteins for energy generation were up-regulated, including hydrogenases, dehydrogenases and ATP synthase. Also, once the electron donors were exhausted, genes involved in the flagella biosynthesis and in stress responses were increased in expression.

Interestingly, two well defined clusters with 14 ORFs each showed significant transcript differences between the two strains and the two media. In the CycA mutant, one of the clusters that included a large operon encoding the enzymes for conversion of pyruvate to succinate was down regulated. The second cluster showed decreased expression in pyruvate medium compared to lactate sulfate medium in both the log and stationary phases for G20 as well as for the CycA mutant. That cluster encodes two operons, one containing the type II c_3 transmembrane complex genes and a second with other energy related protein genes.

Metabolic profiles were obtained for G20 and CycA cultures in respiring and fermenting cultures. Comparing the metabolites from the CycA mutant to those from G20 in lactate/sulfate medium, higher concentrations of glycerol-1-phosphate, stearate, citramalate, aspartate, glycine and nucleotides were detected regardless of growth phase. Lower concentrations of lactate and succinate were found in the mutant. In contrast, in pyruvate fermenting cultures, the concentrations of glutamate, lactate, aspartate, N-acetylaspartate, glycine, and 5-oxo-proline were higher in the CycA mutant than in G20 in log phase and opposite in stationary phase. The reverse was found for concentrations of succinate and trehalose in the mutant, lower than those from G20 in log phase and higher in stationary. These intracellular changes may also reflect overall redirection of substrates and electron flow in the two strains.

Reference

1. Rapp-Giles, B. J., L. Casalot, R.S. English, J.A. Ringbauer, Jr., A. Dolla, and J.D. Wall. 2000. Cytochrome c_3 mutants of *Desulfovibrio desulfuricans*. Appl. Environ. Microbiol. 66:671-677.

Student Presentation

127

Genome Resequencing of Thermoacidophilic Archaeal Carbon Flux Mutants

Y. Maezato* and **P. Blum** (pblum1@unl.edu)

Beadle Center for Genetics, University of Nebraska, Lincoln

Project Goals: The project (DE-FG36-08G088055) supports research on the ability of extremely thermoacidophilic microbes and their proteins to accelerate lignocellulose processing from switchgrass and to maximize sugar release. Delineation of carbon metabolic pathways in *Sulfolobus solfataricus* (*Sso*) is an essential step towards engineering traits associated with lignocellulosic bioprocessing. In *Sso*, carbon catabolism varies between two phenotypic states. In car mutants, growth on hexoses is blocked while pentose catabolism remains unaffected. In light of this characteristic, identification of car was undertaken.

The project (DE-FG36-08G088055) supports research on the ability of extremely thermoacidophilic microbes and their proteins to accelerate lignocellulose processing from switchgrass and to maximize sugar release. Delineation of carbon metabolic pathways in *Sulfolobus solfataricus* (*Sso*) is an essential step towards engineering traits associated with lignocellulosic bioprocessing. In *Sso*, carbon catabolism varies between two phenotypic states. In *car* mutants, growth on hexoses is blocked while pentose catabolism remains unaffected. In light of this characteristic, identification of *car* was undertaken. To genetically map *car* and its regulated targets, a genomic large insert library (BACs) was sorted by selection for clones that restored hexose metabolism in the *car* mutant. Shotgun subclone libraries led to the identification of a glycolyltic gene, *kdgK*, and a chromatin modification gene, *hdac-1*, as components of the *car* mutant glycolytic defect. Genome resequencing was conducted to further characterize the genetic basis for *car*. Nearly 50 differences were identified distinguishing genomes of the *Sso* wild type and *car* mutant derivative. These included the CRISPR associated RAMP module Cas gene cmr2. Additional three-way whole genome comparisons further narrowed strain differences. Detailed genotyping combined with genomic reconstruction is underway to further assess these whole genome sequencing results.

128

Metabolic Fluxes: Quantifying Competition Between Nitrate and Proton Reduction during Fermentative NAD (P)H Formation in Real-Time

Gennady Ananyev,* Nicholas J. Skizim, and **G. Charles Dismukes** (dismukes@rci.rutgers.edu)

Waksman Institute of Microbiology and Dept. of Chemistry and Chemical Biology, Rutgers University, Piscataway, N.J.

Project Goals (Abstracts 128-129): The goal of this project is the design and utilization of tools which will allow a quantitative understanding of the mechanism and kinetics of hydrogen production in cyanobacteria.

The ability to continuously monitor metabolite concentrations *in vivo* as they undergo transient changes in response to environmental perturbations is among the most sought goals of metabolomics. Here we provide direct measures of the quantitative flux of intracellular reductant and its temporal dynamics between competing pathways. We have developed a fluorometric method for monitoring the concentration of intracellular reduced pyridine nucleotides, NAD (P)H, and combined it with a real-time electrochemical assay of dissolved H_2 concentration for simultaneous measurements on whole cells. Here we describe an application to quantitative kinetics in the cyanobacterium *Arthrospira maxima,* the most prolific fermentative H_2 producer in this group. *A maxima* produces H_2 via a bidirectional NiFe-hydrogenase, the sole H_2-metabolizing enzyme in this organism. We demonstrate that the two temporal phases of H_2 production induced following the onset of anaerobiosis arise from distinct metabolic processes responsible for (a) anabolic production of NADPH (Phase-1) and (b) catabolic production of NADH (Phase-2). Phase-1 starts within minutes of anaerobiosis and may extend up to 2-3 h. Phase-1 H_2 is shown to correlate with the residual pool of photosynthetically produced NADPH; it decreases with increasing aerobic dark time prior to anaerobiosis. The second phase starts within 3-20 h and follows the rise of intracellular NAD (P)H with a short lag time (Δt = 24 min) indicative of the time to achieve a redox poise ([NAD (P)H//[NAD $(P)^+$]>100) sufficient for H_2 formation. The major Phase-2 H_2 is produced by autofermentation of carbohydrate reserves. The yield of Phase-2 H_2 is shown to be inversely related to the concentration of NO_3^- in the medium. A positive linear relationship is observed between the NO_3^- concentration and the delay time for onset of Phase-2 H_2 production over more than two decades change in concentration. Experiments carried out at various extracellular concentrations of inhibitors of nitrate reductase (N_3^-, CN^-) provide solid evidence for a direct competition for consumption of cellular reductant by hydrogenase and nitrate reductase (*nar*). This evidence provides a firm basis for proposed metabolic engineering of pathways.

Supported by DOE-GTL DE-FG02-07ER64488 and AFOSR-MURI FA9550-05-1-0365.

129

Real-Time Co-Detection of Dissolved H_2 and Intracellular NAD (P)H Concentrations in Microbes

Gennady Ananyev, Nicholas J. Skizim,* and **G. Charles Dismukes** (dismukes@rci.rutgers.edu)

Waksman Institute of Microbiology and Dept. of Chemistry and Chemical Biology, Rutgers University, Piscataway, N.J.

Project Goals: See goals for abstract 128.

In an effort to probe the kinetics and quantify the concentration of hydrogen produced by living microorganisms, here we describe a home-built combination cell for simultaneous real-time detection of dissolved H_2 and reduced pyridine nucleotide levels, [NAD (P)H]. Electrochemical H_2 detection (LOD: 2 nC = 1 nM H_2) from live cells with sample volumes as small as 5 µl is accomplished with a custom made Clark-type cell (or reverse fuel cell) comprised of a membrane covered Pt/Ir electrode poised at a bias of +220 mV . Total reduced pyridine nucleotide = [NADPH] + [NADH] is assayed by selectively exciting NAD^+ and $NADP^+$ through an ultraviolet light emitting diode (365 nm) operating in pulse mode, and subsequently measuring fluorescence emission at 470± 20 nm with a photodiode. Calibration curves with standards reveal linear responses for both H_2 and NAD (P)H detection using inactivated microbes as background, from which quantitative concentrations could be assessed. Using these tools we have been able to observe for the first time multiple temporal phases of H_2 production in anaerobically poised cyanobacteria which possess a [NiFe]-hydrogenase, as well as those containing a nitrogenase. We show that two main phases of hydrogen production correlate with the availability of residual NADPH produced via photosynthesis prior to the onset of anaerobiosis (phase 1), and NADH produced by anaerobic autofermentation of glycogen reserves (phase 2) in strains containing a [NiFe]-hydrogenase as their only hydrogen metabolizing enzyme. Application of these tools to N_2 fixing cyanobacteria has similarly revealed at least three kinetic phases of dark hydrogen production coupled to fermentative metabolism. Experiments are planned on two diazotrophic *Synechococcus* thermophiles which lack all types of hydrogenase. Together these data shall provide a basis for discriminating between and quantifying the H_2 evolution pathways involving hydrogenase and nitrogenase.

Supported by DOE-GTL DE-FG02-07ER64488 and AFOSR-MURI FA9550-05-1-0365

130

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

Harry Christman,[1] Elsie Campbell,[1] Douglas Rissser,[1] Brett Phinney,[2] Wan-Ling Chiu,[3] and **John C. Meeks**[1]* (jcmeeks@ucdavis.edu)

[1]Dept. of Microbiology and [2]Genome Center, University of California, Davis; and [3]Dept. 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_2, is uncoupled from growth.

We hypothesize that if the frequency of heterocysts, sites of nitrogen fixation and hydrogen evolution, can be increased approximately 3-fold above the less than 10% normally found in filaments, then heterocyst-forming cyanobacteria would be applicable for cost effective photo-biohydrogen production. In *Nostoc punctiforme*, the heterocyst frequency of 8% in the free-living state is increased to 30-35% when in symbiotic association with terrestrial plants, such as the hornwort *Anthoceros* spp. and the angiosperm *Gunnera* spp. We are applying transctriptomic and proteomic analyses with wild-type and mutant strains to identify the regulatory circuits of free-living heterocyst differentiation and how those circuits have been co-opted during symbiotic growth.

Transcriptomics. HetR and HetF are two positive regulatory elements in the differentiation of heterocysts. Mutations in either structural gene in *N. punctiforme* lead to the inability to differentiate heterocysts in free-living cultures or to support N_2-dependent growth of the hornwort plant partner, although the mutants do infect the plant. Conversely, over expression of the genes from a multicopy plasmid in trans, results in the differentiation of multiple heterocysts that are localized in distinct clusters with a nonrandom spacing pattern in the filaments. The heterocyst frequency can reach 40% of the cells, but the heterocysts do not fix nitrogen in support of vegetative cell growth. A question is whether the gene products operate at the same pathway in the regulatory cascade of heterocyst differentiation. Time course (0.5 to 24 h) DNA microarray experiments, using a Nimblegen platform, with the data analyzed in the R statistical environment and subsequently clustered using the Genesis program, yielded two interesting results. A total of 124 genes were up-regulated in both mutants in responsive to combined nitrogen limitation. Only one known up-stream gene for heterocyst differentiation gene (*nrrA*) was up-regulated. The up-regulated genes include those for assimilation of the alternative inorganic nitrogen sources nitrate and urea. Only one amino acid transport gene was up-regulated, implying *N. punctiforme* does not search for organic nitrogen sources during nitrogen starvation. An additional common cluster of 330 genes were down-regulated in the two mutants and these included many encoding proteins of central metabolism, protein synthesis and photosynthetic energy metabolism. Approximately 53% of the up- and down-regulated transcripts encode proteins of unassigned function. The second result of interest is that 130 and 775 genes were uniquely up-regulated, and 29 and 1063 were uniquely down-regulated in the *hetF* and *hetR* mutants, respectively. These results imply that HetF and HetR have physiological roles in addition to heterocyst differentiation, which is consistent with the presence of the genes in filamentous cyanobacteria that do not differentiate heterocysts. We are now examining the transcriptional profiles of wild type strains over expressing HetR and HetF in order to identify genetic targets and any differences in the targets in the two constructs for comparisons to the loss of function mutant data. We have also initiated analysis of a unique pattern mutant in which a high heterocyst frequency (~ 35%) is manifest as multiple singular heterocysts with a 3-4 vegetative cell spacing between heterocysts. This is the pattern we observe in the symbiotic growth state. However, this mutant also does not grow with N_2 as the sole nitrogen source.

Proteomics. We have completed a single Mudpit-based MS/MS run of the total proteome of a N_2-grown culture of *N. punctiforme*. The cell extract was processed into three fractions; an initial 14,000 x g pellet of membrane proteins, a 150,000 x g supernatant of soluble proteins and a 150,000 x g pellet of primarily carboxysomes and phycobilisomes. The fractions were sub-fractionated by PAGE, sliced sections digested with trypsin and the eluted products further fractionated by 2 dimensional LC followed by electrospray injection. This 600 member proteome is comparable to one we previously defined of the supernatant proteins of an ammonium-grown culture, except for an enrichment in membrane proteins, plus proteins associated with heterocysts and nitrogenase function. Notable is the constitutive presence of DNA photolyases and enzymes for metabolism of reactive oxygen species, stress factors to which cyanobacteria are constantly exposed. The value of this proteome will be enhanced by comparison to a N_2 plus fructose grown mixotrophic culture now in progress. In the symbiotic growth state *N. punctiforme* grows as a photomixotroph in *Anthoceros* and most likely as a heterotroph in *Gunnera*. Thus, these proteomes are essential for comparisons to symbiotic growth with the accompanying high heterocyst frequency and high rates of nitrogenase activity.

* Presenting author

Student Presentation

131

Contributors to Light- and Feedstock-Powered Hydrogen Production in *Rhodobacter sphaeroides*

Melanie A. Spero[1]* (spero@wisc.edu), Ugur Sohmen[2]* (sohmen@wisc.edu), Wayne S. Kontur,[1] L. Safak Yilmaz,[2] **Daniel R. Noguera**[2] (noguera@engr.wisc.edu), and **Timothy J. Donohue**[1] (tdonohue@bact.wisc.edu)

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

Project Goals: Our research seeks to generate a quantitative understanding of light- and feedstock-powered hydrogen (H_2) production in the photosynthetic bacterium *Rhodobacter sphaeroides*. We will determine the source (s) of H_2 production in *R. sphaeroides* and determine the effect on fuel production of eliminating or minimizing processes predicted to siphon reducing power from H_2 production. This will be accomplished by deleting genes in pathways predicted to impact H_2 production and assaying for the amount of reductant that is partitioned to H_2, biomass, as well as other electron sinks. Mutations tested to date include those in genes encoding enzymes predicted to generate H_2 (nitrogenase), detract from H_2 production (hydrogenase), and in pathways predicted to compete for reducing power that could be utilized for H_2 production (synthesis of polyhydroxybutyrate, carbon dioxide fixation). This work also compares H_2 production in log-phase batch cultures to late stationary phase cultures, where the amount of reducing power shuttled into biomass production is minimized. Each mutant will be assayed for H_2 production under active or non-growing conditions on a variety of carbon sources to determine the impact of these cellular processes on H_2 production.

These studies are performed in the α-proteobacterium *R. sphaeroides*, the most-studied photosynthetic bacterium, since it is known to produce relatively large amounts of H_2 under photoheterotrophic growth conditions. A working model of the processes predicted to impact H_2 production was developed based on prior knowledge of metabolic pathways, the genome sequence of this bacterium, and global gene expression data (see posters by Yilmaz et. al., and Kontur et al.).

H_2 production in *R. sphaeroides* is proposed to be mainly or completely associated with the nitrogenase enzyme. To determine the contribution of nitrogenase to H_2 production, a mutant strain of *R. sphaeroides* was generated that lacks the structural genes of nitrogenase (*nifHDK*). The growth properties of the ΔNif mutant support the hypothesis that nitrogenase-mediated H_2 production is an important electron sink with some, but not all, carbon feedstocks. Experiments are ongoing to use the ΔNif mutant to determine if other *R. sphaeroides* enzymes can contribute to H_2 production under defined conditions.

One enzyme predicted to detract from H_2 production is uptake hydrogenase (Hup), which can oxidize H_2 into protons and electrons. A mutant strain of *R. sphaeroides* lacking the Hup structural genes (*hupSL*) is being used to determine the impact of Hup activity on H_2 production, especially since global gene expression data predict a wide range of Hup activities present when cells grow in media containing different carbon sources as feedstocks.

We are also testing the effects of other electron sinks on H_2 production, by blocking pathways known or predicted to compete for reducing power. For example, mutant strains with gene deletions in polyhydroxybutyrate synthesis or carbon dioxide sequestration (via the Calvin cycle) are being analyzed to determine effects on H_2 production. Each of these mutants is being assayed for H_2 production on various carbon sources in order to evaluate how altering the flow of reducing power into other electron sinks can alter the distribution of reducing power into various products.

Because biomass is as a significant electron sink in exponentially growing cells, the distribution of electrons into different products is also being analyzed in stationary phase cultures, when electron flow towards biomass production is predicted to be minimized. Preliminary results indicate that partitioning of reducing power to H_2 production varies between growing and stationary phase cultures on some carbon feedstocks. We will report on experiments with wild type and mutant strains that seek to monitor electron partitioning during stationary phase to better understand how feedstock supply impacts the flow of electrons to H_2 and other electron sinks.

132

SurR Regulates Hydrogen Production in *Pyrococcus furiosus* by a Sulfur-Dependent Redox Switch

Gina L. Lipscomb,[1]* Hua Yang,[1,4] Annette Keese,[2] Gerrit J. Schut,[1] B.C. Wang,[1] Michael Thomm,[2] **J**anet Westpheling,[3] Michael W.W. Adams,[1] and **Robert A. Scott**[1] (rscott@uga.edu)

[1]Dept. of Biochemistry and Molecular Biology, University of Georgia, Athens; [2]Dept. of Microbiology, University of Regensburg, Germany; [3]Dept. of Genetics, University of Georgia, Athens; and [4]Centers for Disease Control and Prevention, Atlanta, Georgia

Project Goals: The goal of SAPHyRe (Systems Approach to Probing Hydrogen Regulation) is to develop a detailed systems-level description of the regulatory and metabolic networks controlling hydrogen production in the hyperthermophilic archaeon *Pyrococcus furiosus* (*Pf*). *Pf* will be used as the model organism to investigate its response to various environmental conditions relevant to *all* hydrogen-producing microorganisms, such as carbon and nitrogen sources, metal availability, and oxidative and reductive stresses. The outcome of this project will

serve two purposes: 1) it will bring us one step closer to utilizing *P. furiosus* in development of H_2 as an alternative energy source and 2) it will serve as a model methodology for investigating the regulatory pathways of hydrogen production in other organisms.

We present structural and biochemical evidence for a redox switch in the archaeal transcriptional regulator SurR of *Pyrococcus furiosus*, a hyperthermophilic anaerobe. *P. furiosus* produces H_2 during fermentation, but undergoes a metabolic shift to produce H_2S when elemental sulfur (S^0) becomes available. Changes in gene expression occur within minutes of S^0 addition, and the majority of these S^0-responsive genes are regulatory targets of SurR, a key regulator involved in primary S^0 response. SurR was shown *in vitro* to have dual functionality, activating transcription of some of these genes, notably the hydrogenase operons, and repressing others, including a gene encoding sulfur reductase. This work demonstrates that the activity of SurR is modulated by cysteine residues in a CxxC motif that constitute a disulfide switch. Oxidation of the switch with S^0 inhibits sequence-specific DNA binding by SurR, leading to *de*activation of genes related to H_2 production and *de*repression of genes involved in S^0 metabolism.

133

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

Jonathan Millen[1]* (jonathan_millen@hotmail.com), Ji Hye Jo,[2] Shiv Thammannagowda,[2] Grant Pennington,[2] Matthew Heckman,[1] Pin-Ching Maness,[2] Mary Lipton,[3] **J.H. David Wu**[1] (davidwu@che.rochester.edu)

[1]University of Rochester, N.Y.; [2]National Renewable Energy Laboratory, Golden, Colo.; and [3]Pacific Northwest National Laboratory, Richland, Wash.

Project Goals: The overall objective of this research is to understand H_2 metabolic pathways in *Clostridium thermocellum* and the underlying regulatory network at the molecular and systems levels. We hypothesize that: multiple hydrogenases work concertedly to contribute to hydrogen evolution; transcription factors control hydrogenase expression at the transcription level; and hydrogen metabolism is coupled to cellulolysis and fermentative pathways in this bacterium. We are determining the hydrogenase expression and metabolic network nodes on cells subjected to different culture conditions and metabolic pathway inhibitors, to probe differential expression of the various hydrogenases and their interrelationship with other cellular metabolic pathways. We are also mapping connections in the transcription factor network controlling linked metabolic pathways. Finally, we will purify FeFe-hydrogenases from its native producer and heterologously expressed *E. coli* to determine their subunit compositions, endogenous redox partners, and the direction of reaction (hydrogen production vs. uptake) to shed light on their roles in hydrogen metabolism.

Clostridium thermocellum, a thermophilic, ethanogenic, and cellulolytic anaerobe, produces a complex extracellular cellulolytic organelle called the cellulosome. The cellulosome contains various depolymerizing enzymes that are arrayed on a protein scaffold and effectively degrades complex cellulosic substrates. During cellulose fermentation, the bacterium evolves hydrogen at a high rate. Analysis of its genome sequence reveals the existence of at least three putative hydrogenase genes (CtHydA1, CtHydA2 and CtHydA3) central to hydrogen metabolism. Furthermore, at least 20 genes are potentially related to hydrogen metabolism. The bacterium is thus remarkably versatile in employing various enzymes, some of which are potentially novel, for hydrogen metabolism, indicating the significance of this biological process. Yet little is known concerning the pathway for hydrogen production and the underlying regulatory mechanism/network that control these hydrogenase and the related genes as well as cellulolytic process and other metabolic pathways in the organism.

Due to the novelty of CtHydA3, it was selected for the initial study. Genes encoding CtHydA3, a ferredoxin-like protein (Ct_3004), and three FeFe-hydrogenase maturation proteins (CtHydE, CtHydF, and CtHydG) have been cloned into three plasmids and co-transformed into *E. coli* strain Rosetta (DE3) and BL21 (DE3) for heterologous expression. In addition, a 6X His-tag sequence was fused to either the C- or N-termini of CtHydA3. Protein immunoblots confirmed the expression of the C-terminus His-tagged CtHydA3 (73 kDa band) in *E. coli* Rosetta (DE3), but not in *E. coli* BL21 (DE3), likely due to differences in codon usage between *C. thermocellum* and *E. coli*. The expression is further corroborated by a five-fold increase in *in vitro* hydrogenase activity in the soluble cell extract of the recombinant Rosetta strain, mediated by reduced methyl viologen. However, no difference in *in vivo* hydrogen production was detected in the recombinant Rosetta strain, suggesting an inability of the recombinant hydrogenase to contribute to the host's hydrogen metabolism. The C-terminal His-tagged protein failed to bind to a TALON metal affinity column to facilitate its purification. Work is underway to express the N-terminal tagged His-CtHydA3 and explore its affinity purification and characterization.

To identify transcription factors controlling metabolic pathways, we developed an affinity purification method by immobilizing promoter DNA sequences to a solid support. DNA-binding proteins from the *C. thermocellum* cell lysate, obtained by growing on cellobiose or crystalline cellulose and eluted from the affinity columns, were identified by the MALDI-TOF or LC-MS-MS techniques. Several transcription factor candidates were identified. In a reversed approach, we expressed in *E. coli* various putative DNA-binding proteins found in the genome of *C. thermocellum*. Their target DNA-binding sites will be screened by using a DNA microarray we designed and confirmed by EMSA (electrophoretic mobility shift assay).

The studies will provide important insights into the pathway and regulatory mechanism/network controlling hydrogen metabolism and cellulolysis as well as other pertinent metabolic pathways in this very intriguing cellulolytic and thermophilic bacterium, which catalyzes the rate-limiting cellulose-degradation reaction in a single-step process of biomass conversion (or CBP, Consolidated Bioprocessing). Detailed understanding of the pathway and regulatory mechanism/network will ultimately provide rationales for engineering, alternating, or deregulating the organism for biomass conversion to liquid and hydrogen fuels.

134

Filling Knowledge Gaps in Biological Networks: Integrated Global Approaches to Understand Biofuel Metabolism in *Chlamydomonas reinhardtii*

Venkataramanan Subramanian,[1,2] Alexandra Dubini,[1] Claudia Catalanotti,[3] Wenqiang Yang,[3] Christopher H. Chang,[1] Ambarish Nag,[1] David M. Alber,[1] Peter A. Graf,[1] Glenn Murray,[2] Kwiseon Kim,[1] Arthur R. Grossman,[3] Matthew C. Posewitz,[2] and **Michael Seibert**[1]* (mike.seibert@nrel.gov)

[1]National Renewable Energy Laboratory, Golden, Colo.; [2]Colorado School of Mines, Golden, Colo.; and [3]Carnegie Institution of Washington, Stanford, Calif.

Project Goals: Development of photobiological, biofuels-production processes, a key component of DOE's renewable energy mission, would be accelerated by improved understanding of cellular metabolism and its regulation. Currently, a lack of comprehensive experimental data hinders the development of reliable metabolic models that have robust predictive capabilities. Therefore, we are employing high-performance computing to (1) estimate parameters that delimit the space of stable solutions for experimentally constrained metabolic models, (2) explore network capabilities in silico, and (3) integrate experimental, systems biology data to verify and refine metabolic models. Biochemical reactions are modeled with the fundamental independent variables being enzyme concentrations and Michaelis-Menten parameters. Through iterative model building and an understanding of cellular metabolism obtained from high throughput "omics" data, we are constructing metabolic models that link individual enzyme reactions and the activities of specific metabolic pathways with the production of biofuels in the green alga, *Chlamydomonas reinhardtii*. High-performance simulation and optimization will predict metabolic outputs based on the kinetic parameters governing individual reactions, identify metabolic limitations, and predict specific manipulations that are likely to improve biofuel outputs. The work is envisioned as an important contribution toward strengthening our knowledge of energy-related biosystems.

The goal of this project, jointly funded by the DOE Computational Biology and SciDAC Programs, is to develop a means to globally map, *in silico*, all biological pathways in *Chlamydomonas reinhardtii* (*Chlamydomonas* throughout) that can impact the production of H_2 and other biofuels. *Chlamydomonas* is the first alga for which a fully sequenced genome was available (thanks to the JGI), and this organism has recently emerged as the prototype for investigating the regulation of basic metabolism, particularly fermentative processes. *Chlamydomonas* has a complex anaerobic metabolic network that can be induced in the dark and produces H_2 along with other fermentation products such as formate, acetate, ethanol, and CO_2. Previous studies have focused on determining genes (by microarray and RT-PCR analysis) that were differentially regulated as the result of shifting cultures of the CC-425 parental strain from aerobic growth to dark anaerobiosis [1]. Anoxia led to differential expression of genes involved in fermentation metabolism, and specifically caused accumulation of transcripts encoding pyruvate formate lyase (*PFL1*) and pyruvate:ferredoxin oxidoreductase (*PFR1*). Moreover, *Chlamydomonas* synthesized the metabolites formate, acetate, and ethanol in the ratio 1:1:0.5. Recent experiments have shown that in a strain lacking proteins involved in assembling an active hydrogenase enzyme (*hydEF* mutant), the metabolite ratio shifted to 2:1:1, and succinate accumulated instead of H_2. Interestingly, levels of transcripts encoding proteins involved in fermentation also changed in the mutant relative to parental cells following the imposition of anoxic conditions; the *PFL1* mRNA accumulated to a higher level and the *PFR1* mRNA to a lower level in the mutant than in parental cells [2]. These results have allowed us to generate a physiological model that highlights the flexibility of fermentation metabolism and H_2 production in *Chlamydomonas*. Additional mutants (e.g., *ΔPFL*, *ΔFMR*, *ΔHYDA2*, *ΔHYDA2:ΔPFL*, *ΔHYDA2:ΔMME4*) have been identified this year using a HTP screening procedure, and the analyses of these mutants are enabling us to further refine our knowledge of fermentation metabolism and its regulation in *Chlamydomonas*. Furthermore, transcript, protein and metabolite analyses of various single and double mutants (e.g., *ΔHYDA1, ΔADH1, ΔPDC3, ΔHYDA1: ΔHYDA2, ΔPFL1: ΔADH1)* are already underway.

To integrate the results described above with various experimental results discussed below, we have developed metabolic models which, when simulated using our recently developed software toolkit, enable us to make informed decisions about the best ways to link experimentally measurable parameters to the biochemical reactions and their regulation in *Chlamydomonas*. The multi-compartment model includes carbon metabolism (glycolysis, tricarboxylic acid cycle, starch metabolism, Calvin-Benson Cycle), oxidative phosphorylation, and fermentative metabolism. Steady-state kinetic relationships are expressed as thermodynamically consistent ordinary differential equations. Merging of component pathway models, incorporation of ionization reactions, transformation to C++, compilation of the data, and linking the information to high-performance executable programs is largely automated using our *High-Performance Systems Biology Toolkit* (HiPer SBTK) [3]. The programs being used

are for parameter sampling, data fitting, and local or global optimization within a defined space of kinetic parameters and/or enzyme concentrations. Job configuration is possible through an auxiliary graphical interface or by direct editing of simple text. The code has been modularized, permitting facile incorporation of new techniques for high-dimensional sampling, optimization, and model integration. Key future developments will include investigation of model phase space with respect to initial conditions, and model expansion to include the photosynthetic light reactions and the effects of changes in light intensity on cellular metabolism.

Our current biological research, using global trancriptomic- and proteomics-based approaches, is supporting extensive computational analyses; this research is focused on elucidating the regulation of algal fermentative pathways and identifying various pathways that either directly or indirectly impact H_2 and biofuels production. We recently used 2D-differential gel electrophoresis (DIGE) and shotgun mass spectrometry to identify, quantify and compare proteins present under anoxic, H_2-producing conditions [dark and sulfur-deprived] to those present under oxic, non-H_2-producing conditions. Preliminary DIGE results indicate that the levels of 189 proteins are similar under oxic and anoxic conditions, 41 proteins are higher during anoxic growth and 69 proteins are higher during oxic growth. Furthermore, mass-spectrometry-based, shotgun-proteomic experiments (using the LTQ-Orbitrap system) with whole cell protein extracts detected 1485 proteins under anoxic conditions, 853 proteins under sulfur-deprivation conditions, and 1664 proteins under oxic conditions; analyses of this dataset is underway. Finer resolution will be achieved in the future by proteomic dissection of specific subcellular compartments. In-depth comparative analyses of the specific proteins identified under the various conditions will be discussed.

In summary, high-throughput 'omics' techniques are being used to input transcript, protein, and metabolism knowledge (from both parental and mutant strains) into computational models that explore metabolism (and its flexibility) in the green alga, *Chlamydomonas*. The coupling of experimental results, mutant analyses, and *in silico* modeling is expected to improve our understanding of the complexity of *Chlamydomonas* metabolic networks, including ways in which the cells adjust to changing environmental conditions (i.e., by modulating metabolite fluxes), and how blocking steps in specific metabolic pathways alter metabolite flow. This information as a whole will suggest critical strategies for engineering *Chlamydomonas* metabolism for more efficient production of H_2 and/or other algal biofuels.

References

1. Mus, F., Dubini, A., Seibert, M., Posewitz M.C., and Grossman A.R. (2007) "Anaerobic acclimation in Chlamydomonas reinhardtii: Anoxic gene expression, hydrogenase induction and metabolic pathways", J. Biol. Chem. 282 (35), 25475-25486.
2. Dubini, A., Mus, F., Seibert, M., Grossman, A.R., and Posewitz M.C., (2009) "Flexibility in anaerobic metabolism as revealed in a mutant of Chlamydomonas reinhardtii lacking hydrogenase activity ", J. Biol. Chem. 284, 7201-7213.
3. Chang, C.H., Graf, P., Alber, D.M., Kim, K., Murray, G., Posewitz, M., and Seibert M., (2008) "Photons, Photosynthesis, and High-Performance Computing: Challenges, Progress, and Promise of Modeling Metabolism in Green Algae", J. Phys. Conf. Ser. 125: 012048.

135

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

Alexandra Dubini,[1] Matthew Wecker,[2] and **Maria L. Ghirardi**[1]* (maria_ghirardi@nrel.gov)

[1]National Renewable Energy Laboratory, Golden, Colo. and [2]GeneBiologics, LLC, Boulder, Colo.

Project Goals: See below.

Photobiological H_2 production from water is a clean, non-polluting and renewable technology. Although the potential light conversion efficiency to H_2 by biological organisms is theoretically high (about 10%), the system is currently limited by biochemical and engineering constraints. These limitations include (but are not restricted to) the extreme O_2 sensitivity of the biological hydrogenases and the low availability of reductants to the hydrogenase due to the existence of competing metabolic pathways.

To addresses the O_2 sensitivity issue, our research is developing a new, biologically-based assay to screen large microbial populations for improved H_2-production properties. This novel assay is based on the H_2-sensing system of the photosynthetic bacteria *Rhodobacter capsulatus*. The H_2-sensing system is being optimized as a green fluorescence protein-based reporter-assay for heterologous hydrogen production by the bacteria. It will then be used to screen for O_2 tolerant [FeFe]-hydrogenases generated through directed-evolution techniques. The hydrogenases of *Clostridium acetobutylicum, Chlamydomonas reinhardtii* and *Bacteroides thetaiotaomicron*, along with their respective assembly proteins have been introduced into broad host range vectors and are being shuttled into *R. capsulatus*.

To address the issue of competitive metabolic pathways with H_2 production, we have started using a yeast two hybrid. A complex library of 10^7 preys has been constructed and 8 different baits are being used to screen the library. Those baits include two hydrogenases and the 6 potential partner ferredoxins. We wish to deconvolute the hydrogenase interactors *in vivo* by fishing preys from the library and hopefully discover which of one the ferredoxins is the direct electron donor. We also aim at understanding the ferredoxin interaction network as they are key players in the hydrogenase pathway and are electron donor to many other competitive pathways. The screens are in progress and will help us having a better idea of the metabolic network involved in the hydrogen production pathway.

136

Transcriptome and DNA Methylome Analysis of Algae Using Ultra-High-Throughput Sequencing

David Casero Díaz-Cano,[1] Shawn Cokus,[1] Madeli Castruita,[2] Steven Karpowicz,[2] Sabeeha Merchant,[2] and **Matteo Pellegrini**[1] (matteop@mcdb.ucla.edu)

[1]Department of Molecular, Cell and Developmental Biology and [2]Department of Chemistry and Biochemistry, University of California, Los Angeles

Project Goals: Develop algal biofuels.

High-throughput sequencing has undergone remarkable increases in efficiency over the past few years. Some of the most efficient machines are those produced by Illumina, which sequence tens of billions of bases per week. We are currently using these sequencers in to study the transcriptome of algae such as *chlamydomonas*. *Chlamydomonas reinhardtii*, a unicellular eukaryote in the plant lineage, has been exploited in the laboratory over the last 50 years as a model organism for the study of eukaryotic photosynthesis. The advent of massively parallel short read sequencing technology opens the door to (near) full coverage of the *Chlamydomonas* transcript map via deep sequencing of mRNAs. To evaluate the potential of Illumina's Solexa technology for a) generating a whole transcriptome for *Chlamydomonas*, b) identifying differentially expressed genes, and c) reconstructing gene models *de novo*, we analyzed RNAs isolated from a variety of conditions. We have verified that these libraries may be used to quantitatively estimate transcript fold changes in different conditions using existing gene models. We are also developing a new annotation pipeline using only the short read sequencing data, and have shown that these approaches allow us to accurately reconstruct gene models.

137

Cell-Free Synthetic Pathway Biotransformations (SyPaB) for Producing Hydrogen and Even Fixing Carbon Dioxide

Y.-H. Percival Zhang[1,2*] (ypzhang@vt.edu), Jonathan Mielenz,[2,3] and Mike Adams[2,4]

[1]Biological Systems Engineering Dept., Virginia Tech, Blacksburg; [2]DOE BioEnergy Science Center (BESC), Oak Ridge, Tenn.; [3]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tenn.; and [4]Dept. of Biochemistry and Molecular Biology, University of Georgia, Athens

Project Goals: Demonstrate a new direction for synthetic biology — cell free synthetic pathway biotransformation (SyPaB) Produce 12 moles of hydrogen from low-cost cellulosic materials for the first time Solve several challenges for cell-free SyPaB Propose a novel pathway for CO_2 fixation by using electricty

Cell-free synthetic pathway biotransformations (SyPaB), a new direction of synthetic biology, is to implement complicated biochemical reaction network by assembling a number of purified enzymes and coenzymes.[3] As compared to microbial fermentation, SyPaB has several advantages: (i) high product yields without synthesis of cell mass or formation of by-products; (ii) great engineering flexibility (i.e., easy assembly and control when the building blocks are available); (iii) high product titer; (iv) fast reaction rate; and (v) broad reaction condition. But SyPaB suffers from several obvious weaknesses, such as no ability for self-duplication, costly enzymes, enzyme deactivation, costly and labile coenzymes, and so on.

We have designed the non-natural pathways that completely oxidize starch or cellodextrins by using water as an oxidant for generation of **12 mol of hydrogen per mol of glucose equivalent**[4,5] (Fig. 1). These catabolic pathways comprise substrate phosphorylation mediated by phosphorylases, pentose phosphate pathway, and hydrogenesis mediated by hydrogenase. Also, we have increased overall hydrogen production rate by 10 fold through pathway optimization, high temperature and high substrate concentrations.[4] The above reaction is a unique entropy-driven chemical reaction, i.e., low-temperature heat energy is absorbed and is converted to chemical energy – hydrogen that we can utilize for the first time. Also, we propose to use renewable carbohydrate as a hydrogen carrier to solve hydrogen storage challenge (Fig. 2).[6]

The opinion that SyPaB is too costly for producing low-value biocommodities are mainly attributed to the lack of stable standardized building blocks (e.g., enzymes or their complexes), costly labile co-enzymes, and replenishment of enzymes and co-enzymes. The economical analyses clearly suggest that developments in stable enzymes or their complexes as standardized parts, efficient coenzyme recycling, and use of low-cost and more stable biomimetic coenzyme analogues, would result in much lower production costs than do microbial fermentations because the stabilized enzymes have more than three orders of magnitude higher weight-based total turn-over numbers than microbial biocatalysts, although extra costs for enzyme purification and stabilization are spent[3] (Fig.3).

Fig. 4 clearly suggest that the ultimate hydrogen production costs would be as low as $1.50 per kg, where carbohydrate accounts for 80% of the final product price[2]. Developing thermostable enzymes with TTN_W of > 100,000 are easily reached based on our experiences.[7,8] For example, we have obtained three recombinant thermophilic building blocks — #2 *Clostridium thermocellum* phosphoglucomutase[7], #4 *Thermotoga maritima* 6-phosphogluconate dehydrogenase,[8] and #11 *T. maritima* fructose bisphosphatase expressed in *E. coli*, all of which have TTN_W of > 200,000 at ~60 °C. The above results suggest that discovery and utilization of highly stable thermophilic enzymes from extremophiles that have known genomic sequences are highly operative. Recycling NAD with TTN of 1,000,000 has been reported and the

use of less costly biomimetic NAD will be more economically promising, as shown in Fig. 4.[6,9]

In addition to high-yield generation of hydrogen from biomass sugars, we have designed a novel artificial photosynthesis pathway that can utilize electricity to fix CO_2 for producing ethanol and amylose (Fig. 5). When this process is implemented on large scales, it would solve several challenges for sustainability, such as CO_2 fixation, electricity storage, food production, transportation fuel production, water conservation or maintaining an ecosystem for space travel (the concept paper is under review for publication).

References

1. Y.-H.P. Zhang, Microbe 4 (2009) In press.
2. Y.-H.P. Zhang, J.-B. Sun, A.-P. Zeng, J.-J. Zhong, Curr .Opin. Microbiol. (2010) Invited/In preparation.
3. Y.-H.P. Zhang, Biotechnol. Bioeng. Accepted (2010) http://dx.doi.org/10.1002/bit.22630.
4. X. Ye, Y. Wang, R.C. Hopkins, M.W.W. Adams, B.R. Evans, J.R. Mielenz, Y.-H.P. Zhang, ChemSusChem 2 (2009) 149-152.
5. Y.-H.P. Zhang, B.R. Evans, J.R. Mielenz, R.C. Hopkins, M.W.W. Adams, PLoS One 2 (2007) e456.
6. Y.-H.P. Zhang, Energy Environ. Sci. 2 (2009) 272-282.
7. Y. Wang, Y.-H.P. Zhang, J. Appl. Microbiol. 108 (2010) 39-46.
8. Y. Wang, Y.-H.P. Zhang, Microb. Cell Fact. 8 (2009) 30.
9. J.D. Ryan, R.H. Fish, D.S. Clark, ChemBioChem 9 (2008) 2579-2582.

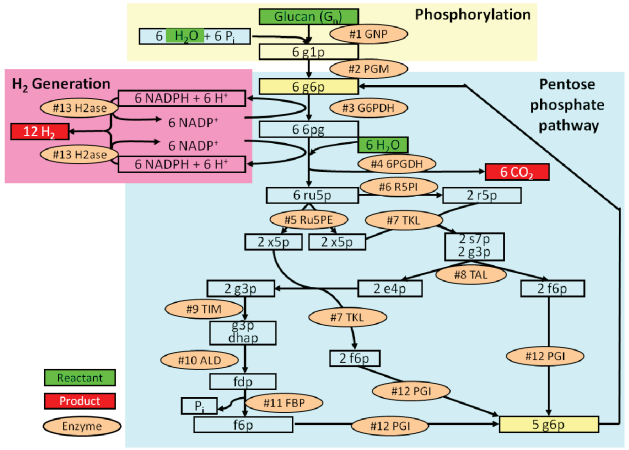


Fig. 1. The non-natural SyPaBs for high-yield hydrogen generation from starch or cellulosic materials [4, 5].

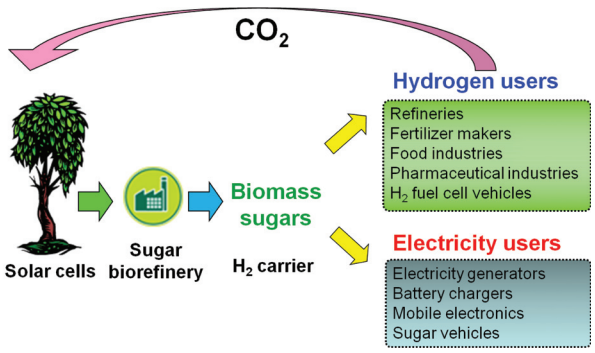


Fig. 2. Scheme of the hydrogen economy based on renewable carbohydrates [1].

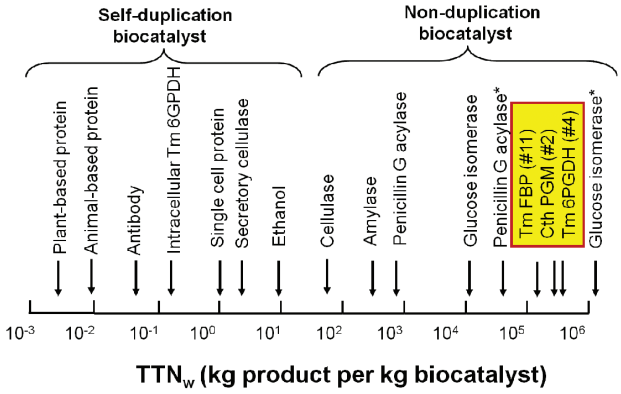
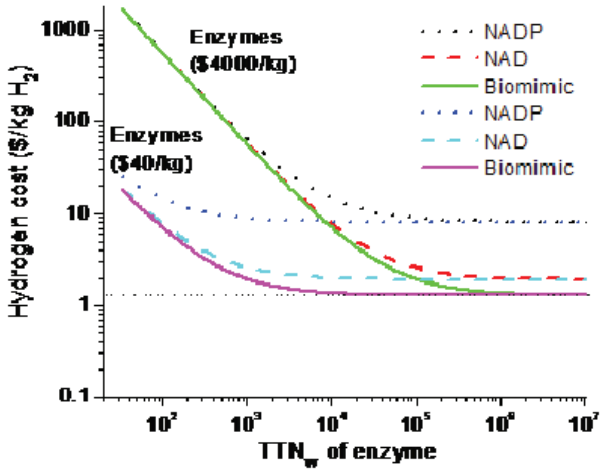


Fig. 3. Comparison of weight-based total turn-over number (TTNW) of self-duplication living biocatalysts and non-duplication enzymes [1, 3].



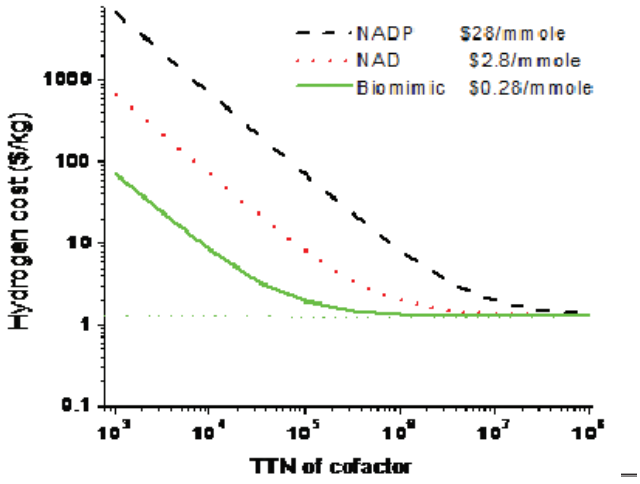


Fig. 4. The estimated hydrogen production costs in terms of turnover number of enzymes and co-enzymes. Carbohydrate prices are $0.18 per kg [2].

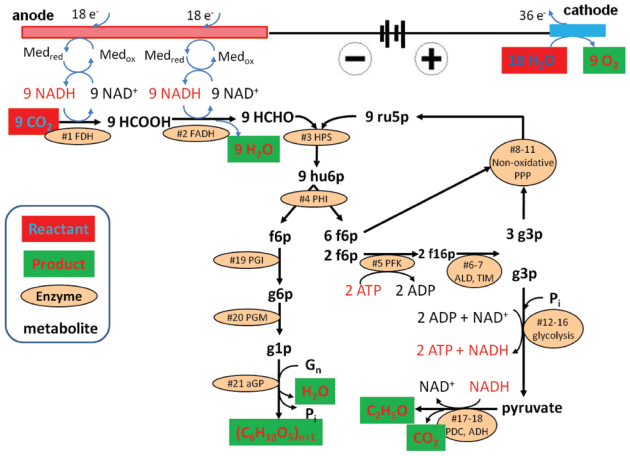


Fig. 5. Artificial photosynthesis that can fix CO_2 by using electricity for producing biofuels and food (under review).

submitted post-press

BioBricks Without Borders: Investigating a Multi-Host BioBrick Vector and Secretion of Cellular Products

Elisabeth Linton* and **Charles D. Miller** (charles.miller@engineering.usu.edu)

Dept. of Biological and Irrigation Engineering, Utah State University, Logan

Project Goals: This research was carried out as part of the 2009 iGEM competition. The overall goal of this project is to use synthetic biological engineering for improved expression and recovery of cellular products. More specific objectives include the conversion of a multi-host vector into a BioBrick compatible format to facilitate expression of BioBricks in variety of organisms and the design BioBrick parts to encourage export of cellular compounds out of the cytoplasm.

The aim of the Utah State University iGEM project is to develop improved upstream and downstream processing strategies for manufacturing cellular products using the standardized BioBrick system. A BioBrick-compatible broad-host vector would facilitate exploitation of advantageous characteristics of various organisms beyond *E. coli*, such as the ability to photosynthetically assimilate carbon. Multi-host vectors were investigated to enable the use of BioBrick constructs in organisms like *Pseudomonas putida*, *Rhodobacter sphaeroides*, and *Synechocystis* PCC6803. For this portion of the project, vector pCPP33 was successfully converted to a BioBrick-compatible format. Following expression, product recovery poses a difficult and expensive challenge. Product purification commonly represents more than half of the total production expense. To counter this problem, secretion-promoting BioBrick devices were constructed through genetic fusion of signal peptides with protein-coding regions. Specifically, phasin protein was targeted for membrane translocation due its binding interaction with polyhydroxyalkanoates (PHAs), which are microbially-accumulated biodegradable plastics. Successful secretion of phasin protein holds potential to lead to an improved recovery mechanism for PHAs. The secretion of green fluorescent protein was studied in parallel due to its ease of detection. A genetic library of more than 50 BioBrick parts has been constructed to carry out this study. Current results indicate that many of these parts are functional and can be used to test production and recovery of cellular products.

Computing for Bioenergy

138

Bayesian Computational Approaches for Gene Regulation Studies of Bioethanol and Biohydrogen Production

Charles E. Lawrence[1] (charles_lawrence@brown.edu), Lee Newberg,[2,3] William Thompson,[1] and Lee Ann McCue[4]*

[1]Center for Computational Molecular Biology and the Division of Applied Mathematics, Brown University, Providence, R.I.; [2]The Wadsworth Center, New York State Dept. of Health, Albany; [3]Dept. of Computer Science, Rensselaer Polytechnic Institute, Troy, N.Y.; [4]Pacific Northwest National Laboratory, Richland, Wash.

http://www.brown.edu/Research/CCMB/

Project Goals: We are extending our phylogenetic Gibbs sampling algorithms to reconstruct the joint posterior space of the ancestral states of regulatory motifs and

developing point estimates and confidence limits for these discrete high-dimensional objects. We are also applying our existing models and technologies, along with the above modifications, to clades of alpha-proteobacterial species, to identify regulatory mechanisms and reconstruct the ancestral states of the regulatory networks for the efficient fermentation of sugars to ethanol and the production of biohydrogen.

Decreasing America's dependence on foreign energy sources and reducing the emission of greenhouse gases through the development of biofuels are important national priorities. These priorities have catalyzed research on cellulosic ethanol as a clean, renewable energy source to replace fossil fuels, and biohydrogen as a carbon-free energy carrier. Turning these biofuels into viable alternative energy sources requires further research into the degradation of cellulose and fermentation of the resulting sugars, and the metabolic and regulatory networks of biohydrogen production. The genomes of many of the microbial species capable of these processes have been sequenced by the GTL and other programs, and many more are expected soon. These sequence data provide a wealth of information to explore nature's solutions for the production of biofuels. In particular, among the over 170 α-proteobacterial species with genome sequence data available are several species with metabolic capabilities of interest, including efficient fermentation of sugars to ethanol and the ability to produce hydrogen. Understanding the regulatory mechanisms and complex interplay of metabolic processes in these species is key to realizing the promise of biofuels. Thus, our research goal is to identify the ensemble of solutions that have been explored by the α-proteobacteria to regulate the metabolic processes key to biofuel production.

The solution space explored by these species spans three scales: molecular (genes/gene products), cellular (genomic), and communities (clades). On the finest scale, the catalytic steps of biofuel production (ethanol or hydrogen) are performed by individual enzymes that are the products of individual genes, each regulated by a set of *cis* and *trans* elements. At the cellular level, the expression of the gene products is often coordinated via a set of *trans* elements (transcription factors) that interact with all or most genes in these pathways to form a regulatory unit called a regulon. Furthermore, the collection of species that encode the genes form clades which have explored a catalytic and regulatory space on an evolutionary time scale. We are developing probabilistic models to represent these multiscale processes, Bayesian statistical inference procedures and computational methods to identify the posterior distributions of these parameters, efficient point estimates of their values, and Bayesian confidence limits for these estimates.

Specifically, we are characterizing the gene/clade interface by extending our phylogenetic Gibbs sampling algorithms to reconstruct the joint posterior space of the ancestral states of regulatory motifs, and developing point estimates and confidence limits for these discrete high-dimensional objects. We are also applying our existing models and technologies, along with the above modifications, to clades of α-proteobacterial species, to identify regulatory mechanisms and reconstruct the ancestral states of the regulatory networks for the efficient fermentation of sugars to ethanol and the production of biohydrogen.

139

Sugar-Salt and Sugar-Salt-Water Complexes: Structure and Dynamics of Glucose – KNO_3 - $(H_2O)_n$

Madeleine Pincu[1,2]* (mpincu@uci.edu), Brina Brauer,[2] **Robert Benny Gerber**,[1,2] and Victoria Buch[2†]

[1]The Fritz Haber Center for Molecular Dynamics, Chemistry Institute, the Hebrew University, Jerusalem, Israel and [2]Dept. of Chemistry, University of California, Irvine
[†]Deceased

Project Goals: To investigate structure, dynamics, and energetics of saccharides and their interactions with ions and biological osmolytes in the presence of microhydration.

Molecular dynamics (MD) simulations are carried out for the complex of glucose with KNO_3 and for complexes of the type glucose – KNO_3 - $(H_2O)_n$, for $n \leq 11$. Structure and dynamic properties of the systems are explored. The MD simulations are carried out using primarily the DLPOLY/OPLS force field, and global and local minimum energy structures of some of the systems are compared with ab initio MP2 calculations. The main findings include: (1) Complexation with KNO_3 leads to an "inverse anomeric effect," with the β-glucose complex more stable than the α-glucose by ~ 1.74 Kcal/mol. (2) As temperature is increased to 600K, the KNO_3 remains undissociated in the 1:1 complex, with the K^+ hooked to the equilibrium site, and the NO_3^- bound to it, undergoing large-amplitude bending/torsional motions. (3) For $n \geq 3$ water molecules added to the system, charge separation into K^+ and NO_3^- ions takes place. (4) For n =11 water molecules all hydroxyl groups are hydrated with the glucose adopting a surface position, indicative of a surfactant property of the sugar. (5) Comparison of DLPOLY with MP2 structure predictions indicates that the empirical force field predicts global and local minimum structures reasonably well, but errs in giving the energy rankings of the different minima. Implications of the results to effects of salts on saccharides are discussed.

submitted post-press

A Multi-Scale Approach to the Simulation of Lignocellulosic Biomass

Goundla Srinivas,[1] Yana Kholod,[2] Loukas Petridis,[1] James W. Evans,[2] Monica Lamm,[2] Theresa L. Windus,[2] Xiaolin Cheng[1,3]* (chengx@ornl.gov), **Mark S. Gordon**,[2] and **Jeremy C. Smith**[1,3]

[1]Oak Ridge National Laboratory, Oak Ridge, Tenn.; [2]Ames Laboratory, Iowa State University, Ames; and [3]University of Tennessee, Knoxville

Project Goals: In concert with the imminent increase in the Department of Energy's (DOE) leadership supercomputing power to the petaflop range, the objective of this project is to develop multiscale methods for extending the time- and length-scales accessible to biomolecular simulation on massively parallel supercomputers. This project also aims to apply the developed multiscale approaches to obtain an understanding of the structure, dynamics and degradation pathways of extended cellulosic and lignocellulosic materials. Information from multiscale simulation, when closely integrated with experiment, will provide fundamental understanding needed to overcome biomass recalcitrance to hydrolysis.

The research involves the development of multiscale simulation methods and their application to solve critical problems needed for understanding the bottleneck in cellulosic ethanol production: the recalcitrance to hydrolysis of lignocellulosic biomass. The multiscale methodologies span from accurate quantum-chemical techniques, needed to understand critical local interactions in biomass, to atomistic and coarse-grained simulations, needed to approach systems-level phenomena. The codes developed will be parallelized for efficient use on petascale supercomputers. The physical simulation models of lignocellulosic biomass derived using the multiscale approaches will serve as a basis for interpreting an array of biophysical experiments, and when closely integrated with experiments, will eventually lead to a description of the physicochemical mechanisms of biomass recalcitrance to hydrolysis, and thus aiding in developing a strategy to overcome the recalcitrance. Our progress presented here mainly focus on: adapting the quantum mechanical fragment molecular orbital (FMO) method to the study of cellulosic/lignocellulosic biomass, developing an adaptive fast multipole based Poisson Boltzmann (PB) electrostatic solvation model for nano-scale biomolecular systems and self-consistently coarse-graining of cellulose/lignocellulose force fields.

This research is funded by the Genomic Science Research Program, Office of Biological and Environmental Research (BER), and the Scientific Discovery through Advanced Computing (SciDAC) program, U. S. Department of Energy (DOE), currently under FWP ERKJE84.

submitted post-press

Processivity of Cellobiohydrolase, Cellulose Structure, and Advanced Methods for Petascale Molecular Dynamics

James F. Matthews,[1] Xiaolin Cheng,[2] Malin Bergenstrahle,[3] Eric H. Knoll,[1] Mark Williamson,[4] Edward C. Uberbacher,[2] John W. Brady,[3] Charles L. Brooks III,[5] Linghao Zhong,[6] Mark R. Nimlos,[1] Michael E. Himmel,[1] and **Michael F. Crowley**[1]* (Michael.Crowley@nrel.gov)

[1]National Renewable Energy Laboratory, Golden, Colo.; [2]Oak Ridge National Laboratory, Oak Ridge, Tenn.; [3]Cornell University, Ithaca, N.Y.; [4]San Diego Supercomputer Center, San Diego, Calif.; [5]University of Michigan, Ann Arbor; and [6]Pennsylvania State University, Mont Alto

Project Goals: The SciDAC project is focused on molecular modeling of molecular and macromolecular systems that are key to the understanding and designing of cellulose degrading technologies acting on biomass. The project is enabling scientific investigation of cellulose structure and properties and enzymatic mechanism (processivity) for degrading cellulose to sugars as a raw material for biofuels. Speed, new sampling methods, and simulation size (scaling) are the aims of code development and performance work with the additional creation of validation suites for code integrity. The main thrust is to enable the specific needs of modeling to produce the methods necessary for thermodynamic, kinetic, and mechanistic insight into cellulose conversion processes at a speed which will make this research possible, mainly through the use of high-performance computers at the petascale level and advanced sampling methods.

The mechanism by which Cel7A cellobiohydrolase degrades cellulose is not currently known nor is the structure and molecular properties of cellulose microfibrils as found in biomass. The degrading of cellulosic biomass to sugars is primarily accomplished both in nature and in biorefineries by enzymes such as Cel7A; the understanding of both the structure of cellulose and of the mechanism of cellulose-degrading enzymes is critical to improving the technology of producing biofuels from biomass. There have been several barriers to studying these molecular systems with modeling, primarily the lack of reliable force fields for cellulose and the lack of highly scalable molecular dynamics programs that can treat these systems and have the particular thermodynamic sampling abilities and force field features to answer the questions which are unique to this difficult problem. Our progress presented here is fivefold: cellulose force field evaluation and characterization, molecular dynamics code validation, enabling highly parallel programs to utilize cellulose force fields, solving parallel bottlenecks in existing codes, and designing highly parallel implementations of sampling methods for studying enzymatic mechanisms and cellulose morphology and decrystalization.

The SciDAC project "Understanding Processivity of Cellobiohydrolases" is supported by the Office of Biological and Environmental Research and OASCR in the DOE Office of Science.

Small Business Innovation Research (SBIR)

140

Genetic Analysis of Cellulose Degradation by *Clostridium phytofermentans*

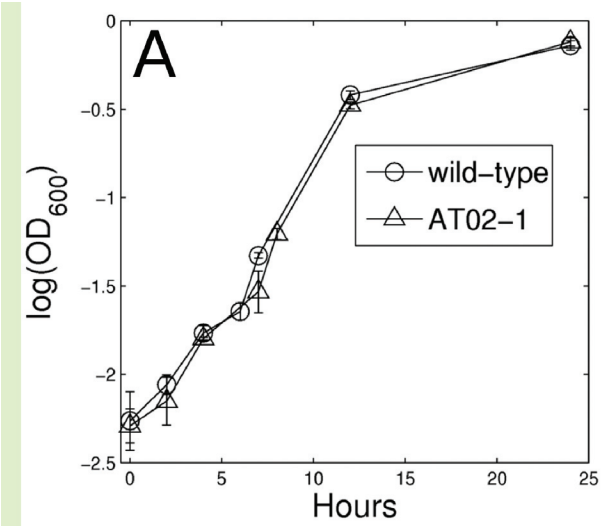
Andrew C. Tolonen[1]* (tolonen@genetics.med.harvard.edu), Amanda C. Chilaka,[1,2] and **George M. Church**[1]

[1]Dept. of Genetics, Harvard Medical School, Boston, Mass. and [2]Dept. of Biology, Northeastern University, Boston, Mass.

Project Goals: See below.

Microbial cellulose degradation is a central part of the global carbon cycle and has potential for the development of inexpensive, carbon neutral biofuels from non-food crops. The major roadblock to the use of cellulosic biomass as a biofuel feedstock is the recalcitrance of cellulosic fibers to breakdown into sugars. *Clostridium phytofermentans* grows on both of the two main components of plant biomass, cellulose and hemicellulose, by secreting enzymes to cleave these polysaccharides and then fermenting the resulting hexose and pentose sugars to ethanol. In order to breakdown cellulose biomass, *C. phytofermentans* has a repertoire of 161 carbohydrate-active enzymes (CAZy), which include 108 glycoside hydrolases spread across 39 families.

Broadly, our goal to understand the genetic mechanisms that permit to *C. phytofermentans* to efficiently convert cellulosic biomass to ethanol. To enable targeted gene inactivation in *C. phytofermentans,* we show that interspecific conjugation with *E. coli* can be used to transfer a plasmid into *C. phytofermentans* that has a resistance marker, an origin of replication that can be selectively lost, and a designed group II intron for efficient, targeted chromosomal insertions without selection. We applied these methods to inactivate Cphy3367, a ß-1,4-glucanase in glycoside hydrolase family 9 (GH9). Cellulolytic *Clostridia* usually have numerous genes for GH9 proteins: the *C. thermocellum* ATCC 27405 genome has 16 GH9 genes, *C. cellulolyticum* H10 has 13 GH9 genes, and *C. cellulovorans* has 5 GH9 genes. In contrast, *C. phytofermentans* has only a single GH9-encoding gene, *cphy3367*. The *C. phytofermentans* strain with an intron insertion in *cphy3367* (strain AT02-1) grows normally on some carbon sources such as glucose, cellobiose, and hemicelluose, but has lost the ability to degrade cellulose (Fig 1). Although *C. phytofermentans* up-regulates the expression of numerous enzymes to breakdown cellulose, this process thus relies upon a single, key hydrolase, Cphy3367. Generally, these results show that targeted gene inactivation can be used to identify key enzymes for the breakdown of biomass by *C. phytofermentans.* Future genetic studies of in *C. phytofermentans* will untangle the roles of additional hydrolases in cellulose degradation.



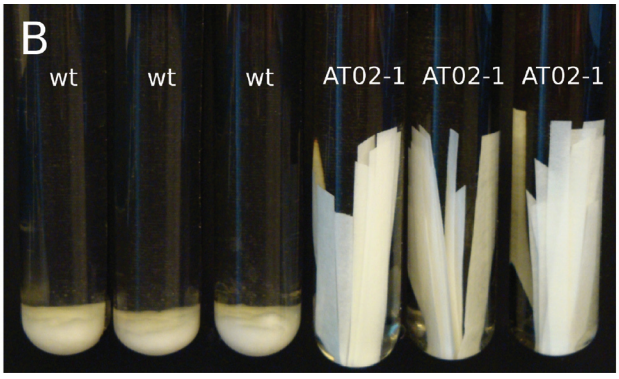


FIG 1 *C. phytofermentans* strain with disruption of cphy3367 (AT02-1) had similar growth rates as wild-type on glucose A, but had lost the ability to degrade filter paper cellulose B. Growth curves are means of triplicate cultures. Error bars show one standard deviation and are smaller than the symbols where not apparent.

141

Microbioreactor Technology for Obligate Anaerobes

Harry Lee (harrylee@pharyx.com) and Paolo Boccazzi* (boccazzi@pharyx.com)

Pharyx, Inc., Boston, Mass.

http://www.pharyx.com

Project goals are 1) determine microbioreactor designs that will support anaerobic inoculation and fermentation, 2) identify optimal materials for fabricating anaerobic bioreactors, 3) determine the range of process parameters where microbioreactor data corresponds to serum tubes and stirred tank fermentors, 4) monitor enzyme activity on-line.

Anaerobic microorganisms have evolved biochemical pathways that can be exploited for industrial applications. These include the ability to breakdown environmental pollutants for bioremediation, the breakdown of cellulose into simple sugars for biofuels, and the production of specialty chemicals. However, there remains a tremendous challenge to the scale-up of bioenzymatic activities to industrial processes. While systems biology approaches and metabolic engineering promise to contribute to our understanding of these systems, a key bottleneck is in conducting controlled experiments to ground these approaches with high quality data. Thus far, experiments are frustrated by the laborious set-up and operation of stirred tank bioreactor systems, which for anaerobic microbiology is further encumbered by the requirement of an anaerobic environment. The absence of easy to use systems also holds back more traditional microbiology approaches such as mutagenesis and screening and directed evolution.

We are developing a parallel bioreactor system, based on microfluidic integration technology and disposable microbioreactor modules, with application specific customizations for anaerobic fermentation. These customizations are aimed to enable up to 32 simultaneous anaerobic fermentations under controlled conditions, with online monitoring of growth kinetics and other phenotypes such as enzyme activity. A unique feature of this system is the ability to operate it in ambient air through careful inoculation port and reactor and control module design, or to operate it within an anaerobic bag, taking advantage of its compact size.

Preliminary anerobic fermentations in microbioreactor devices are shown in Figure 1. Cell growth in the three bioreactor chambers compared to the uninoculated medium control is clear, and exponential growth can be seen for t<5h, indicating obligate anaerobes can be cultured in plastic bioreactors fabricated using our microfluidic integration platform. However, for the *Butyrivibrio fibrisolvens* and *Clostridium acetobutylicum* fermentations, optical density measurements were confounded by the gas bubbles generated by the microorganisms. This was confirmed visually during the fermentation and evidenced in the photographs, shown in Figure 2, of the devices at the end of the fermentation.

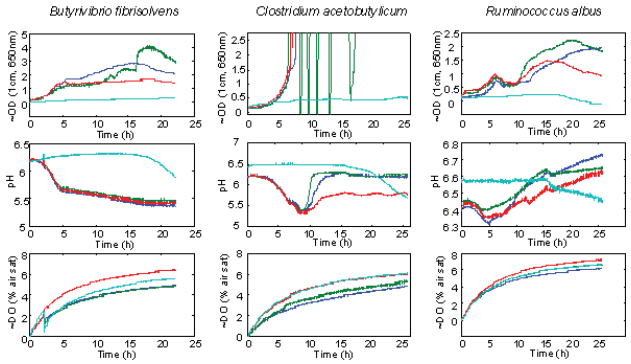


Figure 1. Online sensor measurements of dissolved oxygen (DO), pH and optical density (OD) for anaerobic fermentations of three obligate anaerobes. Growth studies were conducted in our anaerobic bioreactor modules which contained 4 growth chambers each. Light blue lines represent the uninoculated controls, other 3 lines represent 3 independent reactor replicates in each cassette. All strains were grown in modified peptone yeast extract glucose (MPYG) medium (Atlas RM. 2004. Handbook of microbiological media. CRC, Boca Raton). Measurements were taken every 60 seconds.

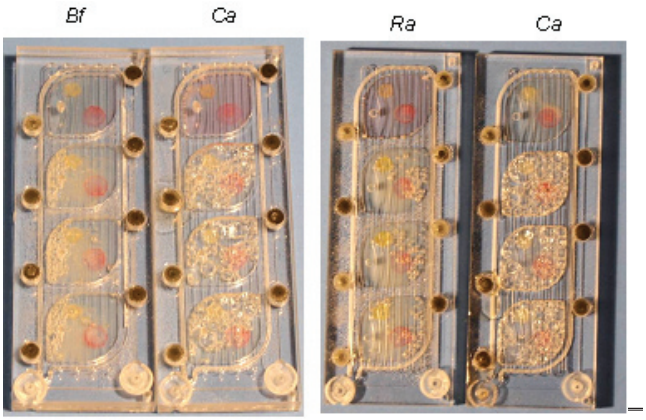


Figure 2. Photographs of integrated bioreactor devices. Pink circles are dissolved oxygen sensors, yellow circles are pH sensors, black plugs are butyl rubber injection ports. A stream of nitrogen flowed through the serpentine channels that surround the cvisible in the chamber are biologically generated gas. Bf (*Butyrivibrio fibrisolvens*), Ca (*Clostridium acetobutylicum*), Ra (*Ruminococcus albus*). First growth reactor from top of each cassette contained uninoculated medium.

142

A Genome-Wide Perspective on the Regulation of Plant Carbohydrate Conversion to Biofuels in *Clostridium phytofermentans*

Elsa Petit,[1] Tom Warnick,[1] Greg Latouf,[1,6] James Hayes,[2,5] Amy Biddle,[1]* Zhiyi Sun,[1] Danny Schnell,[2,4,5] **Sue Leschine**,[1,4,5] and **Jeff Blanchard**[1,4,5] (blanchard@microbio.umass.edu)

[1]Dept. of Microbiology, [2]Dept. of Biochemistry, [3]Graduate Program in Molecular and Cellular Biology, [4]The Massachusetts Institute for Biofuels Research, and [5]Institute for Cellular Engineering, University of Massachusetts, Amherst and [6]Qteros Inc.

Project Goals: Derive genetic models of the conversion of cellulosic biomass to alcohols in *C. phytofermentans* by microarray analysis of the fermentation of simple sugars contained in feedstocks relevant to the agricultural and forestry industries.

The economic costs of degrading cellulosic biomass currently hold back widespread use of plant biomass from agricultural and forestry wastes as a domestic renewable alternative to gasoline. *Clostridium phytofermentans* ferments all major components of the plant cell wall to ethanol and is an emerging model organism for understanding the direct conversion of plant biomass to fuels. The *C. phytofermentans* genome contains an abundance of carbohydrate degradation enzymes, carbohydrate transporters and transcriptional regulators. In order to build a foundation for engineering improvements on industrially relevant feedstocks, microarray experiments were carried out on a number of purified plant cell wall carbohydrates. The results demonstrate that *C. phytofermentans* regulates the stoichiometry of the plant degradative and assimilatory machinery in response to substrate availability. Because of the modularity of the sugar degradation and assimilation systems and the lack of a cellulosome, further improvement in biomass degradation through genetic manipulation is relatively straightforward.

* Presenting author

Systems Biology for DOE Environmental Missions: Systems Environmental Microbiology

143

Metagenomic and Metaproteomic Analysis of Viruses from an Acid Mine Drainage System

Kimberly Pause Tucker[1]* (kpause@marine.usf.edu), Christine L. Sun,[2] Brian D. Dill,[3] Paul Wilmes,[2] Robert Hettich,[3] Brian C. Thomas,[2] Vincent Denef,[2] Nathan VerBerkmoes,[3] Luis Comolli,[4] Birgit Luef,[4] **Jillian F. Banfield**,[2] and **Mya Breitbart**[1]

[1]University of South Florida, Saint Petersburg; [2]University of California, Berkeley; [3]Oak Ridge National Laboratory, Oak Ridge, Tenn.; and [4]Lawrence Berkeley National Laboratory, Berkeley, Calif.

Project Goals (Abstracts 143-145): The aim of this project is to develop integrated genomic and proteomic (proteogenomic) methods to study virus-host dynamics in bioreactor-grown and natural microbial biofilm communities. We will utilize simultaneous metagenomic analysis of CRISPRs (the adaptive microbial immune system) and viruses to examine the co-evolution and dynamics of viruses and their microbial hosts. Due to virus-host specificity, viruses can change the microbial community structure by enriching for resistant strains. Proteomics will be used to examine metabolic interplay via monitoring of protein production before, during, and after viral infection, while genomic analysis of CRISPRs will be used to correlate these findings with changes in community structure. Because viral predation can undermine microbially-based technologies such as microbial fuel cells, bioethanol production, and environmental remediation, methods to study microbe-viral interactions are widely relevant to DOE missions.

Metagenomics has proven to be a useful tool for examining viruses in a variety of natural systems, revealing novel and diverse environmental viral communities. In systems with low species richness, metagenomic sequencing can reveal patterns in microbial diversity and evolution with unprecedented resolution. Concomitant genomic sampling of virus populations and clustered regularly interspaced short palindromic repeat (CRISPR) loci from the low diversity acid mine drainage (AMD) biofilm system from Richmond Mine (Iron Mountain, CA, USA) offers unique insight into virus-host interactions (see Sun *et al.* poster). Five viral genomes (AMDV1-5) were previously reconstructed from metagenomic sequence data from the UBA site, and linked to their potential hosts through the CRISPR loci. Here we report analysis of an additional novel archaeal virus genome, AMDV6, which was reconstructed from a total microbial community metagenome from the C75 site. It appears that there was a bloom of viruses occurring at that time, since many of the total metagenomic sequences were viral in origin (see Sun *et al.* poster). Annotation of the AMDV6 genome revealed open reading frames (ORFs) with functions related to DNA-binding, replication, and modification. Other regions of interest include a phage integrase, as well as small molecule binding domains. The AMDV6 genome also exhibits similarity to some of the other AMD viruses. The largest AMDV6 ORF (1138 aa) shows approximately 30% amino acid identity to hypothetical membrane-associated proteins of AMDV4 and AMDV3. The second largest AMDV6 ORF (672 aa) shares 31% amino acid identity with the DNA polymerase elongation subunit (family B) of AMDV4. The spacers of CRISPR loci from the Archaea concurrently inhabiting the C75 site are being examined to identify potential hosts for the novel AMDV6 virus.

Purification of viral particles from the biofilm prior to metagenomic sequencing allows for the description of the complete viral community, including viruses present at low abundance that might not have been sampled in the total community metagenome, and viruses that are not targeted by CRISPRs. Using a combination of mechanical disruption, filtration, and density-dependent centrifugation, we have developed a protocol to purify viruses from the AMD biofilm for genomic and proteomic analyses. Metagenomic sequencing of viruses purified from the UBA-BS site demonstrated that a high proportion of the sequences had similarity to known AMD viral sequences, confirming the success of the virus purification method. Amongst the sequences with similarity to previously described AMD viruses, many were similar to AMDV4, a virus that infects E-plasma. The majority of the AMDV4 sequences were more than 98% identical on the nucleotide level to sequences in the database, which was generated three years prior. This level of sequence conservation indicates that some of the viral populations in the AMD system are not changing considerably over time despite high virus-host contact rates in the biofilm and an active CRISPR-mediated defense system. The high proportion of virus sequences with similarity to the five previously described AMD viruses suggests that there are only a few dominant viral types in the system, which is consistent with the low microbial species richness at this site. Although the majority of the viral sequences were similar to previously described AMD viruses, a number of sequences were identified that had no similarities to either the NCBI non-redundant database or to the AMD-specific metagenomic database, demonstrating the discovery of novel viral types from metagenomic sequencing of the purified viral fraction. Since these viruses were not previously identified from the whole community metagenomes, it is possible that they are not targeted by CRISPRs. Continued sequencing of the purified viral

fractions will allow us to estimate the percent of the AMD viral community that is targeted by CRISPRs. In addition, reconstruction of the complete genomes of these novel viruses will enable examination of differences in the rates and mechanisms of genome evolution between CRISPR-targeted viruses and those that are not affected by CRISPRs.

Metaproteomic analysis on AMD viral fractions purified from the UBA-BS site demonstrated a 12-fold enrichment of identified viral proteins compared to total community proteomes. Some non-viral proteins were also enriched, including flagellar proteins, which is likely a consequence of their co-purification during the virus particle selection protocol. The majority of identified viral proteins were attributed to AMDV1, a bacteriophage of *Leptospirillum* groups II and III. The amount of variability and rapid evolution of viruses makes identification of viral proteins challenging, prompting the need for *de novo* predictions (see VerBerkmoes *et al*. poster). Inclusion of six-frame translations for the AMDV1 genome lead to a ten-fold increase in protein identification for this virus, demonstrating the need for database expansion to include sequence variants (frameshifts and single nucleotide polymorphisms). Concurrent metagenomic and metaproteomic analyses of purified AMD viral communities will significantly increase protein identification rates. To assess the amount of sequence variability amongst individuals in the viral community and aid in protein identification, the sequence variation in AMDV1 over space and time was examined by PCR and sequencing of gene 1. The AMDV1 population exhibited extensive single nucleotide polymorphisms in this gene, which are likely generated in response to the CRISPR defense. Sequences from the UBA-BS site clustered separately from sequences from the B-drift site, indicating spatial differentiation in the AMDV1.

In conclusion, purification of viral communities from the AMD biofilm has expanded our understanding of the diversity and evolution of viruses in this extreme environment. Metagenomic and metaproteomic analyses of the purified viral communities overcome the difficulties of obtaining sufficient coverage of viral genes/proteins from total biofilm samples. The high levels of single nucleotide polymorphisms and rapid evolution of viruses in the AMD system present a unique challenge for protein identification, which will be a focus of future work (see VerBerkmoes *et al*. poster).

Funding provided by DOE Genomics:GTL Program grant number DE-FG02- 07ER64505

144

Community Proteogenomic Analysis of Virus-Host Interactions in a Natural System

Christine Sun[1]* (christine_sun@berkeley.edu), Brian Thomas,[1] Vincent Denef,[1] Kimberly Pause Tucker,[2] Paul Wilmes,[1] Rodolphe Barrangou,[4] Chongle Pan,[3] Michael P. Thelen,[5] Robert Hettich,[3] Nathan VerBerkmoes,[3] Mya Breitbart,[2] and **Jill Banfield**[1]

[1]University of California, Berkeley; [2]University of South Florida, Saint Petersburg; [3]Oak Ridge National Laboratory, Oak Ridge, Tenn.; [4]Danisco USA Inc., Madison, Wis.; and [5]Lawrence Livermore National Laboratory, Livermore, Calif.

Project Goals: See goals for abstract 143.

Bacteriophages and archaeal viruses (collectively referred to here as viruses) play critical roles in microbial evolution as they can shape the composition and functionality of microbial communities via predation of their hosts and promotion of lateral gene transfer. Few studies have examined virus-host dynamics and viral diversity in natural populations due to difficulties in the recovery of viral population genomic data and the lack of cultivation-independent methods to link viruses to their hosts. However, the recent identification of CRISPRs (clustered regularly interspaced short palindromic repeats) as adaptive, rapidly evolving microbial immune systems provides a crucial cultivation-independent connection between a CRISPR-containing host and its viruses. CRISPR regions consist of repeats separated by short spacer sequences that derive from the genomes of viruses and other mobile elements. Since spacers are incorporated into the CRISPR in a unidirectional manner, the CRISPR region also serves as a historical timeline of virus exposure. We have simultaneously analyzed virus populations and CRISPR loci of the bacterial and archaeal populations they target to study virus-host interaction dynamics in relative low diversity acid mine drainage (AMD) microbial communities. To complement comprehensive examination of the dynamics of the CRISPR loci and viruses in natural populations, we initiated research on an *in vitro* system involving *Streptococcus thermophilus* and its lytic phage. The objective of the *in vitro* simple system research is to calculate the rates and determine the mechanisms of virus and host co-evolution.

We have shown that CRISPR loci in microbial genomes assembled from AMD community genomic datasets are extremely dynamic genomic regions that undergo rapid gain and loss of spacers. Via 454 FLX pyrosequencing of amplified CRISPR loci from two *Leptospirillum* Group II bacterial populations (UBA and 5way type), 419,351 total spacers were recovered. As 454 pyrosequencing generates reads with higher error rates relative to Sanger reads, spacer sequences were placed in groups that share at least 80% identity over 80% length, resulting in 2,841 and 649 spacer groups in the 5way and UBA type *Leptospirillum* Group II, respectively. The rarefaction curves for both samples using collapsed spacer groups demonstrate no approach to

saturation, implying a large diversity of CRISPR spacers in each population. Spacers were used to identify sequences from a viral population that targets *Leptospirillum* Group II (AMDV1). By mapping spacer sequences back onto assembled viral genomes, the "CTT" CRISPR PAM (proto-spacer adjacent motif involved in spacer selection) was identified. The presence of the motif is the only factor that appears to impact selection of the spacer sequence (i.e., there is no significant bias toward the coding vs. non-coding strand). We have identified many cases where multiple spacer sequence variants were sampled from the same locus in the diverse AMDV1 population. Spacers also target the *Leptospirillum* Group II genome, mostly in phage/plasmid-like regions. Interestingly, more 5way unique spacers target the UBA type *Leptospirillum* Group II than the 5way type. This may further support the previously suggested recent lateral transfer of the CRISPR/Cas system from the UBA to the 5way genome. Interestingly, spacers unique to 5way target transposons only located in the UBA *Leptospirillum* Group II type but not to the 5way type.

We compared the *Leptospirillum* Group II locus across seven time points from three locations and spanning five years. Notably, most spacers in the first fifth of loci are conserved in both the UBA and 5way *Leptospirillum* Group II genotypes across all samples, indicating that spacer loss occurs primarily in the middle of loci. In some cases, dissimilarities in spacer context indicate multiple independent sampling events at the same PAM. Five samples from one location did not show unidirectional locus expansion along the 2-year time series, suggesting that the predominant process is selection from among coexisting subpopulations.

At one time point, sequences recovered from one biofilm included a large representation of reads from virus populations. Dominant among these is a population (AMDV3b) related to virus previously described by metagenomics, AMDV3. Other populations present at lower levels include the bacteriophage AMDV1, viruses related to AMDV4, and new archaeal viral populations including AMDV6 and another variant of AMDV3 (AMDV3c). The AMDV3b population was very deeply sampled (>800 times coverage), providing a novel view of natural virus population structure. Levels of sequence heterogeneity and differences in gene content in the AMDV3b population are highly variable across the genome. Extensive heterogeneity is localized in a very large protein (~ 2,000 amino acids) of unknown function that we suspect is a tape measure protein involved in tail or capsid assembly, and thus host specificity. We detect extensive nucleotide polymorphism and small groups of nucleotide insertions/deletions in the gene for this protein, most of which leave the flanking amino acid sequence in frame. There is clear evidence for extensive, fine-scale homologous recombination (tens to hundreds of nucleotide blocks) amongst sequence variants and for construction of protein variants by mix and match involving large, distinct sequence modules.

On a shorter time scale using the *in vitro* system of *Streptococcus thermophilus* and its virus, CRISPR loci and viral sequences were examined at two time points separated by around 42 to 51 generations. No spacers in the original CRISPR loci were able to target the virus. However, at the second time point, new spacers were discovered in two of four CRISPR loci in *Streptococcus thermophilus*. At least 50 of these new spacers match the original viral population perfectly. In at least four instances, the virus population had mutated by single nucleotide polymorphisms within the proto-spacer or the PAM in order to escape the new spacers.

All evidence from both systems indicate that both CRISPRs and viruses are co-evolving rapidly, in an "arms-race" that requires continual acquisition of new spacers to target viral sequences that have been modified by mutation and sequence shuffling. An understanding of virus-host interaction dynamics is generally applicable and directly relevant to the DOE mission through the implications for the maintenance of stable biotechnologies, including bioremediation and industrial bioenergy production.

Funding provided by DOE Genomics:GTL Program grant number DE-FG02-07ER64505

145

Advanced Proteomics Methods to Identify Low Level Viral Signatures in Isolate and Environmental Samples

Nathan VerBerkmoes[1]* (verberkmoesn@ornl.gov), Brian Dill,[1] Jacque Young,[1] Kimberly Pause Tucker,[2] Christine Sun,[3] Brian Thomas,[3] Vincent Denef,[3] Paul Wilmes,[3] Patricia Carey,[1] Manesh Shah,[1] Chongle Pan,[1] Robert Hettich,[1] Rodolphe Barrangou,[4] Mya Breitbart,[2] and Jill Banfield[3]

[1]Oak Ridge National Laboratory, Oak Ridge, Tenn.; [2]University of South Florida, Saint Petersburg; [3]University of California, Berkeley; [4]Danisco USA Inc., Madison, Wis.

Project Goals: See goals for abstract 143.

Genomic and proteomic methods have been developed to interrogate ecological interplay among bacterial and archaeal members of natural microbial communities. The relative simplicity of the acid mine drainage (AMD) biofilm system has made it a prime model for community genomic and proteomic approaches to investigate the ecology of a natural microbial community. While much progress has been made to begin to unravel the ecological and genetic interaction of bacterial and archaeal species in the AMD system, less has been discovered concerning the activity and ramifications of viruses in the AMD biofilms. Interestingly, most AMD bacteria and archaea have active CRISPRs, a microbial system which can convey immunity against viral attack. Viruses (bacteriophage and archaeal viruses) are ubiquitous in microbial communities, and have dramatic effects on species-level and community-level dynamics. By controlling fluctuation of microbial populations, initiating genetic exchange between populations, and directly altering

microbial physiology, viruses can affect carbon, nutrient, and metal cycling.

In situ genomic and proteomic analyses of viruses pose a daunting challenge due to the difficulty of obtaining sufficient sequence coverage given the low abundance of viruses relative to the bacterial and archaeal constituents and the high mutational rate exhibited by viruses. Our aim is to adapt community genomic and proteomic approaches to identify proteins from viruses in order to expand our understanding of physiological, ecological, and genetic viral-microbial interactions. To reach this aim we are developing mass spectrometry based proteomics techniques to confidently identify viral proteins in complex matrices. To develop these techniques, we are using two model systems: the *in vitro* biotechnology system of *Streptococcus thermophilus* and its lytic phage as well as the AMD biofilm community. The first system is well characterized genomically and provides an ideal model system to develop and test proteomic methodology because of the high concentrations of virus particle and host cells. However, although the minimal media is relatively straightforward, the milk proteins create a significant masking issue. The AMD biofilms are the best characterized microbial community via proteogenomic techniques to date, with the presence of viruses confirmed via genomic approaches and imaging). The challenge associated with detection of viral proteins in this system is their low natural abundance and rapidly evolving genomes and thus proteomes.

In the study of the model *Streptococcus thermophilus* and its lytic phage, our first focus is a time series of infection followed by viral enrichment and proteome analyses. Our goal is to fully determine the proteins present in the free-living virus versus proteins specifically expressed in the host during infection. We also hope to determine the host response to the viral onslaught. *S. thermophilus* TSH-466 grown in LM17 media was infected with D2972 phage (M.O.I=.01). At times 0, 30 minutes, 1 hour, 2 hours, 4 hours, and 24 hours post-infection, aliquots were taken from the culture and centrifuged (10,000 g) for the cellular fraction, then the supernatant enriched for virus via PEG precipitation. Cellular fractions were measured by 2d-LC-MS/MS and searched against the host and viral genomes. We have seen a clear enrichment of the intact virus in the later time points with deep coverage over all known structural proteins and the detection of many unknown viral proteins as well. For the whole cell fraction at early time points we identify ~50% of the host proteome with a major increase in viral proteins over time. We are currently analyzing the datasets to understand the host response to viral attack. With the model system we are also developing chemical and antibody depletion technologies to remove milk proteins effectively allowing us to monitor the microbe-viral interactions in milk, the biotechnology media of interest.

In the course of the AMD project we have characterized over 50 unique AMD community samples. We have built a new predicted proteome database which contained all of the predicted viral proteins to date as well as their variants, translated in all 6 open reading frames. We appended this database to the current AMD community predicted proteome database and queried our archived datasets. The new search identified significantly more viral proteins than the previous database. From this informatics exercise we are able to detect viral protein signatures across time and space with high accuracy using existing archived proteomics datasets. Currently we are focusing on the deep proteome characterization of a series of samples from a single location collected over a two year interval (C75 time series). One C75 sample captured a viral bloom (Banfield group). The detection of proteins from such viral blooms in metagenomically characterized samples should be possible. We are employing two new technology approaches to achieve deep detection of the viral signature. The first technology advance is the use of intact protein separations prior to digestion and 2d-LC-MS/MS. The second is the use of a new rapid scanning LTQ-Orbitrap Velos, which should provide a dramatic increase in proteome depth compared to LTQ-Orbitrap. One of the most promising methods to enhance detection of viruses is density gradient centrifugation of viruses extracted from AMD biofilms. These enriched viral protein samples have led to the highest detection rate of viral proteins from AMD biofilms to date. We are currently analyzing a new enrichment set from a time series. Finally we are using *de novo* proteomic sequencing methods to attempt to identify viral strain variants that would have otherwise escaped detection.

Funding provided by DOE Genomics:GTL Program grant number DE-FG02- 07ER64505. Oak Ridge National Laboratory is managed by UT-Battelle, LLC, for the U.S. Department of Energy under Contract No. DE-AC05-00OR22725.

146

Integrating Experimental and Computational Approaches to Enhance Proteogenomic Characterizations of Natural Microbial Communities

Robert L. Hettich[1] (hettichrl@ornl.gov), Nathan VerBerkmoes,[1] Chongle Pan,[1] Brian Erickson,[1] Brian Dill,[1] Alison Russell,[1] Manesh Shah,[1] Doug Hyatt,[1] Ryan Mueller,[2] Vincent Denef,[2] Paul Wilmes,[2] Brian Thomas,[2] Brett Baker,[2] Korin Wheeler,[3] Steve Singer,[3] Michael Thelen,[3] and Jillian Banfield[2]

[1]Oak Ridge National Laboratory, Oak Ridge, Tenn.; [2]University of California, Berkeley; and [3]Lawrence Livermore National Laboratory, Livermore, Calif.

Project Goals: To establish a proteogenomic approach for elucidating how microbial consortia assemble and respond to their environmental pressures. To this end, we have focused on a model environmental microbial community found in acid mine drainage to develop and demonstrate a MS-based proteomic approach to investigate the molecular level activities of microbial consortia. Whole community genomics serves as the underpinning core for the measurements and evaluations of microbial consortia.

The characterization of natural microbial communities at a systems biology level can be addressed in significant detail with a technology platform based on high performance mass spectrometry interfaced closely with computational approaches for mining the raw MS data, compiling protein results, and finally integrating with physiological and genomic information. To this end, we have focused on a model environmental microbial community found in acid mine drainage to develop and establish a proteogenomic approach for elucidating how microbial consortia assemble and respond to their environmental pressures. Whole community genomics serves as the underpinning core for the subsequent measurements and evaluations of microbial consortia.

We have greatly expanded the range of temporal and spatially resolved AMD samples measured by proteogenomic techniques, and have analyzed over 50 distinct AMD biofilms to date, with an average of 2,000-4,000 protein identifications per site. The goal of this work is to explore the genomic and proteomic diversity of the AMD system. The resulting data has significantly expanded the molecular-level view of the AMD microbial system by revealing a deeper level of molecular information for not only the bacteria, but also archaea, nanoarchaea, and viruses as well. Recent work has focused on quantitative environmental proteomics to examine the effects of inter-species interactions on population physiology in natural microbial communities as ecological succession proceeds. For this work we sampled three defined developmental stages (early, intermediate, and late) in triplicate from the same natural environment and measured soluble and membrane proteome fractions for changes in the abundances of proteins from dominant and sub-dominant community members via label free and metabolic labeling quantitative proteomics. We identified approximately 6000 proteins across all samples and replicates (2522 from the membrane fraction and 3437 from the soluble fraction) from fifteen distinct genetic populations of this community. Biological replicates demonstrated high concordance, which deteriorated slightly as communities became more complex. The transition from early developmental stage biofilms to late developmental stage biofilms coincides with a physiological switch in the dominant population of these communities. Functional biases in early developmental stages pertain to rapid growth and protein synthesis (*i.e.*, cell division, ribosome assembly, transcription), whereas biases in the late developmental stage include functions for building block biosynthesis (AA, CHO, Lipids), chemical sensing, and translation. This switch may reflect physiological constraints imposed on this dominant member by the arrival of and competition with secondary colonizers in late stages of biofilm development.

One outcome of the proteogenomic work is the observation of a substantial sequence divergence in the genomes of closely-related environmental microbes, mostly in the form of single nucleotide polymorphisms, which translate into single amino acid variants at the protein level. Clearly, it is important to characterize these amino acid variants in order to understand their functional roles. We have developed a database-searching algorithm, termed *Polyscan*, to identify amino acid polymorphisms by shotgun proteomics. *Polyscan* systematically mutates *in-silico* one residue at a time for every protein to generate tryptic peptides with polymorphisms, and then correlates mutated peptides to the acquired tandem mass spectra. To efficiently search this enormously expanded sequence space, *Polyscan* uses a binary search to assign spectra to candidate peptides by parent masses and then scores peptides with a fast preliminary scoring function followed by a slower, but more accurate, primary scoring function. *Polyscan* was used to analyze a proteome of a natural microbial community in acid mine drainage. In the data analysis, the mutations of N to D, Q to E, and A to S were excluded, as these mutations have the same mass shift as common PTMs on the original amino acids. A total of 1102 polymorphisms were identified at a false discovery rate of 1.2% in an AMD sample measured by high resolution LTQ-Orbitrap-MS. The frequency of these polymorphisms correlates well with their BLOSUM62 propensity. This provides support for the use of *Polyscan* to identify amino acid polymorphisms in shotgun proteomics datasets.

We also have undertaken a characterization of the metalloproteome from the AMD system. Metals play essential roles in cellular metabolism, with both beneficial and toxic effects. Proteins can bind metals in catalytic centers, utilize metals as structural elements, or serve as chaperones to safely transport metals to their appropriate location within a cell. In some microbes, metals can even drive an organism's metabolism, providing the necessary energy for life. Of great interest are the biochemical interactions between metals and proteins that drive geomicrobiological processes. By growing in molar concentrations of iron and millimolar concentrations of copper, zinc and arsenic, AMD microbial biofilms are ideal candidates for investigations into metal-influenced biogeochemical interactions. Investigations into the relationships between the AMD microbial community and the available heavy metals provide insights into the metabolic processes that enable life in this extreme environment. We have completed an initial assessment of the metal-binding proteome for an AMD microbial community. We employed selective enrichment of metal-binding proteins with immobilized metal affinity chromatography (IMAC) across seven different biologically active metals (copper, cobalt, manganese, magnesium, nickel, zinc, and iron) with a specific focus on identifying bound vs. unbound proteins in each case. On average, about 270 proteins were identified in each column fraction, with an average of 9% variability between replicates. In every column except iron, there were more proteins identified in the unbound fraction, with about 54% of bound proteins having no known function. The largest groups of specific metal binding proteins are involved in translation, ribosomal structure, and biogenesis, or post-translational modification, protein turnover and chaperones.

In total, our proteogenomic methodology provides a powerful approach for the systems biology interrogation of natural microbial communities, and should be broadly applicable for microbial systems of relevance for environmental remediation and bioenergy production.

This research sponsored by the U.S. DOE-BER, Genomics Program. Oak Ridge National Laboratory is managed by

UT-Battelle, LLC, for the U.S. Department of Energy under Contract No. DE-AC05-00OR22725.

147

Understanding Carbon Cycling in a Model Microbial Community by Integration of Stable Isotope Probing, Metabolomic, and Transcriptomic Data With Proteogenomic Analyses

Curt R. Fischer[1]* (curt_f@berkeley.edu), Brett J. Baker,[1] Paul Wilmes,[1] Chris Belnap,[1] Chongle Pan,[2] Shufen Ma,[1] Trent Northen,[3] Nathan C. VerBerkmoes,[2] Benjamin Bowen,[3] Susan Spaulding,[1] Pepper Yelton,[1] Nicholas Justice,[1] Ryan S. Mueller,[1] Vincent J. Denef,[1] Brian C. Thomas,[1] Steven W. Singer,[3,4] Michael P. Thelen,[4] Robert L. Hettich,[2] and **Jillian F. Banfield**[1]

[1]University of California, Berkeley; [2]Oak Ridge National Laboratory, Oak Ridge, Tenn.; [3]Lawrence Berkeley National Laboratory, Berkeley, Calif.; and [4]Lawrence Livermore National Laboratory, Livermore, Calif.

Project Goals: Our recent aims have included development of methods to achieve a comprehensive, molecularly-resolved, multi-scale understanding of carbon and energy flow in a model chemoautotrophic microbial community of tractable complexity, in order to lay a foundation for work in more complex environments. The first level of our analysis is to define the metabolic potential of coexisting natural populations through metagenomic sequencing. The second level of our analysis includes whole-community shotgun proteomics and mass-spectrometry-based metabolomics to identify key chemical and biochemical intermediates of carbon and energy transfer in these communities. The third level of our analysis aims at systems-level studies of the community to track the flow of carbon and energy over a range of environmental conditions and relies on cultivation of the entire community.

Microbial communities drive global biogeochemical cycling of energy and carbon, and a molecular understanding of how they respond to environmental stresses can inform predictions of how cycles will change in the future. However, many natural ecosystems are too complex to currently study in molecular detail. Our aim is to develop methods to achieve a comprehensive, molecularly-resolved, multi-scale understanding of carbon and energy flow in a model chemoautotrophic microbial community of tractable complexity, in order to lay a foundation for work in more complex environments. Our studies toward this end can be broken into three levels. The first level of our analysis is to define the metabolic potential of community members through metagenomic sequencing. In past GTL-sponsored research, we have assembled high-quality strain-resolved genomes for the three bacterial iron-oxidizing primary producers of our model community and draft genome assemblies for six archaeal mixotrophs of the order *Thermoplasmatales*. Recent functional annotations indicated that all of these archaea are facultative anaerobes and have complete pathways for glycolysis, beta-oxidation, and polypeptide degradation. We also assembled near-complete genomes for three deeply-branched, low abundance and uncultivated lineages of Euryarchaea (referred to as ARMAN-2, 4 and 5). The small genome size, approximately 1 Mb, is consistent with their small cell size (~500 nm in diameter). A shorter than average gene length allows for a 10% higher coding density than is achieved by other genomes of comparable size. Two of the ARMAN groups have complete glycolytic pathways and the third has a beta-oxidation pathway for fatty acid utilization, thus all are likely capable of heterotrophic growth.

The second level of our analysis includes whole-community shotgun proteomics and mass-spectrometry-based metabolomics to identify key chemical and biochemical intermediates of carbon and energy transfer in these communities. Due to the complexity of the metabolome and lack of appropriate reference spectra, it was necessary to analyze metabolomic data without feature identification. Over 8,000 metabolic features were identified from a collection of 15 samples, representing 7 environment types. Proteomic data, in which detected peptides are matched to proteins predicted from assembled genomes, were also collected for each sample, and metabolomic features were co-analyzed with proteomic data. Correlation and clustering analyses revealed two large, well-resolved groups of metabolites. Each group strongly correlates with proteins from only one of the two dominant bacteria. This finding suggests that evolutionary divergence of these two co-existing bacteria has resulted in divergent metabolite profiles, possibly reflecting differences in genome regulation, kinetics of orthologous enzymes, and the presence of organism-specific metabolites.

The third level of our analysis aims at systems-level studies of the community to track the flow of carbon and energy over a range of environmental conditions and relies on cultivation of the entire community. Laboratory growth of biofilms allows both precise control over environmental conditions and the use of stable isotope as labels or tracers. Our culture system uses continuous-flow peristaltic pumps to move an acidic, iron-rich growth medium through custom-designed open-air Teflon channels situated in a temperature-controlled chamber at 40 °C. We have observed gross metabolic rates of between 1 and 15 W/m^2 in the bioreactor, as calculated from Fe^{2+} oxidation rates. This is comparable to, and sometimes higher than, field-observed values of 0.6 to 3.7 W/m^2 for our model system, and additionally is similar to rates of phototrophic primary production in terrestrial ecosystems.

We have validated that we can track and quantify stable isotopic tracers in the proteomes and metabolomes of our model community. In one proof-of-principle experiment, community proteomes were extracted from two bioreactors, one of which was cultivated with a 50:50 mixture of $^{15}NH_4^+$ and $^{14}NH_4^+$ as the sole nitrogen source, and the other with $^{15}NH_4^+$ as the sole N source. We used the Polyscan algorithm to identify both the sequence and ^{15}N isotopic

enrichment ratio of 1449 peptides and 2404 peptides in the two samples, respectively. The median peptide isotopic enrichment in the isotopically pure $^{15}NH_4^+$ sample was 99%, with the 20th and 80th percentiles being 98% and 100%, illustrating that the vast majority of calculated peptide isotopic enrichments were similar to the experimentally imposed value. The median peptide isotopic enrichment in the 50% labeled sample was 49%, with the 20th and 80th percentiles being 47% and 49%, showing that even the complex mixtures of peptide isotopomers that arise in partially labeled samples can be correctly identified. Separately, in a metabolomics proof of principle experiment, we quantified deuterium incorporation into phosphatidylethanolamine lipids from consortia grown in medium with 10% deuterated water using mass-spectrometry-based metabolomics. The experiments indicate that stable isotopic tracers can be quantified using traditional proteomic and metabolomic mass spectrometry techniques, and that these techniques can be used to trace the flow of stable isotopic tracers throughout community metabolomes and proteomes.

Taken together, the data are beginning to provide a highly integrated picture of carbon cycling in our model community. Carbon is initially fixed by two dominant bacteria (*Leptospirillum* spp. of the Nitrospirae), which also produce cellulose-based polysaccharides and osmolytes that become substrates for the growth of archaea, low abundance bacteria, and fungi. Archaea grow heterotrophically or mixotrophically by beta-oxidation of fatty acids or via other complex organic carbon degradation pathways. Fungi appear in later developmental stage aerobic biofilms. We speculate that fungi help catalyze the biodegradation of cellulose and other polysaccharides in the sulfuric acid-rich pH 1 environment of our model ecosystem, and so we are also targeting these organisms for metagenomic and metatranscriptomic sequencing. Less dominant community members, including actinobacteria and the ultra-small ARMAN archaea are ubiquitous and have intimate interactions that likely involve C transfer with as yet unidentified cell wall-less archaea. Sulfite- and possibly Fe^{3+}-reducing archaea and bacteria may play roles in anaerobic metabolism of organic carbon. Initial evidence in this direction comes from microelectrode-based analysis of waters and sediments from mine and lab bioreactor samples, as well as initial proteogenomic characterization of biofilms sampled from submerged, anaerobic environments. Microelectrode measurements have revealed thiosulfate, sulfite and polysulfides which we believe may be involved in community energy cycling. We have detected shifts in community population structure in sunken biofilms, with increased levels of *Sulfobacillus* and *Thermoplasmatales* archaea relative to surface biofilms. Future transcriptomic, proteomic, metabolomic and stable-isotope labeling experiments will test these hypotheses. Both the methods and the novel integrative bioinformatic architecture that we have developed may become extensible to other, more complex ecosystems as the metagenomics knowledge base for these ecosystems is developed.

Funding provided by DOE Genomics:GTL Program grant number DE-FG02-05ER64134.

148

Sub-Proteomic Analyses and Structural Modeling of Proteins Expressed in Acidophilic Microbial Communities

Yongqin Jiao[1]* (jiao1@llnl.gov), Pepper Yelton,[2] Patrik D'haeseleer,[1] Adam Zemla,[1] Brian Dill,[3] Nathan Verberkmoes,[3] Robert Hettich,[3] **Jill Banfield**,[2] and **Michael Thelen**[1] (mthelen@llnl.gov)

[1]Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory, Livermore, Calif.; [2]Dept. of Environmental Science, Policy, and Management, University of California, Berkeley; and [3]Chemical Sciences Division, Oak Ridge National Laboratory, Oak Ridge, Tenn.

Project Goals: To gain a comprehensive understanding of an extremophile microbial community, its member interactions, biofilm formation, and geochemical impact using an approach that integrates metagenomics, MS proteomics, and biochemistry.

EPS matrix proteome

Extracellular polymeric substances (EPS) are major structural components in biofilms, providing spatial organization and structural stability to the microbial community. The exact composition of EPS varies substantially between different biofilms, and it remains a considerable challenge to provide a complete biochemical profile for interpreting how changes in EPS constituents affect community organization and development. In this study, we performed proteomic analysis to identify proteins that are present in the EPS matrix, and discuss correlations between EPS proteins and their potential microbial functions.

As part of our ongoing investigation of acidophilic microbial communities that generate acid mine drainage (AMD) at the Richmond Mine in Northern California, we examined proteins extracted from a mid-developmental stage biofilm (DS1) and a mature biofilm (DS2). Each biofilm was separated into EPS, whole cell, and mine solution fractions, and analyzed by MS proteomics following trypsin digestion. To determine proteins enriched or associated with the biofilm matrix, we compared the relative abundance of proteins identified in EPS with whole cell and mine solution proteins as a function of biofilm developmental stage.

In order to augment the existing functional annotation of the proteins, we applied a variety of standard subcellular localization prediction tools, and assigned putative enzymatic functions (EC numbers) using PRIAM. BLAST scores from matches within the Carbohydrate Active enZyme database (CAZy) were also used to assign protein families and additional EC numbers. The composition of the proteome in the EPS and mine solution fraction is very different from that in the cellular fraction, with more than 80% of the proteins detected in the cellular fraction underrepresented or undetectable in the EPS or mine solution. In contrast, predicted periplasmic and extracellular proteins are

over-represented by 3-7 fold in the EPS and mine solution compared to the whole cell fraction. Likewise, predicted outer membrane proteins are ~3-fold over-represented in the EPS. Some of these over-represented categories are dominated by a few abundant proteins. For example, much of the overabundance of extracellular proteins in the EPS fraction can be attributed to highly abundant putative flagellin proteins.

Enzymes overall are under-represented in the EPS and mine solution proteome: whereas around 50% of the peptide count for the cell fraction is associated with proteins that have an enzyme annotation, only around 10% of the peptide count of the mine solution and EPS fraction appears to be enzymatic. We also noticed over/under-representation of specific enzymatic functions (EC numbers) in different fractions. For example, among the most abundant enzyme functions, a protein disulfide-isomerase (EC 5.3.4.1), a peptidase (EC 3.4.24.64), and an adenylate kinase (EC 2.7.4.3, 2.7.4.10) are over-represented in the EPS and mine solution compared to the whole cell fraction, although most enzymes (e.g. carbohydrate-active) are estimated at much lower abundance.

Differences between DS1 and DS2 for each fraction are smaller than the differences between the fractions, and are likely due to variation in community composition, in addition to differential protein secretion and degradation during biofilm maturation. Some disparity in overall numbers of proteins belonging to the families of Glycoside Hydrolases, Glycosyl Transferases, and Carbohydrate Esterases were observed between the DS1 and DS2 proteomes, and may shed light on the maturation and carbohydrate composition of the EPS matrix.

The biofilm archaea

To analyze proteins expressed specifically from the Archaeal members of the biofilm community, we used proteomics, gene synteny, and structural homology. Gene order conservation (synteny) over evolutionary time is relatively rare. Genomes tend to be shuffled rapidly except in close relatives and in circumstances where neighboring genes are functionally related. This leads to a strong correlation between synteny and all measures of evolutionary distance. Preservation of synteny over large evolutionary distances should be weighted more strongly in gene function prediction because it is likely to not be due to chance but to result from negative selection against rearrangements. Based on this principle we developed our own weighted synteny-based approach to more accurately annotate poorly annotated genes in a group of co-occurring AMD Archaea. These organisms comprise one lineage consisting of four newly assembled genomes (A-, E-, G-, and I-plasma) in addition to five previously sequenced organisms of the order *Thermoplasmatales*.

Because tertiary structure is of primary importance in protein function, structural homology modeling was used to improve upon synteny-based annotations. Using a modified, high throughput version of an automated homology modeling system AS2TS, we predicted full or partial structures of a number of proteins belonging to these Archaea. Combined with synteny data, models of these poorly annotated proteins have allowed us to better predict protein function and thus further understand the role of these Archaea in the AMD community. Our results indicate that two of these organisms, A- and I-plasma, are capable of making the molybdopterin guanine dinucleotide cofactor and using it in anaerobic energy conservation metabolism. Proteomics confirms that some of these proteins are expressed in anaerobic sunken biofilms. This may help these species to thrive in adverse anaerobic conditions where other AMD Archaea cannot.

We also used this approach to annotate a region of proteins of unknown function found in only one of the AMD Archaea, G-plasma. This region of nine neighboring genes is notable in that most of the genes have no homologs in any of the other Archaea in this lineage or in available gene or protein databases. Nevertheless, proteomic data suggest that eight of the nine genes are expressed in certain conditions. In fact, one of the most highly detected G-plasma proteins is among this group. Protein modeling has been utilized to assign function to some of these unusual proteins. As a whole, these proteins appear to be involved in some type of protein or peptide modification and export or import. They model to reference proteins including a sulfurtransferase, a peptidase, a glutamyl-transpeptidase, and a large pore ABC transporter-like protein. Based on these annotations and proteomic data, we infer that this gene cluster plays an important role in G-plasma's differentiation from its co-occurring close relatives. The novel genes may be exporting peptides for a number of purposes to that effect, including for quorum sensing or as peptide antibiotics that make G-plasma more competitive.

This work was funded by the DOE Genomics:GTL Program grant number DE-FG02-05ER64134, and performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory and University of California, Berkeley under Contract DE-AC52-07NA27344. LLNL-ABS-420907.

149

Bottom Up Genome-Scale (BUGS) Modeling of Microbial Communities Influencing the Fate and Transport of Groundwater Contaminants

R. Mahadevan[2]* (krishna.mahadevan@utoronto.ca), Kai Zhuang,[2] Nadeera Jayasinghe,[2] Srinath Garg,[2] Jiao Zhao,[2] Tae Hoon Yang,[1] Jun Sun,[1] Yilin Fang,[3] Tim Scheibe,[3] Melissa Barlett,[4] Maddalena V. Coppi,[4] and **Derek Lovley**[4] (dlovey@microbio.umass.edu)

[1]Genomatica Inc., San Diego, Calif.; [2]University of Toronto, Toronto, Ontario; [3]Pacific Northwest National Laboratory, Richland, Wash.; and [4]University of Massachusetts, Amherst

Project Goals (Abstracts 149-157): The primary goals of this project are to develop genome-based experimental approaches to elucidate the function of anaerobic microbial communities involved in groundwater bioremediation and novel bioenergy processes and to develop genome-scale models of these communities in order to better understand and optimize bioremediation and bioenergy applications. One area of focus is the subsurface microbial community involved in the in situ bioremediation of uranium-contaminated groundwater. The other focus area are the microorganisms that can exchange electrons with electrodes, which can be employed in the bioremediation of subsurface environments, harvesting electricity from organic matter, and fixation of carbon dioxide into desirable organic products. These studies are demonstrating that with the appropriate genome-scale models it is possible to predict the in situ growth and activity of the relevant microorganisms in the environments of interest solely from relevant geochemical data and that the in situ metabolic state of the microorganisms can be diagnosed via analysis of gene expression of the subsurface community. It is expected that these studies will provide computational tools that can be used to predict the response of the microbial community to environmental manipulations, or manipulation of the genome of the relevant organisms, allowing rational optimization bioremediation and bioenergy applications via environmental or genetic engineering.

Studies at the Department of Energy field study site in Rifle, Colorado have demonstrated that a diversity of microorganisms are competing for resources in the subsurface and that the outcome of this competition can greatly impact the fate and transport of groundwater contaminants such as uranium. In order to better understand these interactions of microorganisms and their environment and their influence on uranium mobility, we have initiated a process known as BUGS (Bottom Up Genome-Scale) Modeling which has the ultimate goal of developing the ability to predict the efficacy of various bioremediation approaches, whether engineered or natural attenuation, prior to their implementation in the field.

In the BUGS Modeling approach, the most abundant, metabolically active microbes in the subsurface are identified with molecular techniques. Representatives of these key members of the subsurface microbial community are recovered in culture and genome scale metabolic models are developed for each microorganism. The genome-scale metabolic models are then coupled with geochemical and hydrological models to determine whether the models can predict the geochemical consequences of the response of the microbial community to various environmental perturbations in laboratory and field studies.

In previous studies it was demonstrated that the BUGS Modeling approach successfully predict the changes in uranium concentrations in groundwater at the Rifle site following the addition of acetate to the groundwater. This was a relatively simple test case because over 90% of the active microbial community was comprised of *Geobacter* species over the time frame modeled. It was also demonstrated that BUGS Modeling could describe the competition for resources between *Rhodoferax* species, another important group of Fe (III) reducers at the site, and *Geobacter* species and predict the relative distribution of these organisms under different geochemical regimes during natural attenuation as well as engineered bioremediation.

The initial bloom of *Geobacter* and successful removal of uranium from the groundwater following acetate addition at the Rifle site is followed by an increase in sulfate-reducing microorganisms that are ineffective in uranium reduction. In order to better and understand this phenomenon and identify strategies for managing this aspect of the bioremediation the interaction between *Geobacter* and sulfate reducers was modeled. Simulations demonstrated that no matter what the starting conditions, *Geobacter* species were able to dominate the initial phase of engineered bioremediation with very little change in timing of the onset of sulfate-reduction. The field-scale simulation was able to accurately predict the data from field experiments. The results suggest that the observed succession of *Geobacter* and sulfate reducers can primarily be attributed to differences in growth rates rather than other factors, such as competition for acetate. The simulations suggest that addition of Fe (III) to the subsurface can extend the time frame for effective uranium removal, but it is essential that the Fe (III) be added early in the bioremediation before sulfate reducers have had a chance to proliferate.

Another environment that strongly selects for *Geobacter* species is the surface of electrodes harvesting electricity from organic matter. Studies in many laboratories have demonstrated that in *Geobacter* species emerge from innocula from a diversity of environments as the predominant microbes on the anodes of the most highly effective microbial fuel cells. Current production in microbial fuel cells is controlled by a variety of factors including rates of substrate consumption and the accumulation of protons within the anode biofilm. It is important to be able to model the spatial gradients of substrates and inhibitors and to account for these heterogeneities on metabolism throughout the biofilm and their impact on current production. *Geobacter sulfurreducens*

produces the highest current densities of any known culture. Therefore, the genome-scale metabolic model of *Geobacter sulfurreducens* was coupled with a biofilm model in order to further investigate metabolism in anode biofilms. With this coupled model it has been possible to determine what regions of the biofilm are limited for substrate under different anode potentials and to evaluate the effect of different maintenance energy requirements on maximum current production and biofilm thickness. A strain of *G. sulfurreducens*, designated KN400, was recently selected in adaptive evolution studies that has substantially higher current-production capabilities than the starting strain, DL1. When estimated differences in cell yield and maximum substrate uptake rates were included in the model, the model accurately predicted the thinner, yet more effective current-producing biofilms of KN400. The model also successfully predicted the impact on current production of genetically engineering a futile cycle in DL1 to increase respiration rates. Additional development of the model to account for the impact of proton accumulation within the biofilm is underway.

The importance of the *G. sulfurreducens* genome-scale metabolic model for predictive analysis and optimization of *in situ* uranium bioremediation and microbial fuel cells justifies further refinement of this model. In order to further validate the model and to gain insights into intracellular metabolic response of *G. sulfurreducens* towards environmental perturbations its metabolism was further analyzed with ^{13}C-based metabolic flux analysis to characterize growth under conditions in which the electron donor (either acetate or hydrogen) or the electron acceptor (either Fe (III) or fumarate) limited growth. The donor and acceptor variations gave rise to differences in the pathways for gluconeogenesis, tricarboxylic acid cycle activities, and amino acid synthesis pathways. For example, cells were able to utilize fumarate as both an electron acceptor and as an additional carbon source when it was provided with acetate. As a result, cells slightly elevated the metabolic fluxes in the tricarboxylic acid cycle, and gluconeogenesis was initiated by phosphoenolpyruvate decarboxylase. In contrast, direct conversion of acetyl-CoA to pyruvate was the main source for the gluconeogenesis when Fe (III) was provided as the electron acceptor. Furthermore, the net flux direction between acetyl-CoA and pyruvate was reversed with fumarate as electron donor compared to Fe (III) as electron acceptor. Fluxes in the tricarboxylic acid cycle were lower with hydrogen as an electron donor compared to when acetate was the electron donor. These results improve the ability of the model to represent the physiology of *Geobacter* species under a diversity of growth conditions.

Additional refinements to the *G. sulfurreducens* model included the development of a thermodynamics based metabolic flux analysis model that successfully identified reactions that are subject to regulatory control, consistent with gene expression data. A novel model of uranium reduction based on a recently discovered electron storage pathways was developed, which improved predictions of uranium reduction in column studies.

These studies demonstrate the value of the BUGS modeling approach for predictively modeling the response of microbial communities to geochemical gradients and environmental perturbations. It is expected that BUGS Modeling will have wide spread application for a range of environmental applications, including the response of microbial communities to climate change.

150

Feeding Microbes Electricity: Gene Expression and Deletion Analysis of Mechanisms for Electron Transfer from Electrodes to Microbes

Sarah M. Strycharz, Kelly P. Nevin* (knevin@microbio.umass.edu), Richard H. Glaven, and **Derek R. Lovley**

Dept. of Microbiology, University of Massachusetts, Amherst

http://www.geobacter.org

Project Goals: See goals for abstract 149.

A rapidly emerging area in bioenergy and bioremediation is the possibility of driving beneficial microbial processes with electrons supplied with electrodes. This is a reverse in electron flow from the more commonly considered microbe-electrode interaction of current production in microbial fuel cells. Potential applications for feeding microbes electrons include bioremediation of waters contaminated with organics, radionuclides, metals, and/or nutrients; as well as the production of fuels and chemicals. Electrons can be supplied from a diversity of sources including wind and solar energy. In order to optimize these applications it is important to understand how microorganisms can use electrons supplied with electrodes as an energy source.

We have identified the first pure cultures capable of directly accepting electrons from electrode surfaces. For example, *Geobacter* species have been identified that can catalyze the reduction of organic acids, nitrate, chlorinated solvents, and metals with electrons supplied from an electrode as the sole electron donor. Other electron acceptors and other microorganisms with practical applications will be revealed at the time of the meeting.

Initial mechanistic studies were carried out with *Geobacter sulfurreducens*. Electrons were supplied with a graphite electrode poised at -500 mV versus Ag/AgCl. Fumarate was provided as the sole electron acceptor. Gene expression in current-consuming cells growing as biofilms on the graphite surface was compared with gene expression in two alternative growth modes: 1) biofilms growing on the same graphite material, but with acetate as the electron donor, rather than electrons from the electrode; and 2) biofilms growing on the same graphite material, but with acetate serving as the electron donor and the graphite serving as an electron-accepting electrode. These alternatives were termed no-current cells and current-producing cells, respectively.

Surprisingly, microarray analysis revealed that gene expression patterns in the current-consuming cells was significantly different than the previously reported gene expression patterns in current-producing cells. For example, current-producing cells have high expression of *pilA*, the gene encoding the structural protein for the pili termed microbial nanowires that are associated with high conductivity through current-producing biofilms. Current-producing cells also have enhanced expression of outer-surface *c*-type cytochrome genes, most notably OmcZ, which is required for optimal current production. However, *pilA*, *omcZ*, and other genes that are highly expressed in current-producing cells were expressed at low levels in current-consuming cells.

Furthermore, deletion of genes for *pilA* as well as *omcZ* and several other outer-surface *c*-type cytochromes had no impact on the current-consuming capabilities of *G. sulfurreducens*. This contrasts with the observed inhibition in current production when *pilA* or *omcZ* are deleted.

The concept that electrons are directly transferred from electrodes to *G. sulfurreducens* is based on several lines of evidence that suggest that hydrogen gas produced at the electrode surface is not an intermediate for electron transfer from the electrode to the cells. If hydrogen was an important electron transfer intermediate then it would be expected that genes for subunits of the uptake hydrogenase, Hyb, would be more highly expressed in current-consuming biofilms than current-producing or no-current biofilms. The gene for HybS, which encodes the small Hyb subunit, did have slightly higher transcript abundance in current-consuming cells versus the no-current control. However, this increase was just above the threshold for significance and the expression of the other Hyb subunits essential for a functional uptake hydrogenase were not upregulated in the current-consuming cells. Furthermore, transcript abundance for hydrogenases in current-consuming cells was not higher than in current-producing cells. Thus, gene expression analysis suggests that hydrogen was not an important intermediate in electron transfer from electrodes.

In both comparative studies cells accepting electrons from an electrode had a greater transcript abundance for a gene (GSU3274) encoding a putative monoheme *c*-type cytochrome. Deletion of this gene completely inhibited electron transfer from electrodes, but had no impact on the capacity for electron transfer to electrodes.

The amino acid sequence of GSU3274 shares homology most closely with putative cytochrome *c* family proteins from *Pelobacter propionicus*, *Thioalkalivibrio* sp., *Leptothrix cholodnii*, *Rhodoferax ferrireducens*, and *Polaromonas sp.*, all of which appear to contain a signal peptide cleavage domain suggesting they are translocated to the outer membrane. However, GSU3274 does not have a signal peptide cleavage site and is predicted to be localized in the periplasm. Therefore, it is unlikely that this protein serves as an electrical contact between the cells and the electrode. The structure of the protein encoded by GSU3274 is predicted to be similar to a cytochrome c_2 from the photosynthetic bacterium, *Rhodopila globiformis*, with a high redox potential. Thus, a potential role for the GSU3274 cytochrome is to serve as an intermediary in electron transfer between the outer cell surface and the inner membrane.

These results suggest that *G. sulfurreducens* has mechanisms for transferring electrons from electrodes that are substantially different than those for transferring electrons to electrodes. Differential gene expression in current-consuming electrodes versus current-producing electrodes may be dictated by the simple fact that the applied electrode potential influences the range of proteins for which electrode-cell interaction is energetically favorable. Furthermore, once electrons are transferred across the inner membrane, the remaining steps in electron transfer to electrodes do not require mechanisms for energy conservation, merely a pathway for electrons to flow down a potential gradient. In contrast, the pathway for electron transfer from electrodes into the cell must be specifically linked to a mechanism for generating a proton-motive force. These are significantly different metabolic demands.

Comparative analysis of gene expression in other organisms capable of accepting electrons from electrodes is underway and are expected to yield further insights into this important pathway for microbe-electrode interactions.

151

Immunocytochemistry Reveals Novel Models for the Role of Outer-Surface Cytochromes in Electron Transfer to Electrodes and Metals

Ching Leang* (leang@microbio.umass.edu), Kengo Inoue, Xinlei Qian, Tünde Mester, Trevor L. Woodard, Kelly P. Nevin, and **Derek R. Lovley** (dlovley@microbio.umass.edu)

University of Massachusetts, Amherst

Project Goals: See goals for abstract 149.

There are multiple competing/complementary models for extracellular electron transfer in Fe (III)- and electrode-reducing microorganisms. Which mechanisms prevail in different microorganisms or environmental conditions may greatly influence which microorganisms compete most successfully in sedimentary environments or on the surface of electrodes and can impact practical decisions on the best strategies to promote Fe (III) reduction for bioremediation applications or to enhance the power output of microbial fuel cells.

The three most commonly considered mechanisms for electron transfer to extracellular electron acceptors are: 1) direct contact between redox-active proteins on the outer surface of the cells and the electron acceptor; 2) electron transfer via soluble electron shuttling molecules; and 3) the conduction of electrons along pili or other filamentous structures. Evidence for mechanism #1 includes the findings that Fe (III)- and electrode-reducing microorganisms display redox-active proteins on their outer surface and that deletion of the genes

for these proteins often inhibits, at least partially, Fe (III) reduction and electron transfer to electrodes. Evidence for mechanism #2 includes the ability of some microorganisms to reduce Fe (III) which they cannot directly contact associated with the accumulation of redox-active soluble substances as well as specific electron shuttling electrochemical signatures of some microorganisms growing on an electrode surface. Evidence for mechanism #3 is more circumstantial and many authorities have questioned the potential for electron transfer along filaments.

OmcS is one of the most abundant cytochromes that can readily be sheared from the outer cell surface of *Geobacter sulfurreducens* and gene deletion studies have demonstrated that it is essential for the reduction of Fe (III) oxide as well as electron transfer to electrodes under some conditions. Therefore, the localization of this important protein was further investigated with immunogold labeling and electron microscopy.

With fumarate as the electron acceptor OmcS was in low abundance and primarily localized on the outer surface of the cells during early to mid-log phase growth. However, in cultures from late-log or stationary phase, the gold particles appeared as strands emanating from the cells. At higher magnification it was apparent that OmcS was associated with filaments with the same diameter and length of the previously described microbial nanowires of *G. sulfurreducens*. *G. sulfurreducens* produces more OmcS when grown under electron acceptor-limiting conditions and the greater abundance of OmcS associated with filaments was readily apparent in these cells. Filaments adorned with OmcS were also abundant in cells grown with Fe (III) oxide as the electron acceptor. When a strain of *G. sulfurreducens* in which the gene for OmcS was deleted was examined in the same manner there were no gold particles associated with the filaments.

This is the first description of cytochromes specifically associating with the filaments of a microorganism capable of extracellular electron transfer. The concept of cytochromes aligning with filaments of *Shewanella oneidensis* was previously proposed from indirect evidence and subsequent studies demonstrated that the cytochromes that were proposed to be aligned with filaments were randomly distributed within the extracellular matrix or associated with the outer cell surface rather than associated with filaments.

The dense packing of OmcS along the filaments suggests that cytochrome-to-cytochrome electron transfer might greatly facilitate electron transfer along the filaments.

These results do not negate the possibility that electrons may be transferred along the length of *G. sulfurreducens* pili without the need for cytochromes, but offer an alternative, possibly complimentary mechanism for long-range electron transfer.

Unlike OmcS, the outer-surface, *c*-type cytochrome OmcZ is not required for Fe (III) oxide reduction, but it is essential for optimal current production by *G. sulfurreducens*. In fumarate-grown cells OmcZ was dispersed throughout the extracellular matrix surrounding the cells of the biofilms that accumulated at the bottom of the culture tubes. When *G. sulfurreducens* grew as a biofilm on a graphite electrode that served as an anode and the sole electron acceptor for growth, OmcZ was highly concentrated at the biofilm-electrode interface. Controls in which the biofilm was grown on the same graphite material, but with fumarate as the electron acceptor, did not have accumulations of OmcZ at the electrode, corresponding with the inability of fumarate-grown biofilms to produce current. Although OmcS was also detected at the biofilm-electrode interface, it was also distributed throughout the biofilm. The specific localization of OmcZ at the anode surface under current-producing conditions, coupled with the previously published finding that deleting the gene for OmcZ dramatically increases the resistance of electron exchange between the anode and the biofilm, suggests that OmcZ may serve as an electrochemical gate promoting electron transfer from *G. sulfurreducens* biofilms to the anode surface. This proposed mechanism is consistent with the finding that the multiple hemes in the recently purified OmcZ protein have a wide range of mid-point potentials that may facilitate electron transfer at the divergent anode potentials that have been noted under different conditions.

The discovery of apparent novel roles for electron transfer for outer-surface *c*-type cytochromes in *G. sulfurreducens* emphasizes the relatively primitive understanding of extracellular electron transfer. Further investigations are warranted if the full potential of these organisms in practical applications such as bioremediation and conversion of organic compounds to electricity are to be realized.

152

Bioinformatic Analysis of Gene Regulation in the Metal-Reducing Bacterial Family *Geobacteraceae*

Julia Krushkal[1]* (jkrushka@utmem.edu), Ronald M. Adkins,[1] Yanhua Qu,[1] Jeanette Peeples,[1] Sreedhar Sontineni,[1] Ching Leang,[2] Peter Brown,[2] Nelson D. Young,[2] Toshiyuki Ueki,[2] Katy Juarez,[3] and **Derek R. Lovley**[2] (dlovley@microbio.umass.edu)

[1]University of Tennessee Health Science Center, Memphis; [2]University of Massachusetts, Amherst; and [3]Instituto de Biotecnología/UNAM, Cuernavaca, México

Project Goals: See goals for abstract 149.

Knowledge of how structural genome differences among microorganisms lead to variation in gene regulation is fundamentally important for our understanding of the functioning of gene regulatory pathways and their individual components. This knowledge is also necessary for our better understanding of the genomic changes leading to adaptation to diverse environments. Our research focuses on *Geobacteraceae*, a metal-reducing family of delta-Proteobacteria, which are capable of harvesting electricity from organic matter and

environmental bioremediation of organic and metal pollutants. We are investigating molecular mechanisms which allow these species to adapt and regulate their responses to environmental stimuli which result in energy production and removal of environmental pollutants. In this presentation, we report on our progress in computational analysis of gene regulation in species from this family. Our ongoing analyses include cataloguing and integrating information about gene regulation, *in silico* prediction of transcription factor binding sites, and investigation of functional effects of genome-scale and single nucleotide and amino acid level changes in transcription factors, operons, and regulatory sequence elements in *Geobacteraceae.*

Computational pipeline for analyses of variation in regulatory organization among species and strains. To investigate the regulatory impact of genome-level as well as nucleotide sequence-level evolutionary changes, we have developed a computational pipeline for analysis of regulatory changes in operon organization among species and strains of *Geobacteraceae.* This computational pipeline uses pairwise all-against-all sequence similarity searches for all genes obtained from two genomes of interest. Comparisons are made by MEGABLAST for nucleotide comparisons between strains of the same species and by tblastx for between-species comparisons at the protein sequence level, with soft filtering and E-value thresholds. We employ bi-directional best hits, coordinate matching, and gene neighborhood analyses to identify pairs of orthologs between the two strains and to validate operon assignments in each strain, while simultaneously recording gene annotations within each strain. After ortholog pairs and operon assignments have been unambiguously established, the pipeline tracks categories of structural gene changes (mutations and domain shuffling) and operon changes (e.g., full or partial deletions and duplications of the individual operons, merging of genes into new operons, splitting of individual operons into novel operons, as well as shuffling of groups of genes among different operons). These changes are further compared to known regulon assignments of individual genes of *Geobacter sulfurreducens*, an extensively studied model representative of *Geobacteraceae,* obtained across different microarray data experiments. This genome scale comparison allows identification of correlations between structural genome changes leading to operon shuffling and changes in gene co-regulation. Subsequent steps integrate these changes in genome sequence, gene co-regulation, and single base pair changes with available information on locations of regulatory sequence elements, which are suggested to influence expression of individual genes. We have applied this pipeline to pairwise comparisons of the *G. sulfurreducens* genome with the genome of the hypermutated strain, KN400, adapted for growth on electrodes [1] and with genome sequences of two other species, *G. metallireducens* and *G. uraniireducens.* This approach allowed us to identify operons which merged, split, or underwent gene reshuffling in the process of evolution. We also used bioinformatic approaches to investigate how single nucleotide level changes in the genome of the strain KN400 affected promoters, multiple transcription factor binding sites, and transcriptional and translational attenuators.

Bioinformatic analysis of transcriptional regulation. We are continuing our research investigating transcriptional regulation of specific biological pathways in *Geobacteraceae* which play important roles in electron transfer and environmental response. In our earlier studies, we investigated multiple regulatory pathways controlled by a variety of sigma factors and other transcriptional regulators. Now we are concentrating our analysis on two regulatory systems, the TetR family of transcriptional regulators and an enhancer binding protein, PilR. Members of the TetR bacterial family regulate expression of genes whose products are involved in a variety of important biological functions, e.g., osmotic stress, catabolic pathways, homeostasis, biosynthesis of antibiotics, efflux pumps, and multidrug resistance. Earlier studies of the TetR family [2] suggested that its members may be particularly abundant and diverse in microbial species exposed to environmental changes, due to their likely role in microbial adaptation to environmental changes. Due to their important roles, we investigated molecular evolution and regulation of 9 TetR family members in *G. sulfurreducens* (including functionally important regulators such as OrfR and AcrR, and other regulators with less established roles), their homologs in other species of *Geobacteraceae*, and their more distant relatives in other microbial species. In total, we performed an in depth analysis of 864 TetR family members across species of Bacteria and Archaea, and also investigated conservation of target operons regulated by TetR family members in *Geobacteraceae.* Our analyses demonstrated that TetR family regulators were not uniformly distributed among *Geobacteraceae,* suggesting either gene loss or horizontal gene transfer of some of these genes. While a number of TetR family members have been duplicated in certain species of *Geobacteraceae,* many of these duplications appeared to be species-specific.

We are also continuing our investigation of an enhancer binding protein, PilR, which regulates the expression of the *pilA* gene encoding structural pilin in an RpoN-dependent manner. *G. sulfurreducens* pili are electrically conductive and are required for Fe (III) oxide reduction and for optimal current production in microbial fuel cells. Previously we predicted PilR regulatory sites upstream the *pilA* gene and of other *G. sulfurreducens* genes whose products participate in biosynthesis, assembly, and function of pili and flagella, in secretory pathways, and in cell wall biogenesis. At present time, we are focusing our research on the biological roles of specific targets located downstream of predicted PilR sites, in order to better understand the regulatory role of PilR in *G. sulfurreducens.* In agreement with studies by other research groups, we are observing some intriguing similarities in genome regulatory sites affecting transcriptional regulation of the type II secretion system, which includes *pilA*, among environmentally important species of *Geobacteraceae*, and a number of bacterial pathogens, in which pili are important for pathogenesis and virulence.

Cataloguing regulatory information. We have developed and are continuing to update our online database, GSEL (*Geobacter* Sequence Elements) [3], which compiles regulatory information for *G. sulfurreducens*. At present time, we are incorporating the wealth of novel information on

transcription factor binding sites, attenuator sites, and riboswitches, which has been obtained by multiple research groups using experimental and computational approaches. The GSEL online server, which integrates information on operon organization, gene annotations, and regulatory sequence elements, is continuing to serve as a publicly available resource allowing users to investigate transcriptional regulation and regulatory interactions in *G. sulfurreducens*.

References

1. Yi, H., K.P. Nevin, B.C. Kim, A.E. Franks, A. Klimes, L.M. Tender, and D.R. Lovley, Selection of a variant of Geobacter sulfurreducens with enhanced capacity for current production in microbial fuel cells. Biosens Bioelectron, 2009. 24: 3498-3503.
2. Ramos, J.L., M. Martínez-Bueno, A.J. Molina-Henares, W. Terán, K. Watanabe, X. Zhang, M.T. Gallegos, R. Brennan, and R. Tobes, The TetR family of transcriptional repressors. Microbiol Mol Biol Rev, 2005. 69: 326-356.
3. Qu, Y., P. Brown, J.F. Barbe, M. Puljic, E. Merino, R.M. Adkins, D.R. Lovley, and J. Krushkal, GSEL v. 2, a genome-wide query system of operon organization and regulatory sequence elements of Geobacter sulfurreducens. OMICS, 2009. 413: 439-449.

153

Experimental Annotation of a Bacterial Genome

Yu Qiu[1]* (yuqiu@ucsd.edu), Byung-Kwan Cho,[1] Young Seoub Park,[1] **Derek Lovley**,[2] Bernhard Ø. Palsson,[1] and Karsten Zengler[1]

[1]Dept. of Bioengineering, University of California, San Diego and [2]Dept. of Microbiology, University of Massachusetts, Amherst

Project Goals: See goals for abstract 149.

Genome sequencing has been paramount to our understanding of biology. Precise genome annotation by computational prediction alone, however, has not been accomplished yet. Here we describe an experimental approach to annotate a bacterial genome that integrates proteomics and transcriptomics data with genome-wide data for RNA polymerase and sigma factor binding and single base-pair resolution transcription start site (TSS) determination. The proteogenomics approach alone yielded 55 new ORFs, 36 ORFs in intergenic regions and 19 pORFs in a different frame or on the opposite strand compared to current annotation. Additionally, we confirmed 241 genes that had previously been predicted only as hypothetical proteins, together representing an increase of ~9% compared to current annotation. The integration of transcriptomic data with RNAP holoenzyme binding regions and TSS data resulted in 748 and 694 RNAP-guided transcription segments (i.e. operons) on the forward and reverse strand, respectively, containing 2.3 genes in average. We identified 115 new RNAP-guided transcription segments that have not been annotated before, 70% of those representing antisense transcripts. In addition, the start of over 50 genes was corrected using TSS information and proteomic evidence. Furthermore, a large number of small non-coding RNAs (34), typically difficult to annotate using computational methods, has been identified and validated by Northern blotting. The data presented here demonstrate that more than 10% of the computationally generated annotation can be corrected using experimental data. More fundamentally, not only a structural annotation, i.e. the ORF structure of the genome, has been accomplished but at the same time the genome has been annotated on a higher operational level by elucidating its transcriptional architecture, i.e. the operon structure.

154

De novo Assembly of a Complete Microbial Genome Using Short Reads

Harish Nagarajan[1,2]* (nh@ucsd.edu), Jessica Butler,[3] Anna Klimes,[3] Yu Qiu,[2] Karsten Zengler,[2] Barbara A. Methé,[4] Bernhard Ø. Palsson,[2] **Derek R. Lovley**,[3] and Christian L. Barrett[2]

[1]Bioinformatics and Systems Biology Graduate Program and [2]Dept. of Bioengineering, University of California, San Diego; [3]Dept. of Microbiology, University of Massachusetts, Amherst; and [4]J. Craig Venter Institute, Rockville, Md.

Project Goals: See goals for abstract 149.

The development of next-generation sequencing technologies has greatly reduced the cost of sequencing per base and opened up a wide range of applications. One of the major applications include *de novo* sequencing of new microbial genomes. De novo sequencing using next-generation technologies have necessitated the development of new algorithms for assembling these short and more error-prone reads. Several de novo assembly algorithms (EULER-SR, Velvet, VCAKE, AllPaths, etc.) that are capable of assembling millions of short-reads from next-generation sequencing technologies into thousands of contigs with varying degrees of efficiency, have been recently developed. The complementary nature of Illumina and 454 reads has been exploited by some recent methods that have produced an assembly of *P. syringae pathovar oryzae* which consisted of 126 scaffolds with 2002 unincorporated contigs and an N50 of 91.5 kb. Another report integrated these two data types using a different approach to assemble an *Acinetobacter baylyi* strain into 10 scaffolds with an N50 of 1Mb.

Despite these recent reports that indicate significant progress by integrating Illumina and 454 technologies, complete de novo assembly of microbial genomes from only short reads and without aid from Sanger sequencing still remains an unsolved challenge. This challenge is critically important, for a single, circular nucleotide sequence of the complete chromosome is a necessary prerequisite for confident and complete research based on a genome. This fact is highlighted by a recent commentary that stresses on the importance and applications of high-quality genome sequences.

As an answer to this challenge, we have developed a successful integrative approach (meta-assembly) for combining next-generation sequencing technologies (Illumina and 454) to assemble a complete microbial genome de novo and applied it to a novel *Geobacter* variant (KN400) that is capable of unprecedented current production at an electrode. In addition to integrating the complementary data types (Illumina and 454), our meta-assembly strategy also leverages the different and complementary results provided by multiple assembly programs to obtain the complete sequence of a microbial genome.

Our meta-assembly approach consists of four distinct phases namely, Hybrid Assembly, Scaffold Bridging and Finishing, Scaffold Ordering and Genome Finishing. We applied this strategy using 50X Illumina GA1 singleton reads and 16X 454 GS-FLX paired-end sequencing reads for the novel *Geobacter* variant (KN400). By early integration of the two data types in the Hybrid Assembly phase, we were able to assemble the reads into just a few scaffolds. We then exploited the complementary assemblies generated by Newbler and EULER-SR in order to resolve the degenerate nucleotides in the scaffolds. We employed a PCR-based search strategy in order to obtain the correct relative ordering and orientation of the scaffolds. We corrected for indels and any errors introduced during our scaffold finishing and scaffold ordering phase by aligning the Illumina reads to the ordered scaffold and obtained a circular genome of length 3,714,259 bp.

KN400 is the first complete microbial genome sequence to be assembled from only short reads without the aid of Sanger sequencing. We found the completed KN400 genome to be collinear over its entire length with no major rearrangements and approximately 97% identical at the sequence level, to *Geobacter sulfurreducens* PCA. We further performed a comparative genomics analysis to identify unique genomic regions and thus provide functional insights into the observed novel phenotype of this particular strain.

This readily applicable strategy will result in a significant increase in the number of complete microbial genomes and should impact the quality of current and future sequencing projects. This strategy can also be applied to sequence all unknown members of a microbial community that can be physically separated thus providing the foundation for systems-level characterization of microbial communities by accelerating the rate of obtaining whole genome sequences of each of the community members.

155

Evidence for Direct Cell-to-Cell Electron Transfer from Adaptive Evolution, Genome Resequencing, and Gene Deletion Studies

Zarath M. Summers[1]* (zsummers@microbio.umass.edu), Heather Fogarty,[1] Ching Leang,[1] Ashley Franks,[1] Bernhard Palsson,[2] Yu Qiu,[2] Christian Barrett,[2] Karsten Zengler,[2] and **Derek R. Lovley**[1]

[1]Dept. of Microbiology, University of Massachusetts, Amherst; and [2] University of California, San Diego

http://www.geobacter.org

Project Goals: See goals for abstract 149.

Syntrophic transfer of electrons between cells of different species is a central feature in the function of many anaerobic ecosystems that impact on the global carbon cycle, as well as for several types of bioremediation and production of methane, an important biofuel. For forty years the primary model for interspecies electron transfer has been interspecies hydrogen transfer in which one syntrophic partner disposes of electrons by reducing protons to produce hydrogen and the other partner uses the hydrogen as an electron donor in respiration.

However, interspecies hydrogen transfer is a rather inefficient mechanism for two microorganisms to share electrons because of potential energy loses in the multiple steps required for hydrogen production and reoxidation. We hypothesize that microorganisms can directly exchange electrons via cell to cell electrical contacts that may involve conductive filaments known as microbial nanowires and/or electron transfer via extracellular, multiheme *c*-type cytochromes.

In order to evaluate how syntrophic cell-to-cell electron transfer might evolve a co-culture of *Geobacter metallireducens* and *Geobacter sulfurreducens* was inititated in a medium that contained ethanol as the sole electron donor and fumarate as the sole electron acceptor. Consistent with genome-scale metabolic modeling, *G. metallireducens* can metabolize ethanol, but cannot respire with fumarate and *G. sulfurreducens* can use fumarate as a terminal electron acceptor, but can not metabolize ethanol. However, it is potentially feasible for the two organisms to grow in ethanol-fumarate medium with *G. metallireducens* metabolizing ethanol to acetate and carbon dioxide with the release of electrons and with *G. sulfurreducens* consuming the electrons as well as the acetate with the reduction of fumarate.

Initially the co-culture grew very slowly and required 30 days to metabolize the ethanol provided. However, with continued transfer of a 1% inoculum the co-culture adapted to utilize the ethanol within 3 days. This increase in metabolic rate was accompanied by the formation of large (ca. **x** mm diameter) spherical aggregates. The formation of

aggregates and faster metabolism of ethanol were associated with less release of hydrogen from the co-culture.

The aggregates were comprised of approximately 15% *G. metallireducens* and 85% *G. sulfurreducens*. Fluorescent *in situ* hybridization (FISH) with species-specific probes of thin sections of the aggregates demonstrated that cells of *G. sulfurreducens* completely surrounded small spherical clusters of *G. metallireducens*. Transmission electron microscopy of aggregate thin sections revealed that there were abundant filaments, with the *c*-type cytochrome OmcS attached, coursing between the cells. This is consistent with the potential for direct cell-to-cell electron transfer.

Sequencing of the genomic DNA from the aggregates demonstrated that there was a single-base pair substitution in the gene for PilR in the *G. sulfurreducens* associated with the aggregate. Previous studies in our laboratory have demonstrated that PilR functions as an RpoN-dependent enhancer binding protein and regulates expression of genes encoding pilin and OmcS. Gene knock-in studies are underway to determine whether this mutation is responsible for the apparent pilin-OmcS network between the cells.

In order to further evaluate the role of hydrogen as the currency for cell-to-cell electron exchange a *G. metallireducens-G. sulfurreducens* co-culture was established with a strain of *G. sulfureducens* in which genes for the uptake hydrogenase, HybB, were deleted. This co-culture formed large, spherical aggregates within a matter of weeks, rather than the 7 months that were required for wild-type cells to form aggregates. These results suggest that preventing interspecies hydrogen transfer may increase selective pressure for direct cell-to-cell electron transfer. Genome resequencing of this syntrophic pair is underway to determine whether there were specific mutations associated with the more rapid development of aggregates.

These studies suggest that direct cell-to-cell electron transfer may be feasible and could be the mechanism of choice for cell-to-cell electron transfer when there is intense selective pressure for rapid syntrophic metabolism. Additional studies are underway with a diversity of mutants in which genes for *c*-type cytochromes, other redox active proteins, or pili have been deleted in order to further evaluate this hypothesis.

156

The Application of Metagenomic and Metatranscriptomic Methods to the Study of Microbial Communities in a Uranium-contaminated Subsurface Environment

Barbara Methé[1]* (bmethe@jcvi.org), Robert DeBoy,[1] Michael Geimer,[1] Peter Rosanelli,[1] and **Derek Lovley**[2]

[1]J. Craig Venter Institute, Rockville, Md. and [2]University of Massachusetts, Amherst

Project Goals: See goals for abstract 149.

Here we report on the application of metagenomic and metatranscriptomic approaches to the study of the microbial community at the Old Rifle site located in Rifle, Colorado (a former uranium ore-processing facility which is currently being managed as a part of the Uranium Mill Tailings Remedial Action (UMTRA) program of the U.S. Department of Energy) The study presented here is meant to complement and contribute to the wider ongoing investigations at this site through the application of next generation sequencing to metagenomic and metatranscriptomic analyses in an effort to reveal genetic information from subsurface microorganisms without the need for prior cultivation.

A metagenomic framework for use in this project was established from the 2007 sampling of the Rifle D05 site groundwater. A hybrid assembly has been constructed using 454 FLX pyrosequencing from which ~121MB of sequence was obtained from a 3-4 kb insert paired-end library and ~58 MB worth of Sanger sequencing paired-end reads from a 4-5kb insert plasmid library. Taxonomic classifications of either reads or assemblies of 454 and Sanger data, identified approximately half can be assigned at the Family level as members of the Geobacteraceae. Further, contigs that can be assigned to the genus *Geobacter* and are members of the Subsurface Clade 1, such as *G. bemidjiensis*, *G, uraniumreducens* and *G.* sp M21, predominate. From a whole genome alignment analysis, coverage of ~71% of the *G. bemidjiensis* genome was also revealed.

An incremental clustering method using the Cd-hit algorithm (a fast clustering algorithm based on shared word counts) was applied to the set of ~238,000 ORFs called from reads from both assemblies, and reads unincorporated into assemblies, to identify and classify protein families. Results from this investigation reveal that the most abundant of the clusters at 90% identity, and containing at least 100 members, are a reverse transcriptase family followed by several transposases indicating the ubiquitous nature of mobile elements in the microbial community. However, other highly abundant protein families of bacterial origin identified include DNA-directed RNA polymerase, a c-type cytochrome family, several families whose functions are most likely related to cell to cell, and cell to surface contact, and a family of heavy metal efflux pumps. In addition, several abundant hypothetical protein families were identified. This latter finding underscores one of the strengths of

* Presenting author

metagenomic analysis as new protein families can be identified that are not currently in the database. Results from the incremental clustering were also used to produce an estimate of library coverage using a rarefaction analysis which revealed that despite the significant amount of sequencing completed, the library has not been saturated (in other words, additional sequencing would reveal new ORFs and protein families).

The hybrid assembly was further evaluated through the JCVI Metagenomic pipeline which provides functional annotation of ORFs from multiple sources and is used to assign gene names, Enzyme Commission (EC) numbers, gene symbol, GO identifiers, and biological role categories. Further annotation and data mining was also performed using MG-RAST tools. An examination of HMM profiles (TIGRfams and Pfams) revealed that among the most abundant profiles are those with functions related to environmental sensing and signal transduction (e.g., PAS domain proteins, diguanylate cyclase domain proteins and sigma-54 interaction and response receiver domain proteins). Other prevalent HMM profiles based on relative abundance include profiles with functions related to exopolysaccharide sorting or transport to the cell surface and efflux transport.

Results from the annotation of ORFs from assemblies were also used to construct metabolic pathways to better understand the metabolic potential of the microbial community. Based on the use of KEGG maps, >50% overall representation of enzymes (including full and partial pathways) was determined in the categories of Purine and Pyrimidine Metabolism, Energy Metabolism, Amino Acid Metabolism and Carbohydrate Metabolism. Other categories of interest for which full and partial pathways could be constructed were the Biosynthesis of Secondary Metabolites and the Biodegradation of Xenobiotics categories. For instance, complete pathways were determined for 1,2-dichloroethane degradation and 3-chloroacrylic acid degradation. While 60% of the enzymes present in the pathway for benzoate degradation via CoA ligation, and 40% of the enzymes present in the pathway for benzoate degradation via hydroxylation, respectively were identified.

We have also begun experimentation designed to elucidate community gene expression through investigation of the metatranscriptome at several Rifle sites (including the 2007 samplings of D05 and D07 and the 2008 sampling of D04). We have used a capture oligonucleotide approach to enrich mRNA which is then used as template in cDNA reactions for subsequent sequencing via 454 pyrosequencing. A comparison of expression profiles from each site reveals an important result of this approach which is the identification of reads most closely related to multiple hypothetical ORFs and enzymes of unknown function from previously sequenced genomes. These findings suggest that these ORFs are true coding sequence whose biological roles remain to be determined. Further, each expression profile reveals evidence of the presence of non-coding RNAs which serve important roles in gene regulation. A comparison of each site in terms of taxonomic diversity reveals that although relatives to Subsurface Clade 1 organisms predominate in the metatranscriptomes, differences in microbial composition do exist. For example, the highest percentage of reads related to Subsurface Clade 1 organisms (~50%) were identified at the D05 site. A wide variety of transcripts with biological role categories has also been identified at each site and these results are currently undergoing additional investigations. Among the functions with greatest abundance of transcripts determined thus far, are ones related to protein synthesis and environmental sensing and signaling. Overall, the efforts presented here are central to several goals of this project especially the investigation of *in situ* microbial communities to generate new insights and hypotheses concerning metabolic processes and interactions between community members. Further, these results may also serve as a framework for predictive modeling of processes relevant to bioremediation.

Student Presentation

157

Tuning the Conductivity and Capacitance of *Geobacter sulfurreducens* Biofilms by Regulation of Gene Expression

Nikhil S. Malvankar[1,2]* (nikhil@physics.umass.edu), Kelly P. Nevin,[2] Ashley Franks,[2] Byoung-Chan Kim,[2] Kengo Inoue,[2] Tünde Mester,[2] Mark T. Tuominen,[1] and **Derek R. Lovley**[2]

[1]Dept. of Physics and [2]Dept. of Microbiology, University of Massachusetts, Amherst

http://www.geobacter.org

Project Goals: See goals for abstract 149.

The surprising finding that microorganisms can transfer electrons over substantial (>100 μm) distances to the surface of electrodes has expanded the potential applications of microbe-electrode interactions. Optimization of these applications requires an understanding of the mechanisms for this long-range electron transfer. One proposed mechanism is that, under the appropriate conditions, microorganisms may be able to form electrically conductive biofilms. However, previous studies have found that microbial biofilms act as insulators rather than conductors.

Geobacter sulfurreducens produces the highest current densities of any known microorganism. In order to investigate the possible conductivity of *G. sulfurreducens* biofilms, cells were grown in a microbial fuel cell that contained two gold electrodes as a "split-anode", separated by 50 μm non-conductive gap. *G. sulfurreducens* grew on the split anode forming a biofilm that bridged the non-conductive gap. Both the dc current-voltage characteristics and the ac impedance response between the two gold electrodes demonstrated that *G. sulfurreducens* biofilms are highly conductive. Various controls revealed that biofilm conductivity depended upon the physiological status of the cells.

To evaluate components that might confer conductivity to biofilms, novel electrode-selected variants and mutants, deficient in different outer membrane *c*-type cytochromes, were grown in the split-anode device. Strains generating more current produced biofilms with higher conductivity, demonstrating that higher biofilm conductivity facilitates higher current density and suggesting that conduction is an important mechanism for long-range electron transfer through the biofilms. Western immunoblot analysis revealed a direct correspondence between biofilm conductivity and the abundance of pilin protein. This is significant because previous studies have suggested that these pili are electrically conductive and may function as 'microbial nanowires'. An electrochemical gating technique further suggested the direct role of pili in biofilm conductance.

There was no correspondence between the number of *c*-type cytochromes and conductivity. Furthermore, biofilm conductance remained unchanged even after the activity of the cytochromes was inhibited by treating the biofilm with a reagent that unfolded the cytochromes. However, there was a direct relationship between capacitance and the abundance of cytochrome hemes in the biofilms. Capacitance decreased two orders of magnitude after cytochromes were treated with a denaturing reagent, further demonstrating that cytochromes contribute to capacitance. This is consistent with the concept that the abundant multi-heme *c*-type cytochromes in the periplasm and outer membrane of *Geobacter* species play an important ecological role, permitting temporary storage of electrons and allowing continued short-term electron transfer across the inner membrane when natural electron acceptors are temporarily unavailable.

These results demonstrate for the first time that biofilms can be electrically conductive and suggest that pili play an important role in conductance. The potential to grow conductive films and alter their electronic properties via genetic engineering is a significant advancement in the emerging field of bioelectronics.

158

CRISPR-Mediated Inhibition of Histidyl-tRNA Synthetase in *Geobacter sulfurreducens* Recreates an Evolutionary Pressure Experienced by *Pelobacter carbinolicus*

Muktak A. Aklujkar* (muktak@microbio.umass.edu) and **Derek R. Lovley**

University of Massachusetts, Amherst

http://www.geobacter.org

Project Goals:
Overall Goal: The primary goal of this research is to develop experimental and computational tools to predictively model the behavior of complex microbial communities involved in microbial processes of interest to the Department of Energy.

Five Year Goal: The five year goal is to provide computational tools to predictively model the behavior of two microbial communities of direct relevance to Department of Energy interests: 1) the microbial community responsible for *in situ* bioremediation of uranium in contaminated subsurface environments; and 2) the microbial community capable of harvesting electricity from waste organic matter and renewable biomass. Based on our current rate of progress, it is expected that within five years it will be possible to predict the *in situ* growth and activity of the relevant microorganisms in the environments of interest solely from relevant geochemical data and to describe in detail the *in situ* metabolic state of the microorganisms from environmental gene expression data. Furthermore, these computational tools will be able to predict the response of the microbial community to environmental manipulations or manipulation of the genome of the relevant organisms, allowing rational optimization of *in situ* uranium bioremediation or electricity harvesting via environmental or genetic engineering.

The family *Geobacteraceae* primarily contains Fe (III)-respiring bacteria capable of uranium bioremediation and electricity production. *Pelobacter carbinolicus*, a member of the family *Geobacteraceae*, cannot reduce Fe (III) directly, or produce electricity, and lacks the abundant *c*-type cytochromes implicated in electron transfer to insoluble extracellular acceptors. The evolutionary basis for these differences is an intriguing problem, and the sequencing of several genomes of *Geobacteraceae* allows it to be addressed. Genome analysis of *P. carbinolicus* revealed a sequence of 32 base pairs (spacer #1) within the clustered regularly interspaced short palindromic repeats (CRISPR) locus, which is identical to a segment of the histidyl-tRNA synthetase (*hisS*) gene. Recent studies have established CRISPR loci as the immunological memory of microbes, with individual spacers matching genes within phage or plasmid entities that previously infected the host, and providing resistance against recurring infection. However, spacers that match host housekeeping genes have not received attention. The model of CRISPR function predicts that spacer #1 should inhibit expression of *hisS* in *P. carbinolicus*. A shortage of histidyl-tRNA, in turn, should impede translation of proteins with multiple closely spaced histidines, predisposing them to mutation or elimination from the genome. A combination of genomics and genetics was used to investigate whether *P. carbinolicus* experienced this evolutionary pressure. Comparison of the *P. carbinolicus* genome with those of four other *Geobacteraceae* confirmed that genes with high histidine demand (computed as the number of histidines divided by the harmonic mean distance between histidines) are fewer in *P. carbinolicus* than in *Desulfuromonas acetoxidans*, *Geobacter bemidjiensis*, *Geobacter metallireducens* and *Geobacter sulfurreducens*. Sixteen gene families that are clearly ancestral have been either mutated to reduce histidine demand or lost by *P. carbinolicus*: these include several *c*-type cytochromes and a subunit of NADH dehydrogenase. Thus, inhibition of *hisS* by spacer #1 could have caused the metabolism of *P. carbinolicus* to shift away from respiration of extracellular electron acceptors. Moreover, when the *hisS* gene of *Geobacter sulfurreducens* was replaced with that of *P. carbinolicus*, growth with Fe

(III) as electron acceptor was totally inhibited by spacer #1 in the context of the *G. sulfurreducens* CRISPR, and growth with fumarate was strongly inhibited. The amount of *hisS* transcript RNA was reduced in this strain compared to a control lacking spacer #1, but the amount of transcript for a control gene, *hisZ*, was similarly reduced, indicating that this effect may be due to the reduced growth rate rather than RNA-level inhibition. Over forty transfers, the inhibited strain adapted to grow as well as the control, and to grow on Fe (III), but spacer #1 was not eliminated. The genome of the adapted strain was resequenced, and appears to bear a deletion of a noncoding region within a cluster of CRISPR-associated genes. Together, these investigations demonstrate that evolution of *P. carbinolicus* from a *Geobacter*-like ancestor, including the loss of multiheme cytochromes, which have high histidine demand, can be attributed in part to CRISPR-mediated inhibition of histidyl-tRNA synthetase. This is the first successful co-introduction of a CRISPR spacer and its target in the same cell, the first application of a hybrid CRISPR construct consisting of a spacer from one species in the context of repeats of another species, and the first report of a potential impact of CRISPR on genome-scale evolution by inhibition of an essential enzyme's expression.

Program Title: Genome-Based Models to Optimize *In Situ* Bioremediation of Uranium and Harvesting Electrical Energy from Waste Organic Matter

159

Genome-Scale Reconstruction of the Bacterial Transcriptional Regulatory Networks

Byung-Kwan Cho[1]* (bcho@ucsd.edu), Eric M. Knight,[1] Young Seoub Park,[1] **Derek Lovley**,[2] and Bernhard O. Palsson[1]

[1]Dept. of Bioengineering, University of California, San Diego; and [2]Dept. of Microbiology, University of Massachusetts, Amherst

Project Goals: See below.

The tremendous amount of novel genomic information is inspiring a new understanding of the bacterial genome on a global scale. Interactions between proteins and DNA for example are fundamental for cellular functions such as transcription, DNA replication, recombination, chromosome packing, and DNA repair. Among these, transcription is an essential step in gene expression and its understanding has been one of the major interests in molecular and cellular biology. By precisely tuning gene expression, transcriptional regulation determines the molecular machinery for developmental plasticity, homeostasis, and adaption.

The primary goals of this project are to develop genome-based experimental approaches to elucidate the function of anaerobic microbial communities involved in groundwater bioremediation and novel bioenergy processes and to develop genome-scale models of these communities in order to better understand and optimize bioremediation and bioenergy applications. It is expected that this study will provide experimental and computational tools that can be used to predict the response of the microbial community to environmental manipulations, or manipulation of the genome of the relevant organisms, allowing rational optimization bioremediation and bioenergy applications via environmental or genetic engineering. To address the issues, the genome-wide transcription factor binding regions (TFBR) - where transcription factors (TFs) bind to modulate the binding of the RNA polymerase (RNAP) - for the global transcription factors in *Escherichia coli* as a model system has been determined using ChIP-chip assays using high-resolution whole genome-tiling microarrays. Our results identified unique and reproducible TFBR from exponentially growing cells under various growth conditions. Most of the TFBR are located within intergenic regions; however significant number of TFBR were found within open reading frames and between divergently transcribed regions.

Changes in the binding levels of TFBR in response to environmental stimuli were strongly correlated with the changes in levels of mRNA transcript and RNAP occupancy. Three distinct regulatory modes (independent, concerted, and reciprocal modes) were identified through the integration of the ChIP-chip, mRNA transcript level, and RNAP occupancy level data. For the further understanding of the regulatory modes, we determined the alternative uses of transcription start sites (TSS) using massive-scale sequencing and subsequently reconstructed the causal relationships between TFBR and the TSS at the genome scale. Previously, we observed that ~35% of promoter regions contain multiple TSSs indicating the presence of alternative TSSs for large portions of the *E. coli* transcription units. Our results identified over several hundreds of direct interactions between global transcription factors (i.e., Fnr, ArcA, Crp, FruR, and Lrp) and TSS under the respective growth conditions. For example, there are two TSSs (3,595,753 and 3,595,778) for the *livKHMGF* operon. Although the TSS (3,595,753) is dominantly used to transcribe the operon, the transcription factor Lrp represses the other TSS (3,595,778) under exponentially grown cells in the absence of leucine. As demonstrated here, the alternative TSSs are used to regulate the bacterial transcriptome in response to different environmental stimuli. The phenomenon is likely to be tightly linked with the transcription regulatory network. Also, the use of alternative TSSs might be widespread in bacteria including the subsurface microbial community.

160

Periplasmic/Extra-Cytoplasmic Sensor Domains of Microbial Two-Component Transmembrane Signal Transduction and Chemotaxis Proteins

P. Raj Pokkuluri,[1] Hendrik Szurmant,[2] Andrzej Joachimiak,[1] and **Marianne Schiffer**[1]* (mschiffer@anl.gov)

[1]Biosciences Division, Argonne National Laboratory, Argonne, Ill. and [2]Dept. of Molecular and Experimental Medicine, The Scripps Research Institute, La Jolla, Calif.

http://www.bio.anl.gov/structural_biology/correlates.html

Project Goals: As sub-project of GTL grant "Genome-based models to optimize in situ bioremediation of uranium and harvesting electrical energy from waste organic matter, Derek Lovley (PI)" our goals are to analyze selected proteins to understand their function in the cell. This includes modeling of structures based on their amino acid sequences, determination of their structures, and the functional interpretation of the structures, such as active sites and surface properties.

Transmembrane signal transduction in bacteria is primarily mediated by the so-called two-component signal transduction system. Some of these proteins have extra-cytoplasmic sensor domains that interact with the environment of the bacteria and trigger responses in the cytoplasmic domains that effect the functioning of the cell. Chemotaxis proteins (Mcp) are part of this system that help the bacteria to sense its environment and move to a more favorable location. To generate an appropriate response to an environmental signal, a sensor histidine kinase (HisKa) adjusts its intrinsic autokinase activity upon detection of ligand(s). To understand the signals that trigger these proteins we have studied their soluble periplasmic sensor domains of these membrane proteins. Elucidating of the stimulatory signals and the corresponding transduction of the message by the above multidomain proteins is relevant in bacteria of both environmental and health significance.

The genome of *Geobacter sulfurreducens*, a Gram-negative organism important for bioremediation, encodes for a large number of signal transduction/chemotaxis proteins. 82 HisKa, 14 with peirplasmic sensor domains; 32 Mcp, 21 with periplasmic sensor domains; 28 GGDEF, 4 with periplasmic sensor domains. We previously determined the structures of two periplasmic sensor domains of chemotaxis receptor proteins that contained a *c*-type heme from *G. sulfureducens* (1) and found that they adopted a PAS-like fold as we had predicted (2). The structure determination by X-ray crystallography for both proteins revealed a novel PAS-like structure formed by two protein chains that involved domain swapping (1). This is the first time that a PAS-like structure was found for a Mcp sensor domain and not a four helix bundle structure found in the aspartate receptor, Tar.

Recently, we are focusing on periplasmic sensor domains of signal transduction proteins from another Gram-negative bacterium of high importance for bioremediation, *Anaeromyxobacter dehalogenans*. *A. dehalogenans* also has a large number of two-component signal transduction molecules: 76 HisKa, 23 with peirplasmic sensor domains; 18 Mcp, 8 with periplasmic sensor domains; 10 GGDEF, none with periplasmic sensor domains. The predicted sensor domains are in the Midwest Center for Structural Genomics (MCSG) pipe line for cloning, protein expression and structure determination.

In an extension of our effort to understand the structure-function of sensor domains of signal transduction proteins in general, we worked with the gram-positive model organism, *Bacillus subtilis*. We used structure-prediction programs to predict that the extra-cytoplasmic sensor domains of HisKa proteins of *B. subtilis* and found that 11 out of 13 have a PAS-like fold (3). This prediction was verified by the determination of the structures of two of these domains as part of the MCSG effort. The determined sensor domain structures showed that our predictions were correct. Because of the large sequence variability of the PAS-like domains are not identified by sequence searches, but we found that the PAS-like fold is a common structural module used by bacteria for detection of its environment (3).

References

1. P.R. Pokkuluri, M. Pessanha, Y.Y. Londer, S.J. Wood, N.E.C. Duke, R. Wilton, T. Catarino, C.A. Salgueiro, M. Schiffer. Structures and Solution Properties of Two Novel Periplasmic Sensor Domains with *c*-Type Heme from Chemotaxis Proteins of *Geobacter sulfurreducens*: Implications for Signal Transduction. *J. Mol. Biol*. 377, 1498-1517 (2008).
2. Y.Y.Londer, I. S. Dementieva, C. A. D'Ausilio, P. R. Pokkuluri, M. Schiffer. Characterization of a *c*-type heme containing PAS sensor domain from *Geobacter sulfurreducens* representing a novel family of periplasmic sensors in *Geobacteraceae* and other bacteria. *FEMS Microbiol. Lett*, 258, 173-181 (2006).
3. C. Chang, C. Tesar, M. Gu, G. Babnigg, A. Joachimiak, P. R. Pokkuluri, H. Szurmant, and M. Schiffer. Extra-cytoplasmic PAS-like Domains are Common In Signal Transduction Proteins. J. Bacteriology, in press.

161

Systems-Level c-di-GMP Signaling in *Shewanella oneidensis* MR-1

Lily Chao, Shauna Rakshe, and **Alfred M. Spormann*** (spormann@stanford.edu)

Depts. of Chemical Engineering, and of Civil and Environmental Engineering, Stanford University, Palo Alto, Calif.

Project Goals: The objective of this research as part of the *Shewanella* Federation, has been to develop a systems-level understanding of c-di-GMP signaling. C-di-GMP

has been recognized as an emerging intracellular prokaryotic signaling nucleotide with perhaps greater implications than c-AMP, and is currently being intensely studied.

Cyclic di-GMP (c-di-GMP) is an important intracellular prokaryotic signaling molecule that is involved in regulating a wide variety of pathways including metabolism, cell division, motility, and pathogenesis. The environmental and cellular factors controlling c-di-GMP signaling are numerous and diverse, but are not well understood. Diguanylate cyclases (DGC) characterized by a 'GGDEF' domain and c-di-GMP-specific phosphodiesterases (PDE) characterized by an 'EAL' domain are known to alter intracellular c-di-GMP concentrations. However, the mechanisms by which these enzymes regulate the cell in response to environmental and cellular conditions are not well understood. In *Shewanella oneidensis* MR-1, many of these enzymes also contain a sensor domain such as the Per-Arnt-Sim (PAS) domain, which can respond to changes in redox potential, oxygen, other small molecular ligands, or light, as well as facilitate protein-protein interactions. Here, we present physiological and biochemical data comparing several PAS-GGDEF-EAL domain proteins present in *Shewanella oneidensis* MR-1. One of these proteins, SO0341, appears to control amino acid metabolism through c-di-GMP signaling. Growth analyses of the deletion mutant in planktonic cultures indicated an extended lag phase compared to that of wild type. Interestingly, growth could be rescued by supplementation of the media, specifically with the amino acids isoleucine, leucine, and valine. Transcriptional microarray analyses of the SO0341 deletion mutant showed that expression of many genes involved in the isoleucine, leucine, and valine biosynthetic pathways are reduced. Enzymatic activity assays of the SO0341 protein indicate that it has both DGC and PDE activity. We have also identified two flavin-binding PAS-GGDEF-EAL proteins, SO0141 and SO3389. Deletion mutants of SO0141 and SO3389 exhibit similar phenotypes in which both strains are less motile under anaerobic growth conditions. Enzymatic activity assays demonstrated that both proteins exhibit DGC and PDE activity. Hence the flavin in these enzymes may be involved in sensing the oxygen or redox status of the environment and modifying the activity of the enzyme accordingly. In contrast to SO0141 and SO3389, a deletion mutant of SO0437 is less motile and exhibits larger biofilms only under aerobic conditions. Furthermore, this phenotype is only apparent under rich medium growth. Our results demonstrate that these enzymes regulate a variety of molecular functions including central metabolism (SO0341) as well as motility (SO0141, SO3389, SO0437).

162

Identification of Genes Involved in Anaerobic Sulfite Reduction by *Shewanella oneidensis* MR-1

Sheetal Shirodkar and **Daad Saffarini*** (daads@uwm.edu)

University of Wisconsin, Milwaukee

Project Goals: One of our goals is to identify components of the electron transport chain that lead to the reduction of sulfur compounds by *S. oneidensis* MR-1. In addition, we aim to identify regulatory proteins that control the reduction of these compounds and elucidate their role in anaerobic respiration.

Shewanella oneidensis MR-1 is a metal reducer that uses a large number of electron acceptors for respiration, including sulfur compounds. Reduction of thiosulfate, polysulfide, and tetrathionate is accomplished using PsrABC that is similar to thiosufate reductases from other bacteria. The mechanism of sulfite reduction by *S. oneidensis* MR-1, however, is poorly understood. We have identified an operon that encodes genes required for sulfite reduction. These include c cytochrome maturation genes required for heme ligation to atypical heme-binding sites. The operon also encodes a c cytochrome and proteins involved in copper transport. Additionally, we have determined components of the electron transport pathway that are needed for sulfite reduction. These include c cytochromes, such as CymA and menaquinones. Analysis of molybdopterin cofactor biosynthesis mutants indicated that sulfite reduction requires a molybdopterin-containing protein. We have also identified a transcriptional regulator, encoded by SO_0490 that regulates sulfite reduction. Mutants deficient in SO_0490 are unable to reduce sulfite. The effect of SO_0490 on the transcription of the genes involved in sulfite reduction, described above, was tested using reverse transcription PCR. The results indicated that SO_0490 does not regulate the expression of these genes. This suggests that additional components are involved in sulfite reduction by *S. oneidensis*. Experiments are underway to identify these proteins and heir role in this process.

163

Investigating Hydrogen Ecology in Marine Microbial Mats Through Integration of Biogeochemistry, (Meta)Transcriptomics and Nano-Scale Imaging

Luke Burow[1,2*] (lburow@stanford.edu), Dagmar Woebken,[1,2] Tori Hoehler,[2] Brad Bebout,[2] Alfred M. Spormann,[1] **Lee Prufert-Bebout**,[2] **Steven W. Singer**,[4] **Jennifer Pett-Ridge**,[3] and **Peter K. Weber**[3]

[1]Dept. of Chemical Engineering and Civil and Environmental, Stanford University, Stanford, Calif.; [2]NASA Ames Research Center, Moffett Field, Calif.; [3]Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory, Livermore, Calif.; and [4]Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, Calif.

Project Goals: The goals of this project are (1) to develop the capability to link identity and function in complex microbial communities and (2) characterize the hydrogen ecology of microbial mats.

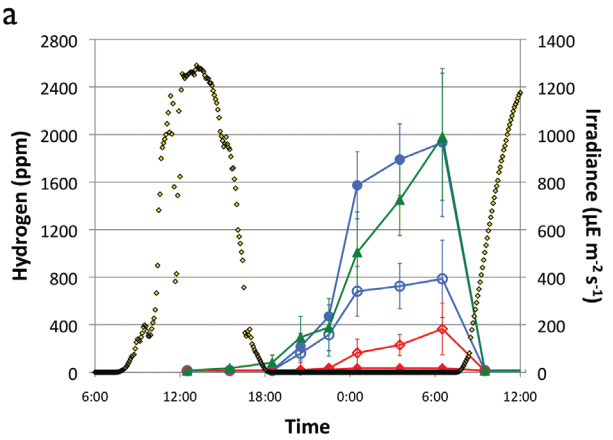
Microbial mats are diverse communities of microorganisms that can produce or consume hydrogen (H_2). Proposed industrial-scale production of biohydrogen and other biofuels will inevitably grow 'unwanted' biofuel-consuming organisms that negatively impact upon biofuel production. Thus, these mat systems represent a valuable model for investigating interactions between microbes that positively and negatively affect biofuel production. We are developing a combination of biogeochemical, molecular and imaging techniques to characterize microbial mat hydrogen ecology.

We have observed H_2 and organic acids to be produced in the upper 2-3 mm layer of mats from Elkhorn Slough, California at concentrations that are orders of magnitude above day levels. Manipulation of the photoperiod by either depriving the mat of light or inhibiting oxygenic photosynthesis using 3- (3,4-dichlorophenyl)-1,1-dimethylurea (DCMU) prior to the H_2-producing dark period indicated that oxygenic phototrophs are likely to be important H_2-producers (Fig. 1a). Spatial localization of H_2-producers and H_2-consumers are a critical determinant of net H_2 production in these systems as disruption of the mat architecture by mechanical homogenization increased H_2 production substantially (Fig. 1a). Addition of molybdate to inhibit sulfate-reducing bacteria (SRB) increased net hydrogen production (Fig. 1a, b) and reduced hydrogen sulfide production suggesting that SRBs are important H_2-consumers in the mats.

Suppression of N_2-fixation using excess ammonium did not decrease net H_2 (Fig. 1b) or organic acid production suggesting that fermentation is not stimulated by a requirement for energetically expensive N_2-fixation and that the main source of H_2 in microbial mats is due to fermentative H_2 production. Although H_2 and organic acid production decreased when the daytime photoperiod was withheld from the mats, it was still substantially higher during the night than during the day. These observations suggest that fermentation is regulated by a circadian rhythm.

Metatranscriptomics approaches are capable of identifying a broad range of genes expressed *in situ* and are exploited in this investigation with a view to capturing key catabolic genes, including fermentative genes related to hydrogen producing pathways. A subtractive rRNA approach using probes complementary to 16S, 23S, 18S and 28S rRNA was used to reduce the high percentage of rRNA often obtained from total RNA isolated from environmental samples. Our method is an adaptation of that currently commercially available from Ambion Inc. (MICROB*Express*). The probe mixture included probes that specifically target cyanobacterial and lower eukaryotic ribosomal sequences in addition to probes already available in the kit. Application of a comprehensive suite of probes reduced rRNA sequences from ~90% to ~40%. In addition to a global metatranscriptomic approach we are using a directed approach that specifically identifies phototrophic and heterotrophic bidirectional hydrogenase (H_2-ase) genes via H_2-ase clone libraries and H_2-ase microarrays. These strategies will provide diversity information of H_2-ases present, which will infer the phylogenetic identity of H_2-producers and/or consumers. Additionally these strategies will be used to evaluate temporal H_2-ase expression dynamics.

The biogeochemical and molecular data we have gathered is enabling nanometer-scale functional and phylogenic imaging by NanoSIMS (NanoSIP-FISH) to identify key organisms involved in fermentation and hydrogen cycling. Daylight incubations with ^{13}C-bicarbonate combined with fluorescence in situ hybridization is being used to identify organisms that fix carbon and generate photosynthate, which is the proposed source of electrons for H_2 evolution. Performing NanoSIP-FISH with samples incubated with ^{13}C-acetate under dark, anoxic conditions will label sulfate-reducing bacteria that consume acetate and hydrogen.



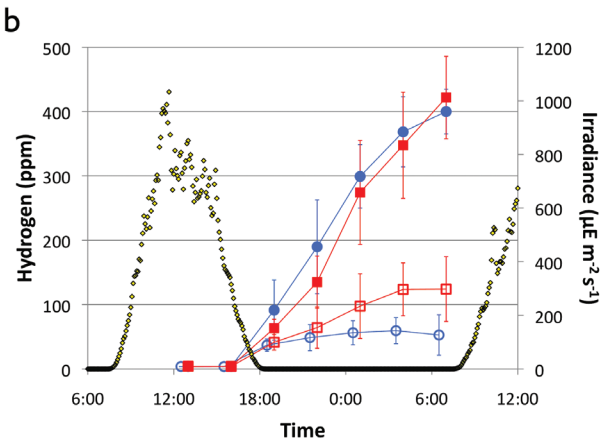


Fig. 1. Net hydrogen production in Elkhorn Slough microbial mats. (a) Effect of photoperiod deprivation (red filled diamonds), 20 μM DCMU (red unfilled diamonds), 30 mM molybdate (filled blue circles) and homogenization (green filled triangles) compared to a control (unfilled blue circles) on H_2 production. (b) Effect of N_2-fixation suppression on H_2 production under control and SRB inhibited conditions. 8.8 mM ammonium added to suppress N_2-fixation (red squares) and compared to N_2-fixing incubations (blue circles). Filled red squares and filled blue circles signify incubations with 30 mM molybdate.

164

NanoSIP: Combining Stable Isotope Probing and High Resolution Secondary Ion Mass Spectrometry to Identify Diazotrophs in Complex, Stratified Microbial Communities

Dagmar Woebken[1,2]* (dwoebeken@stanford.edu), Luke Burow,[1,2] Lee Prufert-Bebout,[2] Brad Bebout,[2] Tori Hoehler,[2] Jennifer Pett-Ridge,[3] Steven W. Singer,[4] Alfred M. Spormann,[1] and **Peter K. Weber**[3]

[1]Dept. of Chemical Engineering and Civil and Environmental, Stanford University, Stanford, Calif.; [2]NASA Ames Research Center, Moffett Field, Calif.; [3]Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory, Livermore,Calif.; [4]Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, Calif.

Project Goals: See below.

Understanding the metabolic processes performed by complex microbial communities requires new methods to observe nutrient fluxes *in situ* and to link the identity and function of microbial community members. Cyanobacterial mats are complex, stratified microbial communities that are often found in coastal marine environments and governed by the diel cycle. During the day, oxygenic photosynthesis accumulates fixed carbon, which is fermented during the night under anoxic conditions. Under these anoxic conditions, these mats evolve large amounts of H_2 as a byproduct of fermentation and N_2 fixation. Therefore, the mats are an excellent model system to understand ecological factors that affect how solar energy is captured, stored in chemical form and then released as a potential biofuel. To link the H_2 ecology of these mats to the flux of carbon and nitrogen, we are developing a new technique, NanoSIP, which combines stable isotope probing (SIP) and nanometer-scale secondary ion mass spectrometry (NanoSIMS) to measure nutrient uptake and assimilation at the single cell level. Previous studies of similar cyanobacterial mats indicated N_2 fixation at high levels and that diazotrophic microbial communities in these mats are comprised of a diverse range of filamentous cyanobacteria, unicellular cyanobacteria and heterotrophic bacteria. The application of ^{15}N-NanoSIP to these mats will allow us to estimate the contributions of individual taxa to the N_2 fixation observed in the sample and correlate ^{15}N incorporation in single cells with *nifH* expression. Additionally, we will be able to correlate carbon and nitrogen flux in samples treated with $^{13}HCO_3^-$ and $^{15}N_2$.

To demonstrate this method, we are studying H_2-evolving cyanobacterial mats collected at Elkhorn Slough in central California. These mats receive substantial inputs of fixed nitrogen from agricultural runoff during the rain season, yet can fix N_2 at high rates during most of the year, suggesting that diazotrophy is an important component of the nitrogen cycle in these mats. Using the acetylene reduction assay, we have identified mat samples dominated by *Microcoleus* spp. that fix N_2 at night and, unexpectedly, mat samples with numerous benthic heterocystous cyanobacteria that fix N_2 primarily during the day. Incubation of these mat samples with $^{15}N_2$ and subsequent analysis by isotope ratio mass spectrometry (IRMS) demonstrated ^{15}N incorporation into biomass. The majority of the label was incorporated into the upper 2mm of the mats, the "green layer" where cyanobacteria are localized. These ^{15}N-labeled samples have been analyzed by NanoSIMS to determine which microbes fix $^{15}N_2$ (Figure 1). Diazotrophic microbes will be identified by linking ^{15}N label observed by NanoSIMS measurements with cell morphology and element labeling using CARD-FISH (Catalyzed Reporter Deposition-Fluorescence *in situ* Hybridization). We use this technique, referred to as EL-FISH, to identify diazotrophic unicellular cyanobacteria and heterotrophic bacteria, which cannot be distinguished by morphology. To complement these single cell techniques we are identifying expressed *nifH* genes recovered from cDNA clone libraries and extracted from metatranscriptomic datasets.

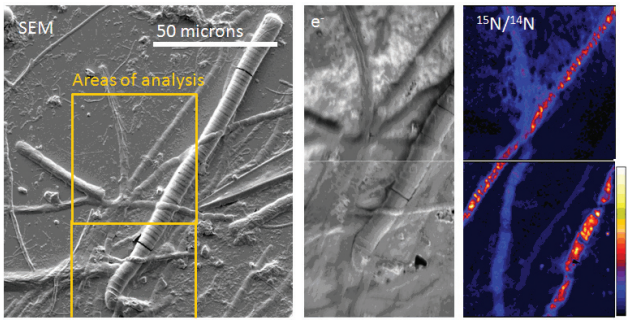


Fig. 1. Scanning electron microscopy (SEM) micrograph and NanoSIMS secondary electron (e-) and nitrogen isotope ratio ($^{15}N/^{14}N$) tiled images of the cyanobacterial community of the Elkhorn Slough microbial mat. The mat was incubated in $^{15}N_2$ for 9 hours during the night prior to sampling. Cyanobacteria that were fixing nitrogen incorporated the ^{15}N tracer. The color scale bar represents $^{15}N/^{14}N$ ratios of 0.003 (black) to 0.09 (white).

165

Population Genomics Reveals a Genetic Basis for Ecological Differentiation in Two Subpopulations of Ocean Bacteria

B. Jesse Shapiro, Jonathan Friedman, Otto X. Cordero, Sarah Preheim, Dana Hunt, Sonia Timberlake, **Martin Polz**, and **Eric Alm*** (ejalm@mit.edu)

Dept. of Civil and Environmental Engineering and Biological Engineering, Massachusetts Institute of Technology, Cambridge

Project Goals: Understand genomic structure, horizontal transfer, and natural diversity in *Prochlorococcus* and *Vibrio* strains

http://proportal.mit.edu/
http://almlab.mit.edu/
http://web.mit.edu/polz/lab/home.html

Microbes adapt to changing selective pressures in their natural environments, leading to a potentially dynamic process of ecological specialization and speciation, even over short periods of time. Yet little is known about the micro-evolutionary processes leading to ecological differentiation of microbial populations in the wild. We sequenced and analyzed complete genomes from 8 closely-related strains of *Vibrio splendidus*, representing two nascent populations, that appear to have recently diversified ecologically: 3 strains found primarily on small particles, and 5 strains found primarily attached to zooplankton. Although gene-flow between populations in the two habitats is common, we observe a significant excess of recent recombination within habitats, suggesting the emergence of ecologically differentiated populations. Gain and loss of DNA is extensive among these strains (each strain contains ~100-300 kb of strain-specific DNA), and a few recently aquired genes may provide habitat-specific adaptive value. For example, a suite of genes involved in O-antigen and mannose-sensitive hemagglutinin (MSHA) biosynthesis are absent in small-particle strains but present in zooplankton-associated strains, perhaps promoting preferential attachment to zooplankton. We also identified a few 'core' genomic regions that, while present in all strains, are highly differentiated between the two habitats. These ecologically associated loci include genes involved in stress response (rpoS), DNA repair (cysteine methyltransferase), and chitin metabolism, leading us to hypothesize that switching between zooplankton-associated (rich in insoluble exoskeletal chitin) and small-particle-associated lifestyles may require fine-tuning of chitin metabolism.

166

Transcriptional-Translational Offsets in the Diel Cell Cycle of *Prochlorococcus*

Jacob R. Waldbauer,[1]* Sébastien Rodrigue,[2] and **Sallie W. Chisholm**[2] (chisholm@mit.edu)

[1]Joint Program in Chemical Oceanography, Massachusetts Institute of Technology and Woods Hole Oceanographic Institution, Cambridge; [2]Dept. of Civil and Environmental Engineering and Dept. of Biology, Massachusetts Institute of Technology, Cambridge

Project Goals: Identify posttranscriptional regulatory processes governing the diel cell cycle of *Prochlorococcus*; link biogeochemical processes to cellular metabolism through quantitative proteomics.

The marine cyanobacterium *Prochlorococcus* is the most abundant oxygenic photosynthetic organism on earth and a key component of oceanic carbon and nutrient cycling. In the ocean, the growth of *Prochlorococcus* cells is tightly coupled to the diel light/dark cycle, and this growth cycle is responsible for primary production on the order of 10 gigatons of carbon per year. Experiments in culture have shown that the mRNA-level expression profiles of most *Prochlorococcus* genes have a strong diel rhythm. We quantified the abundance and periodicity of the *Prochlorococcus* proteome over a diel cycle and compared transcript-level and protein-level expression patterns. Strong diel oscillations in transcript abundance are broadly damped at the protein level, and temporal offsets between the two suggest that posttranslational regulatory mechanisms are important in determining the abundance dynamics of a number of proteins. The overall composition of the proteome is quite stable over the diel cycle, with some proteins consistently among the most abundant in the cell at all times of day, despite major differences in cellular metabolism between light and dark periods. The small variations in protein abundance that accompany significant changes in central metabolic activities imply that *Prochlorococcus* biochemical networks may be poised near balance points that allow for redirection of metabolic fluxes with relatively small shifts in the abundance of their components.

167

Cyanophages Encode and Express the Calvin Cycle Inhibitor CP12 and Pentose Phosphate Pathway Enzymes

Luke R. Thompson, Qinglu Zeng* (qinglu@mit.edu), Libusha Kelly, Katherine Huang, Maureen L. Coleman, and **Sallie W. Chisholm**

Dept. of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge

http://chisholmlab.mit.edu/

Project Goals: Characterize the cellular machinery of *Prochlorococcus* and its viruses as a model system for photosynthetic energy conversion.

Marine cyanophages are known to carry and express genes for the light reactions of photosynthesis. Viral photosynthesis proteins are thought to boost host photosynthesis and help repair damaged photosystems. Notably, no Calvin cycle genes have been reported in cyanophages, suggesting that the products of photosynthetic electron transport are used directly and not stored as glucose. Instead, cyanophages carry genes for the pentose phosphate pathway (PPP), a pathway for glucose oxidation. In one case, the PPP enzyme transaldolase is regulated to be co-expressed with photosynthesis genes, implying that photosynthesis and the PPP operate concurrently in infected cells. These observations led us to the following hypothesis: electron flow through photosynthesis is critical for a productive cyanophage infection, but that energy is not stored as glucose; rather, energy from photosynthesis and from glucose oxidized by the PPP is used to power phage replication. To address this hypothesis, we searched all available cyanophage genomes—and metagenomic sequence of likely viral origin—for any evidence of Calvin cycle and PPP genes. We also measured gene expression of key metabolic genes during infection to address the coordination of transcription. Although no Calvin cycle enzymes were detected, we found widespread incidence of the Calvin cycle inhibitor CP12, particularly in T4-like myoviruses. In cyanobacteria, CP12 binds and inhibits two enzymes in the Calvin cycle, acting as a metabolic switch to direct carbon flux away from the Calvin cycle and toward the pentose phosphate pathway. Three PPP genes were also widely distributed in cyanophages: *zwf* (glucose-6-phosphate dehydrogenase), *gnd* (6-phosphogluconate dehydrogenase), and *talC* (transaldolase). *talC* and *cp12* were the most prevalent of these four genes in cyanophage genomes, followed by *gnd* and *zwf*, and this hierarchy was mirrored in metagenomic databases. PPP, photosynthesis, and DNA biosynthesis genes were co-expressed with known T4-like early genes during myovirus Syn9 infection of *Synechococcus* WH8109. Thus, phage-encoded proteins for all three pathways may play a role early in infection, working in concert. The presence of CP12 and PPP genes and absence of Calvin cycle genes in cyanophages collectively suggests that phage-augmented photosynthesis is used for energy production but not carbon fixation during infection. Rather, glucose is likely oxidized by the PPP, and the NADPH and ribose generated may be used for phage nucleotide production and to relieve oxidative stress induced by infection.

168

Unlocking the Illumina Genetic Analyzer Platform for Improved De Novo Sequencing and Metagenomics

Sébastien Rodrigue* (s_rod@mit.edu), Arne Materna, Sonia Timberlake, Rex R. Malmstrom, **Eric J. Alm**, and **Sallie W. Chisholm**

Dept. of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge

http://chisholmlab.mit.edu/
http://almlab.mit.edu/

Project Goals: Obtain longer sequencing reads from the Illumina Genetic Analyzer platform in order to expand its range of applications to de novo genome assembly and metagenomics while significantly reducing the cost of DNA sequencing projects.

DNA sequencing is one of the most powerful tools for accessing the genetic and metabolic diversity found in microorganisms. The new generation of sequencing instruments allow massive amounts of information to be obtained at an unprecedented speed. Three platforms, the Roche-454 Genome Sequencer, the Illumina Genetic Analyzer, and the Applied Bio-System SOLiD system are currently widely available. The latter two instruments produce a very large number of short reads at a significantly lower cost per basepair. However, their limited readlength creates notorious difficulities in applications such as de novo sequence assembly and metagenomics. Strategies to obtain longer sequencing reads from the Illumina Genetic Analyzer or ABI SOLiD instruments would allow to further exploit their potential, expand their range of applications, and in some cases, significantly reduce the cost associated with DNA sequencing projects.

We describe a simplified procedure to prepare high-throughput sequencing libraries, as well as a strategy to obtain sequencing reads that average 150-210 basepairs long with the Illumina Genetic Analyzer II platform. Following an automatable gel-less library construction, two converging reads are obtained and overlapped to reconstitute a longer composite read. We show that approximately 90% of the paired-end reads can successfully be assembled with this strategy. The approach allowed de novo assembly of single-cell amplified genomes from *Prochlorococcus* cells isolated from the south Pacific ocean. We also demonstrate that the method can be applied to metagenomics sequencing, producing significantly more data at a more affordable cost.

169

Microbial Ecology and Genomics Across Natural Gradients of Light, Nutrients, and Carbon Flow

C.L. Dupont[1]* (cdupont@jcvi.org), Z. Johnson,[2] R.K. Stuart,[1] J. Badger,[1] J. McCrow,[1] L. Zeigler,[1] K. Barbeau,[1] B. Palenik,[3] and **A.E. Allen**[1]

[1]J. Craig Venter Institute, Rockville, Md.; [2]University of Hawaii; and [3]Scripps Institution of Oceanography, La Jolla, Calif.

Project Goals: Compare metagenomic signatures of bacterial and eukaryotic populations of marine microbes across a light and dissolved iron gradient in surface and chlorophyll maximum communities in the Southern California Bight.

The physical, chemical, and biological gradients associated with the transition from the coast to the central Pacific are immense. Offshore of San Diego, water column integrated carbon fixation declines up to 80% within fifteen kilometers. Over the next 700 kilometers, integrated carbon fixation rates fall another one to two orders of magnitude. Nitrate supported primary production, which is likely proportional to carbon export from the euphotic zone, fall a thousand fold over the 700 km. Often thermally stratified, the Southern California Bight also has a subsurface chl a maximum at the nitracline. As such, this area provides a perfect natural laboratory for studying how gradients in light and carbon flow influence microbial community composition and genome content. In the summer of 2007, the gradients in microbial biomass and physiology were characterized over a 700 km oceanographic section leading from coastal San Diego to the edge of the Pacific Gyre. Three distinct oceanographic regimes were identified by their physical and biological characteristics, including light field, depth of thermocline and nitracline, carbon fixation, nitrogen uptake rates, and integrated biomass. For each regime, metagenomic libraries were constructed for both the subsurface chlorophyll maximum and surface microbial communities. Phylogenomic analyses revealed a large enrichment in sequences attributable to viruses and archaea in the subsurface chl max. In general, microbial diversity was higher in the chl max relative to the surface samples. The average bacteria genome size was higher in the subsurface chl max, but also scaled according to cell size and carbon flow. The coastal to open ocean gradient was characterized by an increased abundance in proteins associated with particle adherence, Fe scavenging, and trace metal detoxification.

170

Microbial Metagenomics of the California Current Ecosystem: Southern Upwelling and Northern Oxygen Minimum Zone

Lisa A. Zeigler[1,2]* (lzeigler@jcvi.org), Jonathan H. Badger,[1] John P. McCrow,[1] Ian T. Paulsen,[3] Eric E. Allen,[2] Shannon J. Williamson,[1] and **Andrew E. Allen**[1]

[1]J. Craig Venter Institute, San Diego, Calif.; [2]Scripps Institution of Oceanography, University of California, San Diego; and [3]Macquarie University, Sydney, Australia

Project Goals: To assess microbial community structure and function of the California Current ecosystem using metagenomic sequencing of large volume (200L) water samples. Taxonomic profiles and functional potential are being examined to gain a greater understanding of which organisms are involved in biogeochemical cycling and how they are performing these roles.

Metagenomics has become a powerful tool for enabling the discovery of vast genetic diversity relating to microbial communities; primarily the uncultivated populations. Additionally, the use of metagenomics has led to novel discoveries and new paradigms within marine microbiology. Two datasets taken as a part of the GOS expedition from within the California Current ecosystem are described.

In collaboration with the California Cooperative Oceanic Fisheries Investigations (CalCOFI), which is a partnership between the California Department of Fish and Game, the NOAA Fisheries Service and the Scripps Institution of Oceanography that focuses on the study of the marine environment off the coast of the California, a unique subset of metagenomic samples was collected from the California Current and Southern California Bight (July 2007) to be integrated into the Global Ocean Sampling (GOS) expedition. Exploiting the immense hydrographic metadata acquired by CalCOFI provided a rare opportunity to use metagenomic analyses to describe taxa and metabolism that correspond to the nutrient-rich upwelling waters consistent to this region. These samples are unique to the primarily open ocean sites of the GOS expedition based on their location across a coastal upwelling gradient where carbon, nitrogen, phosphorous and micronutrients are delivered to the pelagic food web. It is our goal to assess the role, through genomics, of different microbial groups in controlling these fluxes. The dataset consists of approx. 4.7Mbp across seven sites from 3 different size ranges, 20-3.0um, 3.0-0.8um, and 0.8-0.1um. These sites span upwelling regions with a relative increase in nutrients and chlorophyll to oligotrophic regions. For taxonomic evaluation of predicted ORFs, we applied a phylogenomic approach using APIS, an automated pipeline that builds trees for every ORF within a dataset. Significant differences in community composition of the microbial populations across the upwelling gradient were revealed. Oligiotrophic sites displayed similar taxa to the GOSI sample dataset, which are typified by a dominance of Alphaproteobacteria and Cyanobacteria, primarily SARII

* Presenting author

and *Prochlorococcus* sp., respectively. Four sites representing regions of higher inorganic nutrient availability displayed clear differences in taxa representation with elevated levels of Gammaproteobacteria, Bacteriodetes, and eukaryotic prasinophytes. One of these upwelling stations was characterized by a Planctomycete bloom. Additional to the taxonomic and nutrient correlation, relationships between functional categories, nutrients, and taxa are being distinguished. Analyses of transporter proteins, for example, suggests variation in substrate utilization between the different water masses; specifically oligotrophic sites appear to encode higher numbers of peptides for ammonium and amino acid transport, whereas, upwelling sites are enriched in nitrate MFS (major facilitator superfamily) and ferrous ion transporters. Ongoing analyses are aimed at linking bacterial and eukaryotic populations present, as well as, further annotation of key functional categories.

Secondly, in collaboration with members of the Coastal Margin Observation and Prediction (CMOP), JCVI researchers (L. Zeigler and A. Allen) have been examining the coastal ecosystem off Oregon pertaining to the Columbia River system (CRS). The CRS affects ecological processes in the Northern California Current System and therefore correlates to other datasets currently being investigated at JCVI. Here four samples were taken that cover freshwater influx into the region to assess the microbial communities; 1) CRE – Columbia River estuary 2) CR plume – Columbia River plume (chlorophyll maxima of ocean just outside plume of freshwater influx) 3) OMZ - 80m sample of anoxic region 4) BTM - 1200m sample of bottom ocean layer. As above, using APIS the following initial taxonomic findings are represented. Briefly, the CRE site has an abundance of actinobacterial sequences and the CR plume waters have an increase in Proteobacteria (primarily "*Candidatus Pelagibacter ubique*"), Euryarchaeota, Bacteroidetes, and dsDNA viral sequences. It appears that the CRE bacterial assemblage is comprised of smaller genomes and the gene pool (particularly in the case of the most dominant group, Actinobacteria) is enriched in functions that encode for processes related to organic nitrogen utilization (xanthine uracil permease, oligopeptide amino acid and urea transporters). Bacterial populations with larger genomes, on the other hand, dominate CR plume waters in the chlorophyll max (e.g., Bacteroidetes) and encode a large repertoire of inorganic nitrogen transporters and a seemingly more diverse complement of signaling genes. Community composition of the OMZ shows a shift from Proteobacteria (primarily *C. Pelagibacter*) to the CFB group (Chlorobi, Flavobacteria and Bacteriodetes) (primarily Polaribacter). Initial screening of functional classifications show a marked increase in the OMZ sample of genes encoding carbamyol phosphate synthase (CPS), as opposed to other GOS environmental datasets. Further investigations are underway to complete functional analyses of this dataset.

171

Phylogenetic Informatics for Indian Ocean Metagenomics

Jonathan H. Badger (jbadger@jcvi.org), John P. McCrow, Chris L. Dupont, Shibu Yooseph, and **Andrew E. Allen**

J. Craig Venter Institute, San Diego, Calif.

http://www.jcvi.org/cms/research/projects/gos

Project Goals: The development of a robust phylogenetic pipeline for the taxonomic binning of metagenomic reads as well as a database and related tools to allow mining of the data in relation to associated metadata such as nutrient levels.

The second phase of the Global Ocean Survey sampling expedition (GOS II), a global circumnavigation survey of ocean surface waters, provided extra metagenomic coverage of sites in the Indian Ocean. These metagenomic reads, which were created by a combination of Sanger and 454-titanium technology, were analyzed by our APIS (Automated Phylogenetic Inference System) pipeline.

APIS uses BLASTP to compare each predicted protein (or protein fragment) against an in-house database of proteins from all completed genomes. The full-length sequences of each matching protein are extracted and aligned, and a bootstrapped neighbor-joining tree of each alignment is created. The trees are automatically analyzed to identify the closest sequence and organism to the query leaf, and summaries over all genes being analyzed are generated.

We have constructed a relational database containing the APIS results as well as available metadata associated with the sampling sites. This database is in the same format as other JCVI metagenomic projects using APIS in order to aid comparison across projects.

Initial analysis of the Indian Ocean data finds that the most prevalent organism by far in the open ocean samples (although it does not comprise a majority) is the SAR11 alphaproteobacterium *Pelagibacter ubique*, an organism known to thrive in low nutrient conditions. Coastal samples show more diversity, as befitting their greater richness of nutrients, and show a greater occurrence of cyanobacteria.

172

Insights and Comparisons of Indian Ocean Metagenomics

John P. McCrow* (jmccrow@jcvi.org), Jonathan Badger, Chris L. DuPont, Shibu Yooseph, and **Andrew E. Allen**

J. Craig Venter Institute, San Diego, Calif.

Project Goals: The development of computational tools to assess and compare marine metagenomic samples on the basis of phylogenetic and functional diversity, and to examine genome level adaptations to resource availability in different size classes of marine bacteria.

http://www.jcvi.org/cms/research/projects/gos

The Global Ocean Survey (GOS) continues to sample and analyze marine microbial life around the world from the Sorcerer II research vessel. Comparisons of microbial populations in the Indian Ocean to those found during GOS phase I have resulted in a better understanding of variation in ecological niches available in marine surface waters.

Improved methods were developed for estimating bacterial genome equivalents from multiple single-copy core genes to normalize genomic samples prior to comparison. Estimation of average bacterial genome size and taxonomic profiling was performed using these same core genes to provide appropriately normalized comparisons between samples. Analyses of gene family distributions in relation to average genome size indicate significant enrichments in transporter abundance in smaller genomes. New methods have been implemented in order to evaluate overall patterns in diversity in relation to system productivity and stability, leveraging independent phylogenetic analysis of all ORFs and of core genes. A subset of sites in the Indian Ocean were selected for metagenomic sequencing of multiple filter sizes of 3.0μm and 0.8μm in addition to the standard 0.1μm filter allowing for comparison between planktonic populations of bacteria and surface attached bacteria unable to pass through a 3.0μm pore.

Several clear trends have emerged; generally, the open ocean, free living or "planktonic" niche is characterized by low overall diversity, small genomes, and disproportionate enrichment for particular genes families such as transporters. Larger size class or attached bacteria and those coming from more productive habitats on the other hand display higher levels of diversity, distinct gene family signatures, and relatively nitrogen rich proteomes.

173

Metagenomic Characterization of Novel *Prochlorococcus* Clades From Iron Depleted Oceanic Regions

Douglas B. Rusch[1]* (drusch@jcvi.org), **Adam Martiny**,[2] **Christopher L. Dupont**,[1] Aaron L. Halpern,[3] and J. Craig Venter[1]

[1]J. Craig Venter Institute, Rockville, Md.; [2]University of California, Davis; and [3]Complete Genomics, Inc., Mountain View, Calif.

Project Goals: We are exploring how marine microbes adapt to variable environmental conditions.

Broadly, we are interested in how marine microbes adapt to different environmental conditions over time and space. This provides insights into the mechanisms driving microbial evolution as well as describing what aspects of the environment are important to life on earth. As part of the Global Ocean Sampling project we have explored microbial life in surface marine waters across the globe. Many of these microbial organisms that dominate this environment are uncultivated and uncharacterized. Identifying and understanding these organisms is important for understanding how these ecosystems behave now and will behave in the future.

Prochlorococcus is the most abundant marine photosynthetic microbe on Earth. It is primarily found in oligotrophic waters across the globe and plays a crucial role in energy and nutrient cycling in the marine ecosystem. The abundance, global distribution, and cultivatability of *Prochlorococcus* have made it a model system for understanding marine microbial diversity and biogeochemical cycling. Analysis of seventy-three metagenomic samples from the Global Ocean Sampling expedition acquired in the Atlantic, Pacific and Indian Oceans revealed the presence of two related but previously unrecognized *Prochlorococcus* clades. A phylogenetic analysis using three different genetic markers places the clades close to known high-light adapted lineages. The two clades consistently co-occur and dominate the surface waters of high temperature, macronutrient replete, low iron regions of the Eastern Pacific Equatorial upwelling and the northern edge of Indian Ocean gyre. These new clades are genetically distinct from each other and other high-light *Prochlorococcus* ecotypes and define a novel ecotype. A detailed genomic analysis indicates that the cells from these clades have adapted to iron-depleted environments by reducing their iron quota through the loss of several iron-containing proteins that likely function as electron sinks in the photosynthetic pathway in other high-light *Prochlorococcus* ecotypes. The presence and inferred physiology of these novel clades may explain why *Prochlorococcus* populations from iron-deplete regions do not respond to iron fertilization experiments and further expand our understanding of how phytoplankton adapt to variations in nutrient availability.

* Presenting author

174

Responses of Soil Microbial Communities to Long Term Elevated CO_2 in Six Terrestrial Ecosystems

Cheryl R. Kuske[1]* (Kuske@lanl.gov), John Dunbar,[1] Gary Xie,[1] Larry Ticknor,[1] Stephanie Eichorst,[1] La Verne Gallegos-Graves,[1] Shannon Silva,[1] Carolyn Weber,[1] Don Zak,[2] Rytas Vilgalys,[3] Chris Schadt,[4] Dave Evans,[5] Patrick Megonigal,[6] Bruce Hungate,[7] Rob Jackson,[3] Andrea Porras-Alfaro,[8] and Susannah Tringe[9]

[1]Bioscience Divison, Los Alamos National Laboratory, Los Alamos, N.M.; [2]University of Michigan, Ann Arbor; [3]Duke University, Durham, N.C.; [4]Oak Ridge National Laboratory, Oak Ridge, Tenn.; [5]Washington State University, Pullman; [6]Smithsonian Environmental Research Center, Edgewater, Md.; [7]Northern Arizona University, Flagstaff; [8]Western Illinois University, Macomb; and [9]DOE Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, Calif.

Project Goals: Our three goals are to: (1) Understand the impacts of long-term elevated CO_2 and other environmental factors (ozone, nitrogen interactions) on the structure and activities of soil microbial communities, at the DOE's six Free Air CO_2 Enrichment (FACE) and Open Top Chamber (OTC) experimental field sites. (2) Establish a multi-tier genomics-based analysis and ecological integration capability that links DOE JGI sequencing technology with the ability to understand functional abilities of soil microbial communities in an ecological setting. (3) Improve our basis for comparison of soil populations involved in carbon cycling and climate change response by expanding the functional genes and proteins we can use to detect and monitor these populations.

Increased plant growth in response to elevated atmospheric CO_2 results in increased carbon inputs to the soil. The collective activities of the complex soil microflora determine whether this additional carbon is sequestered in the soil or released back into the atmosphere. The contributions of soil microbiota to carbon cycling have local, regional and global impacts on terrestrial carbon storage and cycling. However, our current understanding of the composition and activities of microbial biomass, the major processes that represent control points in carbon flux, and the rates at which they occur in terrestrial ecosystems is rudimentary. Accurate climate modeling and carbon management scenarios require an understanding of these soil processes.

For the past ten years, the DOE has operated six large, replicated field experiments (FACE and OTC experiments) designed to test the effects of elevated CO_2 and other factors on terrestrial ecosystems. Our ability to conduct rigorous field comparisons at the 10 yr endpoint of these experiments, and build upon the existing metadata from the sites, provides an unparalleled opportunity to define critical parameters in soil response to climate change variables. For analysis, we have established a collection of replicate soil samples at each of six ecosystem types encompassed by the DOE's FACE and OTC research sites. In some cases, we also have obtained samples across seasons and in multiple years.

Using targeted (rDNA and functional genes) and shotgun metagenomic sequencing strategies, and quantitative PCR methods, we are studying the impacts of over ten years of elevated CO_2 on the soil microbial communities across these six different ecosystems. It is clear from prior research at these sites that soil microflora have responded to the climate change factors and that the populations and mechanisms underlying those responses are likely to be complex. Across the sites, we are attempting to determine if soil community responses are due to changes in total biomass or biomass of major microbial groups, changes in community composition from growth and/or inhibition of specific populations, or changes in respiration and metabolic activity with no change in total biomass or community composition.

Assessment of the bacterial communities in soils across the sites has shown that the bacterial communities have responded to elevated CO_2 in some ecosystems but not in others, and that the nature of the response is specific to that ecosystem. Responses have included changes in relative abundance of the bacterial community relative to the fungal community, as well as changes in community richness and composition. Where multiple factors were included in the field site experiment (e.g. soil depth, plant species, ozone treatment), the observed community responses associated with those factors were often larger than the direct response to elevated CO_2. A similar assessment of the fungal communities has identified very different fungal populations that dominate the different ecosystems, and changes in relative abundance are correlated with elevated CO_2 at some of the sites. In order to facilitate fungal community comparisons, we have developed a new naïve Bayesian classifier to bin fungal LSU sequences into reliable, validated taxonomic units, and are establishing a large sequence training set for fungal LSU sequences. These tools will be made available to the scientific community through a collaboration with the Ribosomal Database Project (Jim Cole, lead).

175

Exploration of Soil Microbial Cellulose Degradation Through a Synthesis of Enzymatic and Targeted Metagenomic Approaches

Cheryl R. Kuske,[1] **Andrew Bradbury**,[1] **Stephanie A. Eichorst**[1]* (seichorst@lanl.gov), Carolyn Weber,[1] Sara D'Angelo,[1] La Verne Gallegos-Graves,[1] Nileena Velappan,[1] Csaba Kiss,[1] Gary Xie,[1] Terri Porter,[2] and Rytas Vilgalys[2]

[1]Bioscience Division, Los Alamos National Laboratory, Los Alamos, N.M. and [2]Duke University, Durham, N.C.

Project Goals: The overall goals are to (a) identify novel cellulolytic enzymes in specific soil bacteria and in soil bacterial metagenome DNA that can be used as molecular markers in field studies, (b) identify dominant, active bacterial and fungal species that metabolize cellulose across different soil types, (c) identify active cellulolytic fungal enzymes in soils under different climate change scenarios.

Through culture-based studies, it is firmly established that many species of soil fungi and bacteria can use complex plant carbon substrates for growth. However, we still know very little about the microbial species capable of using complex plant carbon in soils or the mechanisms through which these substrates are collectively decomposed in the soil. We are studying bacterial and fungal cellulolytic communities through three related experimental efforts. The overall goals are to (a) identify novel cellulolytic enzymes in specific soil bacteria and in soil bacterial metagenome DNA that can be used as molecular markers in field studies, (b) identify dominant, active bacterial and fungal species that metabolize cellulose across different soil types, (c) identify active cellulolytic fungal enzymes in soils under different climate change scenarios.

(a) We are using a phage-display folding-reporter approach to identify novel cellulose-binding domains (CBDs) of proteins in the bacterium, *Clostridium thermocellum*, and in soil DNA enriched in bacterial species capable of degrading cellulose. This approach can be considered to be the protein equivalent of shotgun sequencing, and could be termed the "domainome". The *C. thermocellum* genome is such a rich source of cellulase genes that it provides an excellent model system to experimentally identify novel cellulases or other enzymatic activities linked to CBDs. We have created appropriate display vectors, and have identified suitable cellulose antigens for selection and DNA sources. Large libraries of protein domains from *C. thermocellum* are in the process of being selected against cellulose for the identification of cellulose binding domains. To apply this technique more directly to discovery of cellulolytic enzymes in soil metagenomes, a large phage display library is being generated from a ^{13}C-enriched bacterial DNA fraction from one of the soils in our stable isotope probing experiments (described in the next paragraph).

(b) Using a combination of stable isotope probing and metagenomic analysis of soil microbial communities, we have identified dominant bacterial and fungal species that actively degrade ^{13}C-cellulose in microcosms containing soils that differ in physical and chemical characteristics as well as background organic matter composition. Our results have identified members of the bacterial phyla *Burkholderia* and *Acidobacteria,* and the fungal genera *Arthrobotrys*, *Chaetomium*, and *Trichocladium* as dominant members of ^{13}C-cellulose degrading populations in soils. ^{13}C-enriched DNA fractions were also enriched in a fungal cellobiohydrolase gene. These studies will allow us to correlate the active and dominant microbial species with soil parameters that will guide future field experiments.

(c) Soil fungi are potentially important in mediating carbon cycling in forest soils, both as plant mycorrhizal associates and as primary decomposers of plant litter and belowground plant biomass. We are investigating the combined influence of long term elevated CO_2 and nitrogen fertilization on the biomass and activities of the soil fungal community through gene expression studies of target genes involved in cellulolytic and lignolytic processes and through comparisons of soil metatranscriptomes.

Collectively these studies should expand our knowledge of bacterial and fungal enzymes and genes involved in cellulose degradation in soils, as well as identify species that could dominate these processes in terrestrial carbon cycling.

* Presenting author

Systems Biology Strategies and Technologies for Understanding Microbes, Plants, and Communities

Analytical Strategies for the Study of Plants, Microbes, and Microbial Communities

176

Towards an Understanding of Proteins that Govern the Structure and Function of *Synechocystis sp.* PCC6803 and Multiple *Cyanothece* Strains

Stephen J. Callister,[2]* Jon M. Jacobs,[2] Jana Stöckel,[4] Jason E. McDermott,[3] Lee Ann McCue,[3] Carrie D. Nicora,[2] Ljiljana Paša-Tolić,[1] Louis A. Sherman,[6] **Himadri B. Pakrasi**,[4,5] and **David W. Koppenaal**[1,2] (david.koppenaal@pnl.gov)

[1]Environmental Molecular Sciences Laboratory and [2]Biological and [3]Computational Sciences Division, Pacific Northwest National Laboratory, Richland, Wash.; [4]Depts. of Biology and Energy, [5]Environmental and Chemical Engineering, Washington University, St. Louis, Mo.; and [6]Dept. of Biological Sciences, Purdue University, West Lafayette, Ind.

Project Goals: The primary goal of this project is to apply a systems biology approach to understand the network of genes and proteins that govern the structure and function of *Cyanobacteria*. These microorganisms make significant contributions to harvesting solar energy, planetary carbon sequestration, metal acquisition, and hydrogen production in marine and freshwater ecosystems. *Cyanobacteria* are also model microorganisms for studying the fixation of carbon dioxide through photosynthesis at the biomolecular level. Importantly, this project addresses critical U.S. Department of Energy science needs, provides model microorganisms to apply high-throughput biology and computational modeling, and takes advantage of EMSL's experimental and computational capabilities.

In the initial phase of this project, proteomics characterizations of *Synechocystis sp.* PCC6803 and *Cyanothece sp.* ATCC 51142 were performed to improve our understanding of the proteome makeup of these model organisms under normal and perturbed growth conditions. We have developed the most complete quantitative proteome analysis of *Synechocystis sp.* PCC6803 under various critical environmental perturbations applying a high sensitivity mass spectrometry approach spanning 33 physiological conditions. The resulting proteome dataset consists of 22,318 unique peptides, corresponding to 2,369 unique proteins, covering 65% of the predicted proteins. Quantitative analysis of changes in protein abundance under environmental perturbations has led to the identification of the key proteins required for the maintenance of cellular fitness necessary for cell survival.

We also examined the impact of diurnal rhythms on the protein level of *Cyanothece* 51142. We identified a total of 3,616 proteins with high confidence, which accounts for ~68% of the predicted proteins based on the completely sequenced *Cyanothece* 51142 genome. About 77% of identified proteins could be assigned to functional categories. Quantitative proteome analysis uncovered that ~3% of the proteins exhibit oscillations in their abundance under alternating light-dark conditions. The majority of these cyclic proteins are associated to central intermediary metabolism, photosynthesis as well as biosynthesis of cofactors. Our data also suggest that diurnal changes in activities of several enzymes are mainly controlled by turnover of related cofactors and key players but not entire protein complexes.

While *Cyanothece* sp. ATCC 51142 continues to represent a model organism for proteomics investigations, six additional *Cyanothece* either have finished, or draft genome sequences. This number of strains, having genome sequences, allows for comparison of *Cyanothece* at the level of the core genome and core proteome. While the core genome predicts the common phenotype of *Cyanothece*, the core proteome represents the actual protein phenotype characteristic of *Cyanothece*. As such, all strains were cultured in the laboratory under growth conditions best representing their natural environments and proteins extracted from cells have been analyzed by LC-MS/MS. Results from over 460 LC-MS/MS analyses have been used to develop proteomics databases for strains PCC8801, PCC8802, PCC7424, PCC7425, PCC7822, and ATCC51142. An additional proteomics database for ATCC51472 is pending on the completion of a draft genome sequence for this strain. The Proteomics database for the more distantly related *Synechosystis sp.* PCC6803 was also included in this analysis to better constrain the core proteome to represent *Cyanothece*. While the core proteome of *Cyanothece* is composed of a large percentage of proteins involved in energy production, translation, and amino acid production, a significant portion of the core proteome is also made up of proteins having no predicted function. or only a general assigned function. Of interest was the observation of hypothetical and conserved hypothetical proteins suggesting the importance of these proteins in defining the general lifestyle of *Cyanothece*, yet also suggesting the need for additional functional characterization of these proteins to better understand *Cyanothece* from a systems biology perspective.

The research was performed as part of an EMSL Scientific Grand Challenge project at the W.R. Wiley Environmental Molecular Sciences Laboratory, a national scientific user facility sponsored

by the U.S. Department of Energy's Office of Biological and Environmental Research (BER) program located at Pacific Northwest National Laboratory. PNNL is operated for the Department of Energy by Battelle.

177

Deciphering Microbial Community Dynamics via Observational and Experimental 'Metatranscriptomics': Developments and Applications

Edward F. DeLong[1] (delong@mit.edu), Elizabeth Ottesen[1]* (ottesen@MIT.EDU), Adrian Sharma[1]* (sharma.adrian@gmail.com), Yanmei Shi,[1] Gene Tyson,[1] Frank Stewart,[1] Rex Malmstrom,[1] Sallie W. Chisholm,[1] Jamie Becker,[2] and Dan Repeta[2]

[1]Massachusetts Institute of Technology, Cambridge and [2]Woods Hole Oceanographic Institution, Woods Hole, Mass.

http://openwetware.org/wiki/DeLong_Lab

Project Goals: One of our central project goals is to develop and refine methods for studying microbial community gene expression in the environment, referred to here as 'metatranscriptomics'. There are a variety of diverse applications of these new metatranscriptomic methods. The approach can be used to verify that hypothetical ORFs identified in metagenomic projects are indeed transcribed and expressed. In surveys, 'observational metatranscriptomics' can be used to survey both the nature and abundance of different RNA species (including ribosomal RNA, messenger RNA, and small non-coding RNAs) existing within any given microbial community. Finally, 'experimental metatranscriptomics' can be leveraged to reveal transcriptional responses of microbial communities to environmental perturbation, as well as to discover the specific metabolic pathways involved in matter and energy cycling. In the course of developing methods and protocols for metatranscriptomics, we have explored many of these specific applications. We report here on a few examples, including the refinement of the metatranscriptomic protocols and analyses of biological and technical replicates, the discovery of new microbial small RNAs, and the identification of metabolic pathways involved in the turnover of high molelcular weight dissolved organic matter (DOM) in the marine environment.

Development and quantitative evaluation of an rRNA subtraction protocol

Metatranscriptomes generated by pyrosequencing have great potential use for describing functional processes and attributes of complex microbial communities. Meeting this potential requires protocols that maximize mRNA recovery by reducing the relative abundance of ribosomal RNA, as well as systematic comparisons to identify methodological artifacts and test for reproducibility across datasets. We developed a protocol for subtractive hybridization of small and large subunit RNAs using sample-specific probes, that is applicable across diverse environmental samples. To test the method, rRNA-subtracted and unsubtracted transcriptomes were pyrosequenced from several different bacterioplankton communities ocean, representing ~350 Mbp of metatranscriptomic data. The new subtractive hybridization method reduced bacterial rRNA transcript abundance by 40 to 58%, increasing recovery of non-rRNA sequences up to fourfold. To evaluate this method, we established criteria for detecting sequences replicated artificially via pyrosequencing errors and identified such replicates as a significant component (6 to 39%) of total pyrosequencing reads. Following replicate removal, statistical comparisons of reference genes (identified via BLASTX to NCBI-nr) between technical replicates and between rRNA-subtracted and unsubtracted samples showed low levels of differential transcript abundance (< 0.2% of reference genes). However, gene overlap between datasets was remarkably low, with no two datasets (including duplicate runs from the same pyrosequencing library template) sharing greater than 17% of unique reference genes. These results suggest that current levels of pyrosequencing capture a small subset of total mRNA diversity, underscoring the importance of rRNA subtraction to enhance sequencing coverage across the functional transcript pool.

Novel small RNAs revealed by metatranscriptomics

Previous metatranscriptomic studies have suggested that many cDNA sequences share no significant homology with known peptide sequences, and therefore might represent transcripts from uncharacterized proteins. We found that a large fraction of cDNA sequences detected in a metatranscriptomic datasets are comprised of well-known small RNAs (sRNAs), as well new groups of previously unrecognized putative sRNAs (psRNAs). These psRNAs mapped specifically to intergenic regions of microbial genomes recovered from similar habitats, displayed characteristic conserved secondary structures, and were frequently flanked by genes that suggested potential regulatory functions. Depth-dependent variation of psRNAs generally reflected known depth distributions of broad taxonomic groups, but fine-scale differences in the psRNAs within closely related populations suggested potential roles in niche adaptation. Genome-specific mapping of a subset of psRNAs derived from predominant planktonic species like *Pelagibacter* revealed recently discovered as well as potentially new regulatory elements. Our analyses show that metatranscriptomic datasets can reveal new information about the diversity, taxonomic distribution and abundance of sRNAs in naturally occurring microbial communities, and suggest their involvement in environmentally relevant processes including carbon metabolism and nutrient acquisition.

Metabolic pathways of DOM turnover revealed by metatranscriptomics

To better measure and model the microbial processes associated with the turnover of DOM in the sea, we performed metatranscriptomic analyses in experimental microcosms that were amended with DOM. High molecular weight DOM from surface waters of the North Pacific Subtropical Gyre near station ALOHA was concentrated by ultrafiltra-

tion using a 1 nm pore membrane filter, and added to unfiltered seawater microcosms. The twenty liter microcosms were maintained at *in situ* temperatures and light intensities, and sampled periodically over the course of a 27 hour incubation period. In conjunction with metagenomic datasets obtained at the beginning and end of the experiment, samples for metatranscriptomic analyses were collected over the time course of the experiment in both the unamended control, DOM enriched sample. Subsequent analyses revealed the timing, potential biochemical pathways, microbial species, and potential organic carbon compound intermediates associated with HMW DOM degradation. The results suggested a successional cascade of microbial species related to stepwise metabolic transformations involved in microbially mediated oxidation of DOM in the sea.

178

Identification Characterization of Methanosarcinaceae Cell Surface Proteins

R.P. Gunsalus[1,2*] (robg@microbio.ucla.edu), D. Francoleon,[3] R. Loo,[3] **J. Loo,**[1,3] L. Rohlin,[2] U. Kim,[2] and **M. Arbing**[1]

[1]UCLA-DOE Institute of Genomics and Proteomics, [2]Dept. of Microbiology, Immunology, and Molecular Genetics, and [3]Dept. of Chemistry and Biochemistry, University of California, Los Angeles

http://www.doe-mbi.ucla.edu/people

Project Goals: The goal of this research project is to gain a better understanding of the microorganisms that capture, store and mobilize energy, processes that occur naturally in the Earth's biosphere. We are investigating the molecular biology and biochemistry of several model methane producing archaea and associated hydrogen producing syntrophic bacteria. As participants of anaerobic food chains, they aid in the conversion of complex plant and animal polymers to a variety of small molecular weight carbon intermediates (e.g., alcohols, short chain fatty acids, and various aromatic compounds) to hydrogen, methane and carbon dioxide. However, still lacking is a clear definition of key metabolic pathways, energy conserving complexes, and cell architectures needed to use the above compounds for methane/hydrogen end product formation. We are developing tools to study these anaerobic microorganisms to assess and model hydrogen and methane production. This includes development of community resources for mRNA enrichment/sequencing methods for transcript analysis of model microbes, application of proteomic methods to further define the unique biology, and exploratory metabolomic studies to document the metabolic intermediates in these model microbes. Development and application of such tools will allow better assessment, and modeling of these poorly understood and underutilized hydrogen and methane producing microorganisms for future exploitation.

The cell envelopes of many archaeal species contain a proteinaceous lattice termed the surface layer or S-layer. It is typically composed of only one or two abundant, often post-translationally modified proteins that self-assemble to form a highly organized cell surface-exposed array. Little is known about these proteins in any methanogenic archaean. Surprisingly, over a hundred proteins were annotated to be S-layer or surface associated components in the *Methanosarcina acetivorans* C2A and *Methanosarcina mazei* Gö1 genomes, reflecting limitations of current bioinformatics predictions. To experimentally address what proteins are present, we devised an *in vivo* biotinylation technique to affinity tag all surface-exposed proteins. This overcame several challenges in working with these fragile microorganisms. The two *Methanosarcina* species were adapted to growth under N_2 fixing conditions to minimize the level of free amines that would interfere with the NHS-label acylation chemistry used. A 3-phase separation procedure was then employed to isolate the intact labeled cells from any lysed-cell derived proteins. The Streptavidin affinity enrichment was followed by stringent wash to remove non-specifically bound proteins, and LC-MS-MS methods were employed to identify the labeled surface proteins. In *M. acetivorans* C2A and *M. mazei* Gö1 the major surface layer proteins were identified to be the MA0829 and MM1976 gene products, respectively. Each of the proteins were shown to exist in multiple forms by using SDS-PAGE coupled with glycoprotein-specific staining, and by interaction with the lectin, Concanavalin A. Of the less abundant surface-exposed proteins identified, the presence of all three subunits of the thermosome suggests that the archaeal chaperonin complex is both surface- and cytoplasmically-localized. The above-described techniques provide an alternative strategy to isolate and characterize cell envelope proteins in these archaea.

In related studies we are characterizing the molecular and structural properties of the above surface layer proteins. The *M. acetivorans* MA0829 protein possesses two domains of unknown function that are 78% identical and 86% similar. X-ray crystallography is being used to gain insight into this structure whereby crystallization screening has yielded crystals that diffract to 2.4 A. Structure solution using selenomethioine-labeled protein is in progress. Finally, bioinformatics searches have revealed the distribution of related surface layer proteins in the Methanosarcinaceae and in other archaeal species.

179

Coupling Function to Phylogeny via Single-Cell Phenotyping

Marina G. Kalyuzhnaya[2]* (mkalyuzh@u.washington.edu), Sarah McQuaide,[1] Ekaterina Latypova,[1] Samuel Levine,[1] David Ojala,[1] Michael Konopka,[1] and **Mary E. Lidstrom**[1,2] (lidstrom@u.washington.edu)

[1]Depts. of Chemical Engineering and [2]Microbiology, University of Washington, Seattle

Project Goals: 1) Develop new technology for presorting functional populations and analyze them at the single cell level for both phenotypic and genomic parameters. 2) Apply this approach to populations from Lake Washington sediments to couple functional and genomic datasets at the single cell level.

Rapid advances in modern molecular methods such as the whole genome community sequencing (WGCS) approach open new ways to study microbial ecology. While application of the high-throughput sequencing could result in a blueprint of genomic content of the ecosystem of interest, generally it provides little information about ecological significance of the newly detected functions. To truly understand the role of microbes in the environment, the genomic sequences should be reconsidered in the context of physiological data. Integration of single-cell physiological measurements with genomic data in order to elucidate the functional role of yet uncultivable microbes is the major focus of our current research**.** Overview of our approaches is presented in Figure 1. We use respiration as a core metabolic function to describe methylotrophic capabilities of microbial cells from two natural environments: freshwater lake sediment (Lake Washington) and salt water (Saanich Inlet). Subsequently, cells tested positive for specific functions are targeted for further genomic explorations via whole genome amplification, PCR-surveys for functional genes and whole genome sequencing.

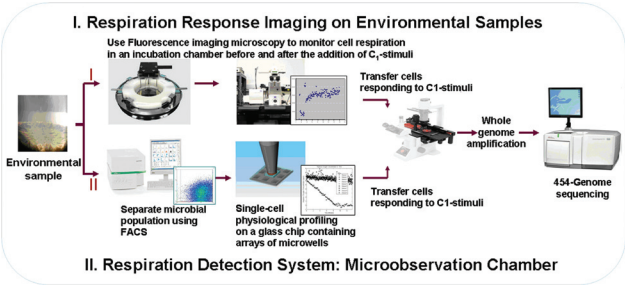


Figure 1. Proposed approaches to couple physiological function with genomics

This single cell platform is a new way to uncover or refine the function for yet uncultured members of natural microbial communities.

180

Proteomics Driven Analysis of Microbes, Plants and Microbial Communities

Mary S. Lipton[1]* (mary.lipton@pnl.gov), Stephen J. Callister,[1] Kristin E. Burnum,[1] Roslyn N. Brown,[1] Haizhen Zhang,[1] Thomas Taverner,[1] Margaret F. Romine,[1] Kim K. Hixson,[2] Janani Shutthanandan,[1] Samuel O. Purvine,[2] Angela D. Norbeck,[1] Matthew E. Monroe,[1] Carrie D. Nicora,[1] Gordon A. Anderson,[1] Steve Giovannoni,[3] Jerry Tuscan,[4] Phil Hugenholtz,[5] Stuart Levy,[6] Derek Lovley,[7] and **Richard D. Smith**[1]

[1]Biological Sciences Division and [2]Environmental Molecular Sciences Laboratory; Pacific Northwest National Laboratory, Richland, Wash.; [3]Oregon State University, Corvallis; [4]Oak Ridge National Laboratory, Oak Ridge, Tenn.; [5]DOE Joint Genome Institute, Walnut Creek, Calif.; [6]Tufts University, Boston, Mass.; and [7]University of Massachusetts, Amherst

Project Goals: This project employs comprehensive global and directed proteomic analyses of microbes, plants and microbial communities to enhance the scientific understanding through improved annotation of genomic sequences, elucidation of phenotypic relationships between environmentally important microorganisms, characterization of higher organisms, characterization of the metabolic activities within microbial communities, and identification of post-translationally modified proteins.

Inherent to exploiting microbial function or utilizing plants as biofuels is the detailed understanding of the physiology of the cell. These cellular functions are dictated by the proteins expressed in the cell, their localization and their modification state. This project exploits the technological and informatics advances in the proteomics pipeline at PNNL (as described in the poster by Anderson et al) to address organism-specific scientific objectives developed in conjunction with biological experts for a number of different microbes and plants. In our poster, we highlight the ability to use proteomics data for genome annotation of microbes and fungi, characterization of microbial communities, advances in the characterization of protein phosphorylation state, and the identification of new proteins important to photosynthesis, and the determination of protein localization in stem, root and leaf tissues of poplar.

Genome sequences are annotated by computational prediction of coding sequences, followed by similarity searches such as BLAST, which provide a layer of (possible) functional information. While the existence of processes such as alternative splicing complicates matters for eukaryote genomes, the view of bacterial genomes as a linear series of closely spaced genes leads to the assumption that computational annotations which predict such arrangements completely describe the coding capacity of bacterial genomes. However, proteomic experiments have shown the expression in *Pseudomonas fluorescens* Pf0-1 of sixteen non-annotated

protein-coding regions, of which **nine were antisense to predicted genes**, six were intergenic, and one read in the same direction as an annotated gene but in a different frame. The expression of all but one of the newly discovered genes was verified by RT-PCR. Few clues as to the function of the new genes were gleaned from informatic analyses, but potential orthologs in other *Pseudomonas* genomes were identified for eight of the new genes. The 16 newly identified genes improve the quality of the Pf0-1 genome annotation, and the detection of antisense protein-coding genes indicates the under-appreciated complexity of bacterial genome organization.

Unique to proteomic studies, the elucidation of post-translational modifications and protein localization lead to a richer understanding of the biological system. We have developed advanced technologies to fractionate proteins from microbial subcellular fractions and have applied this technology to mixtures of dissimilar microbes. The applications of this technology to microbial communities will result in a reduction of the sample complexity and increase characterization of the community. Heme moieties play an important role in microbial respiration, yet to date have remain recalcitrant to proteomic characterization. Application of refined separation strategies have resulted in samples enriched in heme containing proteins and thus aid in the identification of these proteins.

Proteomics characterization is also used to understand more complex systems such as plants and microbial communities. *Populus* is the fastest growing tree species in North America and has been identified as a potentially important crop species for converting plant biomass to liquid fuels. *Populus* species are broadly adapted to nearly all regions of the U.S., and hybrid clones have demonstrated 10 dry tons per acre productivity on a commercial scale. Still, improvements in growth rate, cell wall composition, drought tolerance, and pest resistance are required before this species reaches it potential as an energy crop. We have used proteomics technologies to map the protein expression patterns between root, leaf and stem tissues.

Termites degrade and thrive on lignocellulose with help from the bacterial microbiome harbored within their guts. Recent metagenomic analyses have begun to shed light on the genetic potential of the termite hindgut community, but little is known about which genes are expressed to support the symbiotic relationship. Here, we analyzed the metaproteome of the bacterial community resident in the hindgut paunch of the wood-feeding 'higher' *Nasutitermes* species and identified 886 proteins, 197 of which have known enzymatic function. Using these enzymes, we reconstructed known metabolic pathways to gain a better understanding of carbohydrate transport and metabolism, nitrogen fixation and assimilation, energy production, and amino acid synthesis in this endosymbiotic microbiome.

Additional information and supplementary material can be found at the PNNL proteomics website at http://ober-proteomics.pnl.gov/

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

181

Advances for High Throughput, Comprehensive and Quantitative Proteomics and Metabolomics Measurements; Enabling Systems Biology

Gordon A. Anderson*, Ronald J. Moore, David J. Anderson, Kenneth J. Auberry, Mikhail E. Belov, Kevin Crowell, Stephen J. Callister, Therese R.W. Clauss, Kim K. Hixson, Gary R. Kiebel, Brian L. LaMarche, Mary S. Lipton, Da Meng, Thomas O. Metz, Matthew E. Monroe, Heather M. Mottaz, Carrie D. Nicora, Angela D. Norbeck, Daniel Lopez-Ferrer, Daniel J. Orton, Ljiljana Paša-Tolić, David C. Prior, Samuel O. Purvine, Anuj Shah, Yufeng Shen, Anil K. Shukla, Mudita Singhal, Gordon W. Slysz, Aleksey V. Tolmachev, Nikola Tolić, Karl Weitz, Aaron Wright, Rui Zhang, Rui Zhao, and Richard D. Smith (rds@pnl.gov)

Biological Sciences Division and Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, Wash.

Project Goals: This project is developing and applying new and greatly expanded quantitative, high throughput proteomics and metabolomics capabilities for studying diverse microbial systems and communities, plants, and ecosystems of increasing levels of complexity. Capability advances this past year centered on extending modified protein coverage through the integration of bottom-up and top-down measurements, the development of multiplexed activity-based proteomics, developing broad metabolome measurement coverage, and the integration of proteomics and metabolomics measurements with genomic information. In concert with other measurements and information, these developments are addressing deficiencies in the coverage of biochemical components provided by current measurement capabilities, and thus providing the detailed data and the quality needed to enable truly effective systems biology approaches. FWP number: 40601

Understanding microbial and bioenergy-related systems requires knowledge of the array of proteins and their complement of post-translational modifications, as well as knowledge of the large (and often unknown) range of metabolites and other cellular components. Among the basic challenges associated with gaining this understanding are identifying and quantifying large sets of proteins, modified proteins, and metabolites whose relative abundances

typically span many orders of magnitude, and doing so in a sufficiently high throughput manner. Compounding these analytical challenges is the largely unknown extent and nature of many protein modifications, and the high chemical and structural diversity of metabolites.

PNNL is developing and applying high throughput mass spectrometry-based measurement technologies and associated informatics tools applicable to a broad range of biological studies, many of which are presently conducted in collaboration with a number of BER Genomic Sciences researchers (see poster by M.S. Lipton et al.). Our high throughput proteomics/metabolomics analysis pipeline is based on high resolution nano-liquid chromatography separations combined with high mass measurement accuracy mass spectrometry measurements. This poster highlights several developments that build upon this foundation:

1. Providing much broader protein coverage, including modification states to which current measurements are effectively 'blind'. We developed new approaches that combine top-down and bottom-up measurements to extend quantitative proteome coverage to a large range of protein modification states, and that integrate measurements from targeted post-translationally modified sub-proteomes. As part of these efforts, we combined a "RePlay" chromatography method for on-line reanalysis of the separated proteome components with an ultra-fast post-column pressure digestion system to attain nearly continuous mass spectrometer utilization and simultaneous acquisition of both top-down and bottom-up proteomics data from a single analysis, which avoids many of the present ambiguities associated with data interpretation. In conjunction with this approach, we are exploiting the increased throughput of bottom-up measurements (see below, and poster by R.D. Smith et al.) to provide detailed measurements for targeted sub-proteomes. We also are commencing development of new informatics approaches to integrate these complementary data sets.

2. Multiplexed activity-based proteomics. To augment the more detailed proteomics measurements noted above, we are implementing measurements that directly measure enzyme *activities* rather than abundances, and thus measurements that account for changes in protein modification, structure, localization etc. The approach involves the synthesis of *in vitro or in vivo* multiplexed (for different activities) and isotopically coded chemical probes that can be applied simultaneously to discover and quantitatively follow enzymatic activities. The approach allows isolation, enrichment, and analysis of large sets of labeled 'signature' peptides that in turn enable protein identification, as well as provide direct and quantitative data on a large range of biological activities in any targeted biological system.

3. Increasing '-omics' coverage by broad nanoLC measurements of the metabolome. We have adapted the new high throughput nanoLC-ion mobility-MS platforms noted below to obtain broad and quantitative measurements of the broad range of metabolites and small molecule components of biological systems. These measurements provide much more comprehensive data sets, which are needed to support the development of computational models for biological systems and effective systems biology approaches.

A new fast separation liquid chromatography-ion mobility-mass spectrometry platform has been developed (and several versions now implemented) that benefits all of the above efforts by providing high levels of data quality in conjunction with an order of magnitude increase in measurement throughput (see poster by R.D. Smith et al.). The information garnered from improved global coverage of protein modifications and metabolites, and obtainable with increased throughput, is expected to have a profound impact on our ability to develop computational models of biological systems.

Gaining the full benefits of these extended measurement capabilities requires a significantly different and extended computational infrastructure. Thus, we have expanded the PNNL informatics pipeline to incorporate a suite of data analysis tools, data consolidation applications, and statistical packages, as well as visualization software for data interpretation. Through the development of these new tools and the enhancement of existing tools, we have implemented a framework that will support integration of the enhanced proteomics and metabolomics data sets. This framework further supports integration of genomics data from public repositories and provide the needed infrastructure to interoperate with the GTL Knowledgebase.

This research is supported by the Office of Biological and Environmental Research of the U.S. Department of Energy. Pacific Northwest National Laboratory is operated for the U.S. Department of Energy by Battelle Memorial Institute through Contract No. DE-AC05-76RLO 1830.

182

A New Platform for Much Higher Throughput, Comprehensive, and Quantitative Proteomics and Metabolomics Measurements and Data Analysis

Richard D. Smith* (rds@pnl.gov), **Gordon A. Anderson**, Erin S. Baker, Mikhail E. Belov, Eric Y. Choi, Kevin L. Crowell, William F. Danielson III, Yehia M. Ibrahim, Ryan Kelly, Brian L. LaMarche, Andrei V. Liyu, Divakara Meka, Da Meng, Matthew E. Monroe, Daniel J. Orton, Jason Page, David C. Prior, Thomas A. Seim, Anuj Shah, Gordon W. Slysz, Keqi Tang, and Aleksey V. Tolmachev

Biological Sciences Division and Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, Wash.

Project Goals: This project aims to develop and apply greatly improved capabilities for proteomics and metabolomics. In this context, we have developed and demonstrated a new fast separation liquid chromatography/ion

mobility/mass spectrometry platform for quantitative high throughput proteomics and metabolomics measurements that achieves high levels of data quality in conjunction with an order of magnitude increase in measurement throughput. In addition to significantly improved sensitivity and extremely large data generation rates, the new platform provides the basis for effectively generating and combining data from multiple measurements to attain broader coverage of both protein modification states and chemically diverse metabolomes. The platform considerably speeds large-scale applications and thus enables previously impractical studies, e.g., of diverse microbial systems, communities, and ecosystems. FWP number: 40601

A major challenge underlying the successful development and application of systems biology approaches is the large numbers of measurements needed to accommodate experimental constraints, e.g., derived from available sample sizes, variability in measurements, and/or practical measurement throughput limitations. Although much greater than feasible with classical approaches (using 2D-PAGE), the proteomics measurement throughput now provided by LC-MS- and LC-MS/MS-based approaches is grossly inadequate for characterizing very large numbers of samples, e.g., involving many perturbations, or spatially and/or temporally distinct samples. Metabolomics is similarly constrained, and faces additional challenges because of the broad chemical diversity of metabolites and the greater difficulties associated with identification. Additionally, sample recovery and enrichment methodologies can limit proteome and metabolome coverages.

To address these shortcomings, we developed a new platform at PNNL that demonstrates greatly improved measurement throughput, sensitivity, robustness, and quantitative capability for proteomics and metabolomics measurements in a range of biological research applications.

The new measurement platform incorporates fast multiplexed nanocapillary LC separations coupled via a greatly improved electrospray ionization interface to an ion mobility spectrometer (IMS) stage interfaced to a high speed, accurate mass, and broad dynamic range time-of-flight mass spectrometer (TOF/MS). The automated fast nanocapillary LC system incorporates high pressure LC pumps, an autosampler, and a multiplexed 4-column fluidics system. Each 10-cm-long capillary LC column is operated at 10,000 psi to provide both fast and high resolution separations. Electrospray ionization (ESI) generated ions are accumulated at the end of the second stage of a dual electrodynamic ion funnel trap before being injected into an IMS separation drift tube stage where peptide or metabolite ion separations occur on a time scale of <50 msec. To increase IMS-TOF/MS sensitivity, we developed a novel multiplexing approach that increases the number of ion injection pulses into the IMS separation stage by >30-fold, and thus the S/N levels achievable in a given analysis time, without any loss of separation or MS data quality. Downstream of the IMS separation drift tube, spatially dispersed ion packets are efficiently collected by another electrodynamic ion funnel and transferred for analysis to an orthogonal acceleration TOF/MS analyzer stage. A high-performance data acquisition system based on a high speed analog-to-digital converter developed to ensure high mass accuracy, high dynamic range measurements is being used in conjunction with a real-time multi-dimensional spectral averaging capability developed under a new CRADA with Agilent Technologies.

Detailed evaluation of the LC-IMS-TOF/MS platform has confirmed significantly improved performance compared to the best currently available proteomics platforms. The new platform provides more than an order of magnitude increase in data generation rates, and initial studies confirm more than an order of magnitude improvement in sensitivity, as well as lower limits of detection. Further improvements in performance are expected from the use a more intense ion source based upon an advanced ESI multi-emitter design used in conjunction with a dual stage ion funnel interface. These advances are being complemented by the development of a new informatics pipeline for rapidly processing and analyzing the greatly expanded data volumes. In combination with improved informatics tools, application of the new platform is expected to enable much more comprehensive coverage of proteins (and e.g., modified proteins) and chemically diverse metabolites

This research is supported by the U.S. Department of Energy Office of Biological and Environmental Research at Pacific Northwest National Laboratory (operated for the U.S. Department of Energy by Battelle through Contract No. DE-AC05-76RLO 1830).

183

Stable Isotope Probing of RNA Combining Phylogenetic Microarrays and High Resolution Secondary Ion Mass Spectrometry to Link Composition and Function in Microbial Systems

Xavier Mayali[1]* (mayali1@llnl.gov), Peter K. Weber,[1] Eoin L. Brodie,[2] Todd Z. DeSantis,[2] Ulas Karaoz,[2] Gary L. Andersen,[2] Meredith M. Blackwell,[3] Stephanie R. Gross,[3] Shalini Mabery,[1] Paul D. Hoeprich,[1] Ian D. Hutcheon,[1] and **Jennifer Pett-Ridge**[1]

[1]Lawrence Livermore National Laboratory, Livermore, Calif.; [2]Lawrence Berkeley National Laboratory, Berkeley, Calif.; and [3]Lousiana State University, Baton Rouge

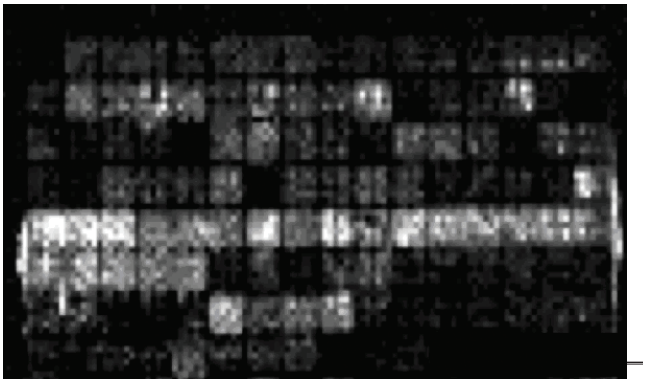
Project Goals: To develop nano-scale stable isotope probing, complex community meta-transcriptomic analysis, and the translation of genome scale data into biogeochemical and metabolic flux network models.

A fundamental goal in microbial ecology is to understand the biogeochemical role of individual microbial taxa in their natural habitat. This rather simple concept is in actuality a complex problem because 1) most microbes remain uncul-

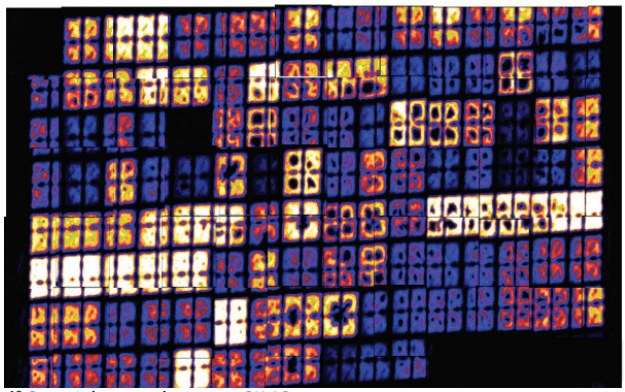
tivated and 2) the majority of microbial communities are very diverse. The former makes the direct testing of isolated strains for biogeochemical activity a limited approach. The latter impacts culture-independent methods like metagenomics as many biogeochemical processes cannot be directly inferred from sequence data alone, even when assembly of complete genomes is possible.

Our approach to this problem (Chip-SIP) involves the combination of high-density phylogenetic microarrays ("chips") and stable isotope probing (SIP) to directly link identity and function. Microbial communities are incubated in the presence of stable isotope-enriched substrate (s), RNA is extracted and hybridized onto a microarray synthesized on a conductive surface, and the array is imaged using high resolution secondary ion mass spectrometry (SIMS) with a Cameca NanoSIMS 50 to detect isotopic enrichment. We have successfully validated this approach utilizing RNA from a single pure culture with varying degrees of isotopic enrichment using two different substrates (^{15}N-labeled ammonium and ^{13}C-labeled glucose). We show that isotopic enrichment of individual probe spots as detected by nanoSIMS is positively correlated with fluorescence as detected by a traditional microarray scanner (figure 1). This allows the relationship between hybridization efficiency and relative isotopic enrichment to be determined. Further, we have successfully detected ^{15}N enrichment in an estuarine bacterial community following incubation with ^{15}N-NH_4, demonstrating the utility of the method in mixed natural communities.

Current efforts are aimed at elucidating the major players in nitrogen fixation and carbon transformation in the gut of the wood-eating passalid beetle *Odontotaenius disjunctus*. The gut of this organism is spatially segregated into at least 4 distinct compartments (foregut, midgut, anterior hindgut, and posterior hindgut; see figure 2) each differing physically, chemically and microbiologically. We hypothesize that the sequential biogeochemical activities required to derive energy from lignocellulosic materials are partitioned across the gut sections. We are employing the Chip-SIP approach to determine which organisms at each gut location are involved in these processes.



Fluorescence by microarray scanner



^{13}C enrichment by nanoSIMS

Figure 1: Visual comparison of ^{13}C-enriched RNA from *Pseudomonas stutzeri* hybridized to an array comprised of *Pseudomonas*-specific probes, showing corresponding signal between fluorescence (top) and ^{13}C isotopic enrichment by nanoSIMS (bottom)

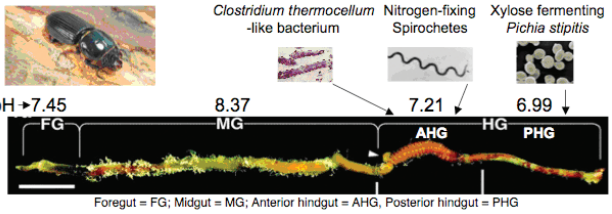


Figure 2: Dissected *Odontotaenius disjunctus* beetle gut showing the 4 sections (foregut, midgut, anterior hindgut and posterior hindgut); adapted from Nardi et al., Arthropod Struct. Dev. 35 (2006) 57-68. Scale-bar = 10 mm

184

Synthetic Genomics: Progress on Construction of a Synthetic Bacterial Cell

John Glass* (jglass@jcvi.org)

Synthetic Biology Group, J. Craig Venter Institute, Rockville, Md.

Project Goals: Synthesize a minimal mycoplasma genome that has all of the machinery for independent life. Our goal in this aspect of the program is to create a minimal bacterial cell based on *Mycoplasma genitalium*, which has the smallest genome of any bacterial cell that can be grown in pure culture. A minimal cell contains only essential genes and can be grown in pure culture under defined conditions. It lacks synthetic capacity for small molecules or metabolites that can be supplied in the medium. Thus it is stripped down to core functions for macromolecular synthesis and cell division. The rational for this is that through creation and analysis of a cell with perhaps fewer than 400 protein coding genes we will be better able to learn the first principals of cellular life. Such a cell would have less than one tenth as many genes as *Escherichia coli* and the lack of complexity would enable an uncluttered perspective on how cells work.

* Presenting author

Bacteria and yeast have been widely used as hosts for cloning segments of DNA from a variety of organisms. Cloning of large DNA segments is limited by size and toxicity to the host. Reports of *Escherichia coli* DNA clones larger than three hundred kilobases have been infrequent, whereas yeast has been commonly used to clone megabase-sized DNA. We cloned whole bacterial genomes from *Mycoplasma genitalium*, *Mycoplasma pneumoniae*, and *Mycoplasma mycoides* as circular centromeric plasmids in yeast. Once cloned, the bacterial genomes can be modified using efficient, well-established methods for DNA manipulation in yeast. Next, for one of those cloned genomes, *M. mycoides*, we introduced such modified genomes back into a different bacterial cell, *Mycoplasma capricolum*, by genome transplantation. To do this, restriction barriers had to be overcome. These methods should enable future transplantation of a synthetic genome, and also facilitate the engineering of bacteria with poorly developed genetic systems.

185

Characterization of Microbial Strains Important in Biofuels and Biomass Conversion

Christopher L. Hemme,[1]* Matthew W. Fields,[2] Qiang He,[3,4] Ye Deng,[1] Lu Lin,[1,5] Qichao Tu,[1] Housna Mouttaki,[1] Xueyang Feng,[6] Zheng Zuo,[6] B.D. Ramsey,[2] Zhili He,[1] Kerrie Barry,[7] Elizabeth Saunders,[8] Hui Sun,[7] Miriam Land,[9] Yun-Juan Chang,[9] Liyou Wu,[1] Joy Van Nostrand,[1] Loren Hauser,[9] Alla Lapidus,[7] Cliff S. Han,[8] Jian Xu,[5] Yinjie Tang,[6] Juergen Wiegel,[10] Tommy J. Phelps,[11] Eddy Rubin,[7] and **Jizhong Zhou**[1,12]

[1]Institute for Environmental Genomics, University of Oklahoma, Norman; [2]Dept. of Microbiology, Montana State University, Bozeman; [3]Dept. of Civil and Environmental Engineering and [4]Center for Environmental Technology, University of Tennessee, Knoxville; [5]Qingdao Institute of BioEnergy and Bioprocess Technology, Chinese Academy of Sciences, Qingdao, China; [6]Dept. of Energy, Environmental and Chemical Engineering, Washington University, St. Louis, Mo.; [7]DOE Joint Genome Institute, Walnut Creek, Calif.; [8]DOE Joint Genome Institute, Los Alamos National Laboratory, Los Alamos, N.M.; [9]Genome Analysis and Systems Modeling Group, Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tenn.; [10]Dept. of Microbiology, University of Georgia, Athens; [11]Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, Tenn.; [12]Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, Calif.

Project Goals: To explore the genomic and physiological properties of stable bacterial co-cultures important for biomass conversion and biofuels production in consolidated bioprocessing schemes.

Comparative Genomics of *Clostridia*

Genomic sequencing of 20+ *Clostridia* strains related to biofuels production and biomass conversion were sequenced, including multiple strains from Cluster III thermophilic and mesophilic cellulolytic *Clostridium* species and multiple strains of saccharolytic *Thermoanaerobacter* species. This dataset represents a significant improvement in the genomic knowledge base of bacteria important to biofuels production. The genomes of two strains of *Thermoanaerobacter*, *T. pseudethanolicus* 39E and *Thermoanaerobacter* sp. X514, have been finished and comparative genomics analysis has been conducted. Experimental studies have shown that when either of these *Thermoanaerobacter* strains are grown in coculture with *Clostridium thermocellum* LQRI, cellulose degradation rates and ethanol production yields are increased compared to the *C. thermocellum* monoculture. Furthermore, this effect is greater for the X514 co-culture compared to the 39E co-culture. Comparative genomics and experimental analysis revealed several potential mechanisms by which such physiological effects could manifest. First, it was noted that X514 encodes a complete *de novo* Vitamin B_{12} biosynthesis operon whereas 39E encodes only a partial operon. Experimental analysis showed that X514 monocultures are largely insensitive to the addition of exogenous B_{12}, while ethanol yields are severely impacted in 39E monocultures when no exogenous B_{12} is added. This effect can be alleviated in 39E when 2-3X B_{12} is added to the culture. This effect is magnified in coculture with *C. thermocellum*, suggesting that X514 synthesizes its own B_{12} while 39E does not and that B_{12} is a critical nutrient in determining ethanol yields. Metabolic flux analysis also revealed that absolute flux through the central carbon metabolism pathway is greater in X514 than in 39E. Finally, 39E and X514 encode distinctly different xylose uptake systems and the X514 genes in general are more highly expressed under xylose growth conditions compared to 39E. Thus, substrate uptake, metabolic flux rates and vitamin synthesis likely contribute greater to variable ethanol production and cellulose degradation rates in *Thermoanaerobacter-C. thermocellum* co-cultures.

Hydrogen Production from *Desulfovibrio vulgaris-Clostridium cellulolyticum* Cocultures

Experimental analysis shows that the stable coculture of *D. vulgaris-C. cellulolyticum* produces significantly higher concentrations of molecular hydrogen when grown on cellulose compared to *C. cellulolyticum* monocultures. Functional genomic and experimental analyses were conducted to identify the mechanisms behind this observation. SEM images suggest that cellular binding to cellulose fibers is greater in the coculture than in the *C. cellulolyticum* monoculture, suggesting that the addition of *D. vulgaris* increases binding to the cellulose fibers and may in turn increase cellulose degradation rates. Preliminary microarray analysis also shows that *C. cellulolyticum* cellulosome genes are more highly expressed in co-culture than in monoculture as well as NiFe-hydrogenase genes and other genes related to hydrogen production.

Transcriptional Profiles of X514 at Different Carbon Substrates

The transcriptional profiles of *Thermoanaerobacter* sp. X514 at different carbon substrates have been conducted. Experimental studies show that X514 is able to metabolize

hexose (glucose, fructose, ribose, galactose and so on), pentose monosaccharides (including xylose) and some complex carbohydrates (sucrose, cellobiose, starch). When X514 are grown in glucose, xylose, fructose and cellobiose, the corresponding genes in carbon uptake system are more highly expressed. Moreover, X514 metabolized these four sugars by Embden-Meyerhof-Parnas (EMP) pathway and pentose phosphate (PPP) pathway. X514 encodes carbohydrate active enzymes for catabolism of fructose, xylose and cellobiose. In contrast to glucose metabolism, growth on fructose, xylose and cellobiose resulted in upregulation of carbohydrate metabolism genes which shift carbon fluxes head towards ribose. These observations suggest that when X514 is grown on fructose, xylose and cellobiose, more ribose should be synthesized as the substrate of nucleotide and amino acid metabolism. Experimental analysis shows that energy metabolism of X514 on fructose is more active than that on other sugars, with higher concentrations of ethanol, acetate and lactate generated. Furthermore, the V-type ATPase genes and a large number of genes involved in inorganic ion transport and metabolism (such as sodium-translocating decarboxylase enzyme genes, Na^+/H^+ antiporter and sodium/hydrogen exchanger genes and so on) were significantly up-regulated. The data indicate that under fructose growth conditions, electrochemical ion gradient at the cytoplasmic membrane is much more actively established than when grown on other sugars. Thus, more ATP should be generated under these conditions. For the alcohol generation, the results show the three characterized *adh* genes are all expressed at similar levels when grown on these four sugars. But for the additional six lineage-specific *adh* genes, the expression levels greatly varied under different growth conditions, indicating differential expression of the *adh* genes in X514 under different growth conditions.

186

Nitrate Reduction and Functional Characterization of *c*-type Cytochromes in *Shewanella*

Dongru Qiu,[1]* Haichun Gao,[1] Zhili He,[1] Jingrong Chen,[1] Yili Liang,[1] Soumitra Barua,[1,2] Margaret F. Romine,[3] Samantha Reed,[3] Dave Culley,[3] David Kennedy,[3] Yunfeng Yang,[2] Kenneth H. Nealson,[5] James M. Tiedje,[4] James K. Fredrickson,[3] and **Jizhong Zhou**[1,2] (jzhou@ou.edu)

[1]Institute for Environmental Genomics and Dept. of Botany and Microbiology, University of Oklahoma, Norman; [2]Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, Tenn.; [3]Pacific Northwest National Laboratory, Richland, Wash.; [4]Center for Microbial Ecology, Michigan State University, East Lansing; and [5]Dept. of Earth Sciences, University of Southern California, Los Angeles

Project Goals: As part of the GTL program, our research is focused on nitrate reduction and functional characterization of c-type cytochromes in *Shewanella* *Shewanella* species are characterized by their respiratory versatility and psychrophility. Their ability to utilize a wide range of electron acceptors for respiration is due to a large number of *c*-type cytochromes in their genome. The dissimilatory metal reduction capacity of *Shewanella* and *Geobacter* provides a potential opportunity for the efficient bioremediation and electricity generation.

The NapC- and CymA-Dependent Nitrate Reduction in *Shewanella*

Nitrate respiration systems are highly diverse among *Shewanella* species. Bioinformatics analysis revealed three types of nitrate reduction systems in *Shewanella* genomes. *S. oneidensis* MR-1 harbors only the CymA-dependent nitrate reductase and the NapGH ubiquinol oxidase. Most *Shewanella* species, including *S. putrefaciens* W3-18-1, have both CymA- and NapC-dependent nitrate reductases, as well as the NapGH ubiquinol oxidase. The *S. baltica* strains have both the CymA- and NapC-dependent nitrate reductases but lack the NapGH ubiquinol oxidase. MR-1 appears to be atypical because it lacks both *napC* and *nrfBCD*, whose gene products act to transfer electrons from the quinol pool to terminal reductases NapA and NrfA. In *E. coli*, NapC and NrfBCD are essential for catalyzing reductions of nitrate to nitrite and the subsequent reduction of nitrite to ammonium, respectively. Our previous results revealed that CymA is likely to be a functional replacement of both NapC and NrfBCD in the nitrate and nitrite reduction in MR-1.

Our results suggest that the two-step manner of nitrate reduction found in MR-1 may be common among *Shewanella* species. Deletion of the *nap1* (*napDAGHB*) or *nap2* (*napDABC*) operon did not significantly affect cell growth, but the double mutant could not grow on nitrate, suggesting that the two *nap* operons are functionally redundant. In addition, the in-frame of *cymA* and *napC* of W3-18-1 deletion mutants did not show severe growth inhibition on nitrate, though deletion of *cymA* resulted in the loss of nitrate and nitrite reduction and growth in MR-1. Furthermore, the *cymA* deletion mutant showed little growth on nitrite in contrast to the *napC* deletion mutant, indicating that CymA was involved in nitrite reduction in both W3-18-1 and MR-1. The *cymA* gene from W3-18-1 complements the MR-1 *cymA* in-frame deletion mutant and allows reduction of ferric ions, nitrate, and nitrite when expressed *in trans*. The *napC* gene from W3-18-1 can also complement the MR-1 *cymA* deletion mutant and allows ferric iron reduction but it failed to allow nitrite reduction. These results support the hypothesis that the NapC-dependent and CymA-dependent periplasmic nitrate reduction systems allow an efficient dissimilatory reduction of nitrate and nitrite. Deletion of *narP* and *narQ* resulted in the growth inhibition on nitrate, suggesting that nitrate reduction is also regulated by the NarQP two- component system in W3-18-1. Our competition assays showed that W3-18-1 had a competitive advantage over MR-1 when grown together on nitrate.

Characterization of *C*-type Cytochromes and Their Role in Anaerobic Respiration in *Shewanella*

The arsenal of *c*-type cytochromes is also highly diversified across the 21 sequenced *Shewanella* genomes and only twelve of the 41 c-type cytochrome of *S. oneidensis* MR-1 are present in all other sequenced strains. Only a few *c*-type cytochromes have been characterized. To discern the functions of unidentified *c*-type cytochrome genes in bacterial energy metabolisms, we generated 37 single mutants with an in-frame deletion of each individual cytochrome gene in MR-1. Reduction of a variety of electron acceptors was measured and the relative fitness was calculated for these mutants based on competition assays. This revealed that SO0610, SO1777, SO2361, SO2363, and SO4360 were important under aerobic growth conditions, and that most *c*-type cytochomes play a more important role in anaerobiosis. The *petC* gene appeared to be important to both aerobiosis and ananerobiosis. Our results regarding functions of CymA and MtrC are consistent with previous findings. We also assayed the biofilm formation of these mutants and results indicate that SO4666 might be important for pellicle formation.

S. putrefaciens W3-18-1 lacks orthologues for the secondary metal reductase and accessory proteins (MtrFED) of *S. oneidensis* MR-1. Sputw3181_2446 encodes a decaheme *c*-cytochrome, orthologous to the outer membrane primary metal reductase OmcB of MR-1 (60% similarity) while another reductase similar to OmcA in MR-1 was also found in W3-18-1. Sputw3181_2445 encodes an 11-heme *c*-type cytochrome OmcE, which only shares 40% similarity with the decaheme cytochrome OmcA. Single and double in-frame deletion mutants of *omcB* and *omcE* were generated for functional characterization of *omcE* and metal reduction in W3-18-1. Reduction of solid-phase Fe (III) and soluble Fe (III) in *S. putrefaciens* W3-18-1 was mainly dependent on OmcB under anaerobic conditions (with 50 mM lactate as electron donors and Fe_2O_3, α-FeO (OH), β-FeO (OH) and ferric citrate as electron acceptors. W3-18-1 catalyzed a more rapid reduction of α-FeO (OH) as compared to MR-1, suggesting that other genes may be involved in Fe (III) reduction in W3-18-1. As previously observed in MR-1, the deletion of both OmcE and OmcB led to a severe deficiency in reduction of solid-phase Fe (III) in W3-18-1 and and an even greater deficiency in the reduction of soluble iron. The *omcB* and *omcE* genes of W3-18-1 have been expressed with the pBAD vector in *E. coli*. Heme staining assays also demonstrated that the disappearance of specific protein bands in the SDS-PAGE gels were consistent with *omcB* and *omcE* deletion in three mutant samples. These results suggest that *omcE* and *omcB* are actually expressed as cytochrome proteins and could play a central role in metal reduction in *S. putrefaciens* W3-18-1.

187

Expression, Purification, and SAXS Structural Studies of the *Caulobacter* Chromosomal Segregation Machinery Complexes and Components

Jian Zhu,[1] Ping Hu,[1] Gary Anderson,[1] Thomas Earnest[1]* (TNEarnest@lbl.gov), and **Harley McAdams**[2]

[1]Lawrence Berkeley National Laboratory, Berkeley, Calif. and [2]Stanford University, Palo Alto, Calif.

Project Goals: The purpose of this project is to: 1) map the protein-protein interaction network of *Caulobacter crescentus* including the implementation of TAP tagging to identify interaction partners for proteins that form biomolecular complexes; 2) isolate these complexes by co-expression in *E. coli*, overexpression in *Caulobacter* and/or cell-free systems and to automate these steps to achieve high-levels of throughput for complexes that will be applied on a pathway- to genome-scale; 3) characterize these complexes biochemically, functionally, and structurally at the molecular to cellular level.

Caulobacter crescentus is stalked α-proteobacteria with significant demonstrated potential as a bioremediative agent, as well as being a well-characterized model organism for microbial systems biology. Approaches for the expression, purification, stabilization, and complex formation have been developed that allow for a high-success rate for the production of high-value target proteins and protein complexes. Chromosomal segregation during cell division is highly coordinated spatially and temporally through a set of proteins whose interactions drive the process of properly positioning the chromosomes to the two cells poles. In *Caulobacter* this utilizes the ParA and ParB proteins that interact with *parS* (proximal to the origin of replication). Additionally in *Caulobacter* these proteins interact with the novel cell polarity protein, PopZ, to attach the daughter chromosomes to the two cell poles through interactions of PopZ with ParB and *parS* (Bowman *et al.*, 2008). We express PopZ, ParA, and ParB individually in *E. coli* in sufficient quantities for structural and functional studies. SAXS studies on these proteins individually and in complex with each other reveals sets of interactions for these proteins. SAXS and native gel studies on PopZ alone indicates that in solution this ~20 kD proline-rich polypeptide forms a large assembly of ~ 300 kD which can be disassembled with urea and reassembled by urea removal *in vitro*. ParB forms a homo-dimer that becomes more compact with the addition of *parS*. ParA is an ATPase that oligomerizes upon ATP binding and interaction with ParB (Figure 1), and is thought to provide the force required to move the chromosomes to their proper position. The ParA/ParB/*parS* complex demonstrates a significant increase in the radius of gyration upon addition of ATP. The SAXS data have generated models for ParB, ParA and their complexes that can be compared and fitted with crystal structures of the components. These approaches, combined with the genetic, biochemical, and microscopic

data, are utilized to address structural and functional studies on a number of protein complexes involved in cell polarity, cell cycle control, transcriptional regulation, and bioremediation.

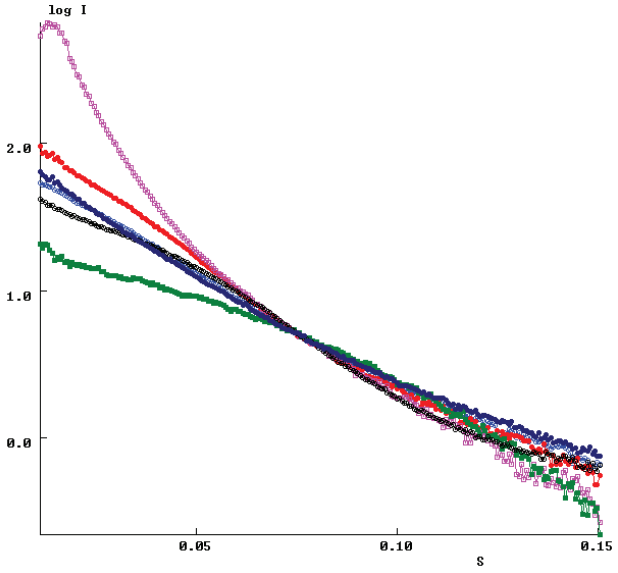


Figure 1. SAXS experiments on ParA, ParB, and *parS* demonstrate oligomerization upon addition of ATP (purple curve) compared with the sample without ATP (red curve). We thank Greg Hura of the ALS SIBYLS beamline 12.3.1 for assistance with the SAXS experiments.

Biological Systems Interactions

188

PNNL Foundational Scientific Focus Area—Biological Systems Interactions

Jim Fredrickson[1]* (jim.fredrickson@pnl.gov), **Gordon Anderson**,[1] **Scott Baker**,[1] **Alex Beliaev**,[1] **Mary Lipton**,[1] **Jon Magnuson**,[1] **Margie Romine**,[1] **Thomas Squier**,[1] **H. Steven Wiley**,[1] Don Bryant,[2] Frank Collart,[3] William Inskeep,[4] Francois Lutzoni,[5] Andrei Osterman,[6] Margarethe Serres,[7] Harvey Bolton Jr.,[1] and David Ward[3]

[1]Pacific Northwest National Laboratory, Richland, Wash.; [2]Pennsylvania State University, University Park;[3]Argonne National Laboratory, Argonne, Ill.; [4]Montana State University, Bozeman; [5]Duke University, Durham, N.C.; [6]Burnham Institute for Medical Research, La Jolla, Calif.; and [7]Marine Biological Laboratory, Woods Hole, Mass.

Project Goals: The primary FSFA objectives include—develop a mechanistic understanding of metabolic interactions among key members of microbial mats using the tools of genomics and systems biology; understand the collective energy, carbon, and nutrient processing in laboratory-based microbial systems that contributes to their stability and efficient utilization of resources using a systematic application of –omics (transcriptomic, proteomic, and metabolomic) approaches; determine interspecies co-adaptations and functional innovations that contribute to robustness and functional efficiency; explore microbe-microbe and microbe-environment interactions that control genome evolution; determine the functional content of the mobile pool of genes in microbial mats and corresponding mechanisms by which they are disseminated; and understand cellular strategies that permit a system of interacting organisms to control the excess generation of reactive oxygen species to promote adaptive responses that enhance their survival; and systems biology investigations of the culturable lichen *Cladonia grayi* to understand mechanisms of resilience against environmental stress.

The PNNL Genomic Science Foundational Scientific Focus Area (FSFA), initiated at the beginning of FY10, is addressing critical scientific issues on microbial interactions, investigating how microorganisms interact to carry out, in a coordinated manner, complex biogeochemical processes such as the capture and transfer of light and chemical energy. The primary research emphasis will be on associations between autotrophic and heterotrophic microorganisms with the additional objective of obtaining a predictive understanding of how interactions impart stability and resistance to stress, environmental fitness, and functional efficiency. The main scientific objectives of the FSFA include: development of a mechanistic understanding of interactions among key members of microbial autotroph-heterotroph associations (AHA) using the tools of genomics and systems biology; understanding the collective energy, carbon, and nutrient processing in AHAs that contributes to their stability and efficient utilization of resources; probing interspecies co-adaptations and functional innovations that contribute to robustness and functional efficiency and exploring the types of microbe-microbe and microbe-environment interactions that control genome evolution; understanding cellular strategies that permit a system of interacting organisms to control the excess generation of ROS to promote adaptive responses that enhance their survival; and systems biology investigations of the culturable lichen *Cladonia grayi* to understand mechanisms of resilience against environmental stress.

Autotroph-heterotroph microbial associations formed the foundation of the biosphere nearly 3 billion years ago with oxygenic photosynthetic prokaryotes (cyanobacteria) and their associated heterotrophic partners colonizing shallow ocean zones. The photolithotrophs use sunlight for energy to fix CO_2 and N_2 and produce O_2, H_2, and organic molecules that supported the growth and metabolism of their heterotrophic partners that facilitate recycling of carbon and nutrients. Autotroph-heterotroph associations are common planet-wide, representing metabolically interactive, self-sustaining communities that are often pioneering and can represent the only biota in extreme environments. These associations are well-adapted to a range of harsh conditions

that include extremes of temperature, salinity, desiccation, irradiance, high O_2, and nutrient deprivation. Microbial associations, inclusive of photolithotrophs (e.g., light energy) and chemolithotrophs (e.g., inorganic chemical energy), are highly relevant to DOE mission areas including bioenergy, carbon cycling/sequestration, and contaminant fate and transport. Further, interacting microorganisms provide key services such as carbon, nutrient, and metal cycling to the biosphere, have considerable potential for a wide range of biotechnological applications, and present challenging and exciting new basic research opportunities.

The PNNL FSFA is utilizing genome-enabled systems biology approaches on three levels—molecular, cellular, and community—to elucidate the underlying design principles of microbial associations, emphasizing interactions between microorganisms for which there are established or hypothesized interdependencies. As part of this approach, the FSFA is developing the experimental tools and data necessary to quantitatively understand and predict causal relationships between environmental change, microbial associations, and cellular functions. The FSFA is using a combined top-down/bottom-up approach where bioinformatics-based genome functional predictions are made using a range of tools and resources; evaluations of evolutionary and ecological adaptation processes are made at the genome scale, high-throughput expression analyses and functional genomics are used to uncover key genes and proteins as well as metabolic and regulatory networks. The bottom-up component uses genetic, physiological, and biochemical approaches to test or verify predictions made by the top-down approaches. The top-down experimental component includes the generation of large amounts of data from biological perturbation experiments that support computational analyses to develop models of various cellular networks. The FSFA is utilizing a series of lab-based model systems consisting of constructed consortia with engineering potential, natural communities, and consortia derived from natural communities for hypothesis testing. Natural communities include microbial mats, biofilms, and lichens. These systems include associations that have evolved to permit successful colonization of extreme environments through effective utilization of solar and chemical energy and scarce nutrients. Research involving these systems will guide our ability to understand and predict how biological associations function with a high degree of efficiency and resiliency. A significant advantage is afforded by using a combination of mechanistic and systems-level investigations of representative associations cultivated in the laboratory under controlled conditions and analyses of natural assemblages using analytical and computational tools of systems biology.

189

Transcriptional Regulation of *Shewanella* Central Carbon Metabolism by HexR

Dmitry Rodionov[1,4*] (rodionov@burnham.org), Xiaoqing Li,[1] Samantha Reed,[2] Margaret Romine,[2] James Fredrickson,[2] Pavel Novichkov,[3] Semen Leyn,[4] and Andrei Osterman[1,5] (osterman@burnham.org)

[1]Burnham Institute for Medical Research, La Jola, Calif.; [2]Pacific Northwest National Laboratory, Richland, Wash.; [3]Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, Calif.; [4]Institute for Information Transmission Problems RAS, Moscow, Russia; and [5]Fellowship for Interpretation of Genomes, Burr Ridge, Ill.

Project Goals: This project was started as a component of the *Shewanella* Federation studies aimed at better understanding of the ecophysiology and speciation of this important genus. It is now continued in the framework of the PNNL Foundational Scientific Focus Area (FSFA) "Biological Systems Interactions" (PI, J. Fredrikson). This FSFA has a scientific focus on understanding interactions between microbes using systems biology approaches including state-of-science technologies, and it is focused on interactions between microbes and their extracellular environments, with an emphasis on acquiring an understanding of microbial autotroph-heterotroph associations. The predictive understanding of these biological systems will be acquired by integrating experimental and computational approaches that exploit the expertise from multiple disciplines working in a synergistic manner. Genomics and proteomics approaches will be applied to predict gene occurrence and function and to identify, quantify, and characterize individual proteins and complexes. Microbial physiology, ecology, biochemistry, metabolic profiling and genetic approaches will be used to verify predictions and test specific hypotheses.

The absence of genes encoding phosphofructokinase in all sequenced Shewanellae, leads to the Entner-Doudoroff (ED) and pentose phosphate (PP) pathways being the major routes of sugar utilization rather than a glycolytic route for central carbohydrate metabolism (CCM). Not surprisingly, such a redistribution of catabolic flux is associated with a completely different regulatory strategy compared to those used for classical *glycolytic* metabolisms found in other proteobacteria such as *Escherichia coli*. Using a comparative genomics approach, we have identified a novel *Shewanella* regulon that is controlled by HexR and that encompasses ~30 genes from the CCM pathways, as well as the deoxynucleoside and glycine utilization. The HexR-binding motif was predicted to be a 17-bp palindromic sequence with the consensus tTGTAATwwwATTACa. Assay of purified HexR protein by electrophoretic mobility shift analysis confirmed recognition of the predicted binding motifs by this regulator. The ED pathway intermediate, 2-keto-3-deoxy-6-phosphogluconate, functions as a HexR antagonist releasing it from its target operator. Analysis of the relative

position of the HexR binding sites and candidate promoters in multiple *Shewanella* genomes suggested a dual mode of HexR action; negative regulation (repression) of some of the target genes and positive regulation (activation) of others. This observation is in agreement with the expression patterns of 27 predicted HexR regulon genes observed in the ~200 *S. oneidensis* MR-1 microarray experiments available in the M3D database (http://m3d.bu.edu/).

Overall, three distinct groups of highly correlated HexR-regulated genes were revealed: (i) *zwf-pgl-edd-eda*, *pykA*, *tal-pgi*, *gapA2*; (ii) *phk*, *deoAB*, *cdd*, *nqrABCDEF*; and (iii) *ppsA*, *gapA3*, *gcvTHP*. Remarkably, the third group of genes showed a strong anti-correlation with the first two groups supporting the proposed dual mode of HexR regulation. This observation was directly supported by qPCR-based comparison of the expression of HexR regulon genes in the wild-type and a targeted *hexR* deletion mutant of *S. oneidensis*. The most significant differences in WT vs. mutant gene expression patterns were observed between genes involved in catabolic pathways and in gluconeogenesis (repressed or activated, respectively). For example, of the two genes, *pykA* and *ppsA,* that encode enzymes catalyzing phosphoenolpyruvate to pyruvate interconversion in opposite directions, the former is repressed whereas the latter is activated by HexR. Comparison of growth phenotypes of mutant and wild type strains on various carbon sources (N-acetylglucosamine, glycerate, inosine, and lactate) showed that *hexR* deletion leads to an inability of *S. oneidensis* to grow on lactate as a single carbon source. This finding confirmed the observed positive mode of action of the HexR regulator on the gluconeogenic gene, *ppsA*, whose activity is known to be essential for the growth of *E. coli* on lactate. The detailed results of our HexR regulon reconstruction, including the predicted transcription factor binding sites, are presented in a recently developed RegPrecise database (http://regprecise.lbl.gov).

Additional physiological studies and metabolomic profiling analyses are in progress to further investigate the role of HexR in the regulation of CCM in *Shewanella*. The HexR regulon in *Shewanella* may be considered as a partial functional replacement of a classical 6-fructose-phosphate regulon FruR, which is known to control fructose utilization and CCM in *E. coli*. The sequenced Shewanellae lack FruR and are not able to grow on fructose. Reconstruction and comparative analysis of HexR regulons was expanded to a broader set of genomes from γ-, and β-proteobacteria and some Firmicutes contributing to better understanding of evolutionary history of HexR and its role in the regulation of CCM.

These studies demonstrate the value of applying comparative genomics and complementary experimental analyses to predict and validate regulatory networks in previously uncharacterized biological systems. As part of the new Foundational Science Focus Area project team led by the Pacific Northwest National Laboratory we intend to continue applying such strategies to explore regulatory networks in individual species and communities.

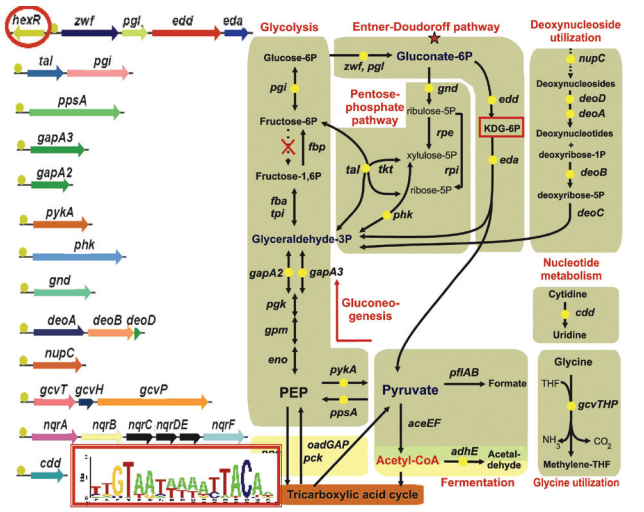


Fig. 1. The reconstructed HexR regulon and target pathways in *Shewanella oneidensis* MR-1.

190

Finding Function for Fungal Glycoside Hydrolases

Jon K. Magnuson,[1] Frank Collart,[2] Deanna Auberry,[1] Sarah Zerbs,[2] Ellen A. Panisko,[1] Justin Powlowski,[3] Adrian Tsang,[3] and **Scott E. Baker**[1]* (scott.baker@pnl.gov)

[1]Pacific Northwest National Laboratory, Richland, Wash.; [2]Argonne National Laboratory, Argonne, Ill.; [3]Concordia University, Montreal, Quebec, Canada

Project Goals: The goal of this project is to develop a pipeline for functional annotation of fungal glycoside hydrolases.

Plant biomass is efficiently broken down and used a source of carbon by many microbes, including fungi. Many enzymes that are involved in the deconstruction of cellulose and hemicelluloses are members of a large group of enzymes called glycoside hydrolases (GHs). On average, the genomes of filamentous fungi contain well over 100 genes encoding different GH family enzymes. Currently, GHs are classified into families based on sequence and predicted structure. GH families may contain multiple enzymatic activities. The goal of our project is to develop a pipeline for functional characterization of GHs. Functional information GHs is generated by genome annotation, proteomic analysis of the secretome and enzymatic assays of expressed GHs of interest. We have cultured on different substrates and subsequently processed a variety of fungal secretomes for proteomic analysis. We compare protein expression profiles of GHs across different culture substrates. Finally, we have compared the activity and stability of expressed *Aspergillus niger* enzymes produced by different hosts: *E. coli*, *Pichia* and *Aspergillus niger*. Our data allow us to add functional data to annotations of fungal

GHs there were previously characterized based solely on predicted amino acid sequence analysis.

191

High-Temperature Chemotrophic Microbial Communities of Yellowstone National Park: Metagenomics Provides a Foundation for Dissecting Microbial Community Structure and Function

W. Inskeep[1]* (binskeep@montana.edu), Z. Jay, M. Kozubal,[1] J. Beam,[1] R. Jennings,[1] H. Bernstein,[2] R. Carlson,[2] D. Rusch,[3] and S. Tringe[4]

[1]Dept. of Land Resources and Environmental Sciences and Thermal Biology Institute and [2]Dept. of Chemical and Biological Engineering, Montana State University, Bozeman; [3]J. Craig Venter Institute, Rockville, Md.; and [4]DOE Joint Genome Institute, Walnut Creek, Calif.

Project Goals: The goals of this project are to to study microbial interactions in model geothermal microbial communities. Extreme geochemical conditions and high temperature result in low-diversity microbial communities where metagenomic sequence data can be used to dissect microbial community structure and function.

Microbial communities are a collection of interacting populations. However, a significant fraction of our knowledge base in microbiology originates from organisms grown and studied in pure culture, in the absence of other members of the community who may compete for resources or provide necessary co-factors and or substrates. Moreover, many of the organisms studied in pure culture have not necessarily represented the numerically dominant members of microbial communities found in situ. The advent of molecular tools (e.g., genome sequencing) has provided opportunities for assessing the predominant and relevant indigenous organisms, as well as their likely function within a connected network of different populations (i.e., community). High-temperature microbial communities are often considerably less diverse than mesophilic environments and constrained by dominant geochemical attributes such as pH, dissolved oxygen, Fe, sulfide, and or trace elements including arsenic and mercury. Consequently, the broader goal of our work is to utilize extreme high-temperature geothermal environments including acidic Fe-oxidizing communities as model systems for understanding microbial interactions among community members. Recent metagenomic sequencing of high-temperature, acidic Fe-mats of Norris Geyser Basin, Yellowstone National Park (YNP), conducted as part of a DOE-Joint Genome Institute (Community Sequencing Project) reveal communities dominated by novel members of the *Archaea*, bacterial members of the deeply-rooted Order Aquificales as well as other less-dominant Bacillales and Clostridiales. Phylogenetic and functional analysis of metagenome sequence is providing an excellent foundation for establishing the role of individual populations in a network of interacting community members, and for directing hypotheses regarding the importance of specific biochemical pathways responsible for material and or energy cycling. For example, we are using metagenome sequence in combination with information available from reference strains to identify protein-coding sequence of importance in the oxidation and or reduction of Fe, S, O, and As, as well as central C metabolism (including fixation of CO_2). Genes coding for proteins with hypothetical or putative roles in electron transfer, C-capture and C-transformation have been prioritized for design of quantitative-reverse transcriptase-PCR (Q-RT-PCR) primers to evaluate functional capacity quantitatively in both pure-culture and subsequent mixed communities. Future proteomic and transcriptomic analyses, as an element within the PNNL Foundational Scientific Focus Area, will focus on both pure-culture experiments under different electron donor and acceptor conditions, as well as natural thermophilic mats. Proteomic results will be used to assess and confirm the importance of specific proteins and to improve microbial community models. Application of genomic, proteomic, and metabolic information to dissect microbial community structure and function is tractable within high-temperature geothermal systems in part due to the relative simplicity of the community and the dominance of several key geochemical variables (i.e. pH, Fe, O_2).

192

Comparative Genome Analyses of Members of the Ecologically Versatile Genus *Shewanella*: Searching for Sequence Signatures That Reflect Environmental Adaptation

Margrethe H. Serres[1]* (mserres@mbl.edu), Raghu P.R. Metpally,[1,2] and Margaret F. Romine[3]

[1]Marine Biological Laboratory, Woods Hole, Mass.; [2]University of Iowa, Iowa City; and [3]Pacific Northwest National Laboratory, Richland, Wash.

Project Goals: This project is a component of the *Shewanella* Federation and, as such, contributes to the overall goal of applying the genomic tools to better understand the ecophysiology and speciation of respiratory-versatile members of this important genus.

Shewanellae are an environmentally important group of bacteria whose members have been isolated from many different ecosystems (fresh and marine waters and sediments, a deep sea iron mat, subsurface sedimentary rock, and squid gland) that vary in atmospheric pressure, temperature, and salinity. These organisms thrive in red-ox interfaces in the environment and are well known for their versatile respiratory capability, using over 20 different compounds as electron acceptors. Complete genome sequences have been obtained for 19 *Shewanellae*. The strains were selected for sequencing based on their phylogenetic distance, some representing closely related sub-species clusters and others being more distantly related. Overall they represent a genetic gradient in which speciation and environmental adaptation

can be studied. A recent comparative analysis of the genome sequences, phenotypic characteristics, and proteomic expression profiles of the first ten strains sequenced showed that phenotypic and genotypic similarities largely correlated with phylogenetic distances despite the evidence of laterally transferred genes (1). Some of the phenotypic and genotypic traits were more conserved with increasing evolutionary distance (i.e. predicted metabolic pathways) than others (i.e. protein expression patterns).

Our comparative analysis has extended to 19 completed *Shewanella* genome sequences, including more distantly related strains that are obligately marine or that thrive in lower or higher temperatures than those previously studied. The protein sequences from all 19 strains have been analyzed for their domain content (Pfam, TIGRfam) in order to detect differences (functional, protein family compositions) that correlates with the environment in which the strains were isolated. Protein families involved in responses to environmental factors (chemotaxis proteins, two-component regulators, signaling proteins) appear to be large and more diverse among the *Shewanellas*. A curated table of orthologous proteins has been generated for the sequenced shewanellae allowing for categorization of *Shewanella* proteins as core (present in all strains), dispensable (absent in one or more strains), or strain-specific. A significant amount of curation relating to gene calling and function assignments has been done for this dataset, which is available in the *Shewanella* Knowledgebase (2). Of the 16612 orthologous groups of *Shewanella* genes (redundant genes removed), 11% are core genes, 54% dispensable genes and 35% unique genes, the two latter categories encoding genes involved in adaptation to and survival in select environments.

In addition to studying the presence and absence of genes, we are now searching the *Shewanella* genomes for evidence of selective changes in the sequences that can be linked to our knowledge of specific functions or specific environmental conditions of the strains. Sequence changes at non-synonymous vs. synonymous sites have been identified to find genes that are under a purifying or a diversifying selection pressure. Amino acid replacement ratios, radical vs. conservative changes, have also been determined (3). Such codon usage analyses have been used to identify genes that are under biochemical or ecological constraints. We have also calculated the Codon Adaptation Indexes for the *Shewanella* sequences, an estimate of the synonymous codon usage bias and of gene expression levels. The *Shewanella* ortholog table is used as a framework for our studies allowing us to separate the genes into sets that are common to all of the *Shewanellae* or that vary among the strains giving them their unique characteristics. The sequence changes are also evaluated relative to the gene product functions, locations, and protein family memberships.

Our comparative analysis of members of the *Shewanella* genus is forming the foundation for studying other groups of related organisms as well as consortia of microbes in selected environments.

References

1. Konstantinidis KT, Serres MH, Romine MF, Rodrigues JL, Auchtung J, McCue LA, Lipton MS, Obraztsova A, Giometti CS, Nealson KH, Fredrickson JK, Tiedje JM. 2009. Comparative systems biology across an evolutionary gradient within the *Shewanella* genus. Proc Natl Acad Sci U S A. 106 (37):15909-14.
2. www.shewanella-knowledgebase.org
3. Hanada K, Shiu SH, Li WH. 2007. The nonsynonymous/synonymous substitution rate ratio versus the radical/conservative replacement rate ratio in the evolution of mammalian genes. Mol Biol Evol. 10:2235-41.

193

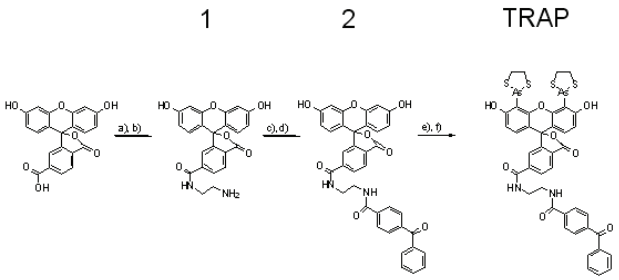
In-Vivo Trapping and Structural Interrogation of Transient Protein Complexes

M. Uljana Mayer, Ping Yan, Ting Wang, Yijia Xiong, Diana J. Bigelow, and **Thomas C. Squier*** (thomas.squier@pnl.gov)

Pacific Northwest National Laboratory, Richland, Wash.

Project Goals: Identification of new imaging methods that permit high-throughput measurements of protein complexes that will allow the development of predictive models for bioenergy applications and with regard to how microbes respond to environmental change.

Protein cross-linking, especially coupled to mass-spectrometric identification, is increasingly used to determine protein binding partners and protein-protein interfaces for isolated protein complexes. The modification of cross-linkers to permit their targeted use in living cells is of considerable importance for studying protein-interaction networks, which are commonly modulated through weak interactions that are formed transiently to permit rapid cellular response to environmental changes. We have therefore synthesized a targeted and releasable affinity probe (TRAP) consisting of a biarsenical fluorescein linked to benzophenone that binds to a tetracysteine sequence in a protein engineered for specific labeling (Scheme 1). Here, the utility of TRAP for capturing protein binding partners upon photoactivation of the benzophenone moiety has been demonstrated in living bacteria and eukaryotic cells. In addition, ligand exchange of the arsenic-sulfur bonds between TRAP and the tetracysteine sequence to added dithiols results in fluorophore transfer to the crosslinked binding partner. Following isolation of protein complexes, the facile release of TRAP from the original binding site permits the identification of the proximal binding interface through mass spectrometric fragmentation and computational sequence identification.



Scheme 1. *Synthesis of TRAP.* A) EDC, Et_3N, NHS, dry DMF, 30 min; B) *N*-Boc-ethylenediamine, 16 h; C) 20 % TFA/CH_2Cl_2, 2 h, D) 4-Benzoylbenzoic acid, EDC, NHS, iPr_2EtN, DMF, 16 h; E) HgO, TFA, 70 °C; F) 1) $AsCl_3$, PdOAc, iPr_2EtN, NMP, 4 h; 2) EDT, 20 % acetone/H_2O (overall yield: 2 %).

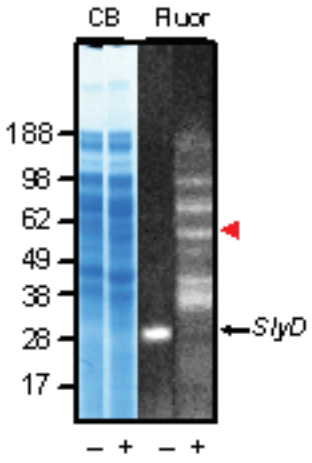


Figure 1: *Selective in-vivo labeling of chaperon SlyD and Photo-dependent Trapping of Binding Partners that Include Elements of [Ni-Fe]-Hydrogenase Maturation Pathway.* Coomassie Blue (CB) protein stain (left panel) and fluorescence image (right panel) before (-) and following (+) photodependent cross-linking. Positions of SlyD and HypB (red arrow), previously shown to bind SlyD, are indicated.

One example of the utility of using TRAP to identify molecular pathways associated with the maturation of [Ni-Fe]-hydrogenases involves the introduction of a tetracysteine tag into the chaperone SlyD expressed in *Shewanella*, permitting its selective labeling (Figure 1). Following photoactivation and reduction, the fluorescence TRAP moiety is readily transferred to binding partners – resulting in a considerable simplification in the identification of the unique molecular interfaces following proteolytic digestion and mass spectrometric analysis (Figure 2).

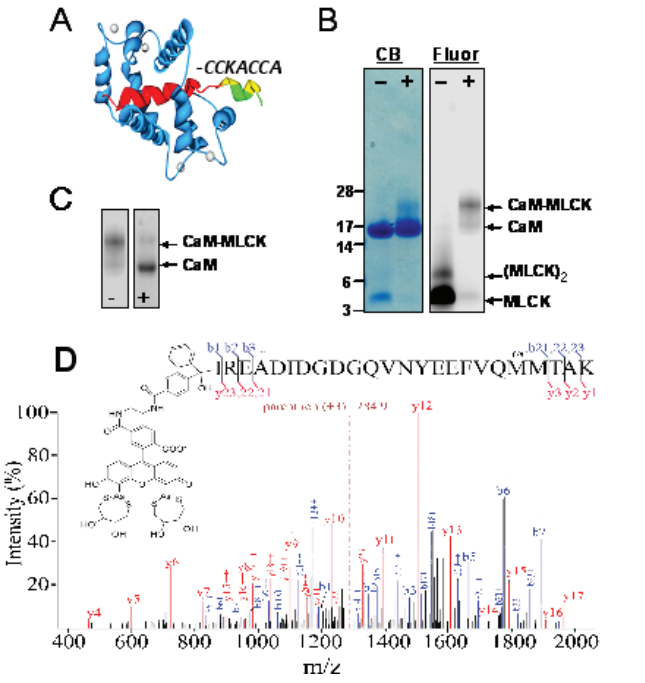


Figure 2: *Mass Spectrometric Identification of Interfacial Binding Site Following Crosslinking and Fluorophore Transfer.* (A) Depiction of the structural interface between tagged MLCK protein (red) containing engineered TRAP binding site CCKACCA (yellow and green) in complex with binding partner CaM (blue). (B) Coomassie Blue (CB) protein stain and fluorescence from TRAP reagent (Fluor) prior to (-) and following (+) photo-induced crosslinking. (C) Cross-linked complex before (-) and following (+) transfer of TRAP to binding partner upon addition of DTT (5 mM). (D) Mass spectrum of TRAP-bound peptide, where all b (blue) and y (red) fragment ions are identified. Site of TRAP binding in CaM bait protein is shown above mass spectrum.

Conclusions: We have synthesized a new class of photocrosslinker (i.e., TRAP) that can be targeted to a tetracysteine tag to capture (that is, trap) protein binding partners upon light activation following the straight-forward introduction of a unique tetracysteine binding sequence onto a protein of interest. The simplicity of this method will facilitate the high-throughput identification of protein complexes in a range of different organisms. The small size of TRAP compared to other targeted and multifunctional crosslinkers enables facile identification of the site of cross-linking by searching for the added molecular weight of the transferred crosslinker using commonly available mass spectrometers and publically available software. Optimal conditions associated with cross-linking complexes of interest are facilitated by the high-fluorescence yield of the TRAP reagent, enabling facile visualization of TRAP fluorescence on SDS-PAGE gels.

Future Measurements: TRAP will be used in combination with targeted in vivo photocrosslinking and mass spectrometry to identify interfacial binding sites, permitting an understanding of how environmental conditions affect protein-protein interaction networks. Coupled with the use of complementary multiuse affinity reagents (MAPs)

that permit the visualization of protein localization within microbial communities, we propose to use these reagents to identify regulatory elements that modulate energy flux through key metabolic pathways associated with biomass and the generation of biofuels.

Reference

1. Yan P., T. Wang, G.J. Newton, T.V. Knyushko, Y. Xiong, D.J. Bigelow, T.C. Squier, and M.U. Mayer (2009) *A targeted releasable affinity probe (TRAP) for in vivo photo-crosslinking*. Chembiochem. 10: 1507-1518.

Plant-Microbe Interfaces

194
Plant-Microbe Interfaces

Mitchel J. Doktycz[1]* (doktyczmj@ornl.gov), Gerald A. Tuskan,[2] Christopher W. Schadt,[1] Gregory B. Hurst,[3] Edward Uberbacher,[1] Dale A. Pelletier,[1] Jennifer Morrell-Falvey,[1] Timothy J. Tschaplinski,[2] David J. Weston,[2] Scott T. Retterer,[1] Andrey Gorin,[4] Yunfeng Yang,[1] Robert Hettich,[3] Udaya C. Kalluri,[2] Xiaohan Yang,[2] Abhijit Karve,[2] Mircea Podar,[1] Steven D. Brown,[1] Robert Cottingham,[1] Tatiana Karpinets,[1] Chongle Pan,[4] Guru Kora,[4] Denise Schmoyer,[1] and Susan Holladay[1]

[1]Biosciences Division, [2]Environmental Sciences Division, [3]Chemical Sciences Division, and [4]Computer Science and Mathematics, Oak Ridge National Laboratory, Oak Ridge, Tenn.

http://PMI.ornl.gov

Project Goals (Abstracts 194-205): Understand the genome-dependent molecular and cellular events involved in establishing and maintaining beneficial interactions between plants and microbes. *Populus* and its associated microbial community serve as an initial test system for understanding how these molecular events manifest themselves within the spatially, structurally, and temporally complex scales of natural systems. To achieve this goal, we will focus on 1) characterizing the natural variation in *Populus* microbial communities within complex environments, 2) elucidating *Populus*-microbial interactions at the molecular level and dissecting the signals and pathways responsible for initiating and maintaining microbial relationships, and 3) performing metabolic and genomic modeling of these interactions to aid in interpreting the molecular mechanisms shaping the *Populus*-microbial interface.

Rapid progress in biological and environmental sciences has been enabled by the availability of genome sequences and the tools and technologies involved in interpreting genome function. As our understanding of biological systems grows, it becomes increasingly clear that the functional expression of individual genomes is affected by an organism's environment and the community of organisms with which it associates. The beneficial association between plants and microbes exemplifies a complex, multi-organism system that is shaped by the participating organisms and the environmental forces acting upon it. These plant–microbe interactions can benefit plant health and biomass production by affecting nutrient uptake, influencing hormone signaling, effecting water and element cycling in the rhizosphere, or conferring resistance to pathogens. Studying the integral plant–microbe system in native, perennial plant environments, such as *Populus* and its associated microbial community, provides the greatest opportunity for discovering plant–microbial system functions relevant to DOE missions related to bioenergy and carbon-cycle research and understanding of ecosystem processes.

The functional attributes of *Populus* depend on the microbial communities with which it associates. Bacteria and fungi can be found within *Populus* tissues and closely associated with the roots in the rhizosphere. Understanding these communities, and the interfaces between organisms, is critical to realizing fundamental scientific knowledge that may enable increased plant productivity, ecosystem sustainability, disease resistance, drought tolerance, and ecosystem carbon budgets. This interface can also influence the processes, or mechanisms, by which adaptive traits arise from genetic variation and community function. Microbial rhizosphere structure, plant root bacterial and fungal colonization patterns, and the microbe–plant signaling pathways inherent in each type of association are all found within *Populus* and can be functionally translated hierarchically across scales into ecosystem patterns and processes.

Understanding the mechanisms by which plants and microbes interact represents a grand challenge facing biological and environmental science. How microbial selection and colonization occurs, what reciprocal benefits are bestowed upon the plant and microbe, and how these interactions ultimately affect, and are affected by, the environment are just some of the intrinsic scientific questions. The multiple spatial and temporal scales involved in these interfaces, the complexity of the component systems, and the need for better tools that use and build upon growing genomics resources to probe and interpret these combined systems represent some of the essential technical challenges. The variety and magnitude of these challenges are only offset by the impact and benefit of overcoming these challenges and in applying this understanding to issues as diverse as efficient energy transformation and carbon cycling.

Student Presentation

195

Plant-Microbe Interfaces: Characterization of Cell Surface Properties in *Azospirillum brasilense* Wildtype Cells and Che1 Pathway Mutants Using Atomic Force Microscopy

A. Nicole Edwards,[1]* Piro Siuti,[1] Jennifer L. Morrell-Falvey[1,2] (morrelljl1@ornl.gov), Amber N. Bible,[3] Gladys Alexandre,[1,3] Scott T. Retterer,[2,4] and **Mitchel J. Doktycz**[1,2,4]

[1]Graduate School of Genome Science and Technology, University of Tennessee-ORNL, Knoxville; [2]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tenn.; [3]Dept. of Biochemistry, Cellular, and Molecular Biology, University of Tennessee, Knoxville; [4]Center for Nanophase Materials Sciences, Oak Ridge National Laboratory, Oak Ridge, Tenn.

http://PMI.ornl.gov

Project Goals: See goals for abstract 194.

In order to compete in complex microbial communities, bacteria must quickly sense environmental changes and adjust cellular functions for optimal growth. Such responses are detected and initiated by various two-component signal transduction pathways. Chemotaxis-like signal transduction pathways, which are typically responsible for modulating the swimming motility patterns of chemotactic bacteria, have also been implicated in the modulation of other cellular responses, including cell-to-cell aggregation. Microbial cell-to-cell aggregation is an important behavior which serves to enhance cell survival in adverse environmental conditions. It is particularly advantageous for microbes to organize into aggregative communities not only for protection against predation or antimicrobials, but also for associative metabolic interactions. Plant associative bacteria harbor mechanisms which lead to the remodeling of cell surface and extracellular structures to promote cell-to-cell aggregation and plant microbial interactions. Recently, the Che1 chemotaxis signal transduction pathway from the alpha-proteobacterium *Azospirillum brasilense* was shown to modulate the propensity of cells for cell-to-cell aggregation that leads to flocculation under certain growth conditions. *A. brasilens*e are soil diazotrophic bacteria that colonize the roots of many economically important grass and cereal species. Under conditions of high aeration and limiting availability of combined nitrogen, *A. brasilense* cells differentiate into aggregating cells and form dense flocs that are visible to the naked eye. Flocs are formed by cells embedded in a dense polysaccharide matrix and by cell-to-cell aggregation. Optical and electron microscopy (EM) approaches have been previously applied to compare and identify specific cell surface changes that accompany aggregation and flocculation in *A. brasilense* wild-type strain Sp7 and its Che1 pathway mutant strain derivatives that are affected in their propensity to flocculate under nutritional and aeration stresses. However, no specific extracellular structure could be identified using these techniques, despite the preliminary observation of changes in extracellular polysaccharide (EPS) production detected by growing colonies in the presence of Congo Red. Although optical and EM techniques have revealed many insights into bacterial aggregative behavior, resolution limitations and fixative procedures can inhibit visualization of extracellular structures. Therefore, atomic force microscopy (AFM) was selected as a unique alternative to imaging *A. brasilense* Che1- dependent flocculation behavior at nanometer resolution in an effort to directly visualize changes in cell surface properties that correlate with flocculation. In this study, we investigated *A. brasilense* Sp7 and its Che1 mutant strain derivatives, Δ*cheA1* and Δ*cheY1,* utilizing AFM imaging techniques to gain insight into molecular and regulatory role of Che1 in cell-to-cell aggregation and flocculation. We demonstrate that AFM identifies a distinctive remodeling of the cell surface and extracellular matrix, likely via changes in EPS production, in the Δ*cheA1* and Δ*cheY1* strains concomitant with flocculation under nitrogen-limiting conditions and high aeration.

196

Plant-Microbe Interfaces: Initial Proteome Characterization of the *Populus* Rhizosphere Community

Chongle Pan[1,2]* (panc@ornl.gov), Gregory B. Hurst,[2] Robert L. Hettich,[2] Patricia K. Lankford,[3] Manesh B. Shah,[3] Edward C. Uberbacher,[3] Timothy J. Tschaplinski,[4] Sara Jawdy,[4] Gerald A. Tuskan,[4] Lee E. Gunter,[4] Udaya C. Kalluri,[4] Christopher W. Schadt,[3] Neil R. Gottel,[3] Dale A. Pelletier,[3] and **Mitchel J. Doktycz**[3]

[1]Computer Science and Mathematics Division, [2]Chemical Sciences Division, [3]Biosciences Division, and [4]Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, Tenn.

http://PMI.ornl.gov

Project Goals: See goals for abstract 194.

A complex consortium of microorganisms lives in the plant rhizosphere – a narrow region of soil surrounding the roots. The rhizospheric microbial community feeds on proteins and sugars released by the roots and plant material sloughed off from the roots. In return, the community helps plants to acquire water, nitrogen, and other minerals, and to suppress disease. Here, we describe the development of proteomics methods for understanding this symbiotic relationship. Samples were acquired from both greenhouse-grown and wild *Populus*, the latter including specimens growing in clay and in sandy soil in the Caney Fork State Park in Middle Tennessee (see poster by Schadt et al. for more details). Proteins were extracted from fine roots and associated microorganisms, followed by shotgun proteomics analysis using 2-dimensional liquid chromatography and tandem mass spectrometry. Peptides and proteins were identified by searching mass spectral data against a composite protein sequence database comprised of *Populus* proteins and

genomes of relevant sequenced microorganisms. Because the membership and physiology of the rhizosphere community are influenced by growth conditions and genotypes of *Populus* and environment variables associated with soil, results from these different sites and specimens provided a broad survey of the proteomes of *Populus* rhizosphere community. We identified a core rhizosphere proteome present across all samples and sets of proteins unique to soil types or *Populus* growth conditions. Experiments to improve the depth of the community proteome coverage through optimization of protein extraction methods and employment of high-performance mass spectrometers are ongoing. The observations from these field samples will be followed up with measurements on controlled, reconstituted small soil-less or soil-based model systems. Informatics tools to search proteomic data using the metagenomic sequence data collected from the same samples are also being developed.

197

Plant-Microbe Interfaces: Development of a Knowledgebase for Exploring Plant-Microbe Interactions Using Metabolic Reconstructions

Tatiana Karpinets[1]* (karpinetstv@ornl.gov), Michael Leuze,[2] Guruprasad Kora,[2] Dale A. Pelletier,[1] Mustafa Syed,[1] Byung Park,[2] Denise Schmoyer,[2] Ed Uberbacher,[1] and **Mitchel J. Doktycz**[1]

[1]Biosciences Division and [2]Computer Science and Mathematics, Oak Ridge National Laboratory, Oak Ridge, Tenn.

http://PMI.ornl.gov

Project Goals: See goals for abstract 194.

The Plant-Microbe Interfaces knowledgebase provides an integration framework for microbial, fungal and plant genome data along with information about the characteristics of plants, soils and weather conditions from field studies, their geographical location, diversity and the phenotypic characteristics of microorganisms from soil and plant samples at each location. This macro level information on the samples will be later supplemented by proteomics, metabolomics and transcriptomics data. In this study we have utilized the database to explore metabolic signatures for different classes of organisms interacting with *Populus*. In natural environments a tripartite interaction among plants, fungi, and bacteria is essential to plant growth. On one hand, the intimate relationships within the plant/bacteria/mycorrhizal fungus network supplies plants with nutrients, promotes their growth, and increases their resistance to stress. On the other hand, pathogenic fungi and bacteria can be harmful to plants and can lead to diseases and to altered production of desired traits. Additionally, some microorganisms, also known as biocontrol agents, can protect plants by reducing the number of pathogenic microorganisms. Molecular mechanisms underlying phenotypic differences among plant-associated microorganisms are not clear. The microbial species responsible for beneficial and pathogenic effects on plant growth that have been sequenced provide an opportunity to discover the genomic determinants of the various phenotypes through comparative analysis of the genomes. Metabolic capabilities inferred from genome annotations can be especially helpful in fingerprinting phenotypic differences at the level of enzymes and metabolic pathways. In this study we perform a comparative analysis of phenotypes by developing Pathway Genome Databases (PGDBs) for two fungal species, *Eremothecium gossypii* ATCC 10895 (a plant pathogen) and *Laccaria bicolor* S238N-H82 (a plant symbiont) and a set of sequenced bacterial species including four endophytes (*Pseudomonas putida* W619, *Stenotrophomonas maltophilia* R551-3, Enterobacter sp. 638, *Methylobacterium populi* BJ001), six plant pathogens from the genera *Agrobacterium*, *Pseudomonas*, and *Burkholderia*, and five species from the same genera that are used for biocontrol of phytopathogens.

PGDBs were generated by the Pathologic program from the Pathway Tools software. Because the quality of the metabolic reconstruction by this software depends on the genome annotation, we have developed an automated pipeline to improve the enzyme annotation and to make it consistent across studied organisms. The primary input file to the pipeline contains all of an organism's RefSeq files downloaded from the NCBI website. This file is parsed to build input files for Pathologic and to augment them with the enzyme information from the KEGG orthology annotation of the organism's genome. The PGDB is built with Pathologic running in batch mode, and the PGDB is then refined, to predict transcription units and transporters. Additionally, mySQL tables are created to characterize each protein coding sequence in the genome by a product name, EC numbers, pathways, and protein domains. The domain annotations are generated by searching each sequence against a set of databases (CDD, Pfam, SMART, TIGRFAM, and COG) using RPSBLAST. Thus, the pipeline allows us to quickly incorporate the latest annotation information into the KnowledgeBase, supports metabolic reconstructions, provides a means for improving their quality, and facilitates a comparative analysis of the organisms.

A preliminary analysis of the information generated by the pipeline has revealed some interesting metabolic differences between pathogenic and beneficial microbes at the level of specific enzymes and metabolic pathways. Across analyzed microbial species, biological control agents have a significantly larger number (~10-20%) of metabolic enzymes and pathways when compared with either plant pathogens or endophytes ($p<0.05$). No statistically significant differences, however, were observed when plant pathogens were compared with plant endophytes, indicating a closer metabolic relationship between these phenotypes. In addition to increased metabolic versatility, all of the analyzed biocontrol agents encode in their genomes two specific enzymes that clearly distinguish their metabolic capabilities from the analyzed pathogenic bacteria and endophytes. One enzyme, 1,6-dihydroxycyclohexa-2,4-diene-1-carboxylate dehydrogenase (EC 1.3.1.25), is involved in benzoate degradation, namely in the conversion of benzoate to catechol. It is pres-

ent in biocontrol agents but is not found in any pathogen. The other enzyme, 1-aminocyclopropane-1-carboxylate (ACC) deaminase (EC 3.5.99.7), is absent from endophytes, but is common for biocontrol bacteria. This enzyme catalyzes the conversion of ACC, a precursor of ethylene synthesis in plants, to α-ketobutyrate and ammonia. A variety of beneficial effects on plant growth has been linked to this enzyme, including enhanced nodulation and increased resistance to stress.

198

Plant-Microbe Interfaces: Novel Navigation Techniques to Study Plant-Microbe Associations Utilizing Google Maps API

Guruprasad Kora[1]* (koragh@ornl.gov), Tatiana Karpinets,[2] Denise Schmoyer,[1] Michael Leuze,[1] Byung Park,[1] Mustafa Syed,[2] Dale A. Pelletier,[2] Gerald A. Tuskan,[3] Christopher W. Schadt,[2] Ed Uberbacher,[2] and **Mitchel J. Doktycz**[2]

[1]Computer Science and Mathematics, [2]Biosciences Division, and [3]Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, Tenn.

http://PMI.ornl.gov

Project Goals: See goals for abstract 194.

Efficient integrative analysis, navigation and visualization of information are important prerequisites to understanding dynamic functional interactions between plants and microbes in the environment. The information that must be collected and analyzed includes a diverse set of data characterizing plants, microbe and their environments at different levels of resolution. The collected data will include characteristics of plants, soils and weather conditions from field studies, their geographical location, diversity and phenotypic characteristics of microorganisms from soil and plant samples at each location. This macro-level sample information will be supplemented by proteomics, metabolomics and transcriptomics data, and by genome annotations of the sequenced microorganisms. These micro-level measurements will be also supplemented by results of statistical analyses and by computational predictions produced by developed mathematical models. The unique diversity of the collected data and their composite nature necessitates a more flexible way of managing the information in the project knowledgebase than commonly used relational databases and related visualization and navigation tools.

To address the complex nature of the collected information in studies of plant-microbe associations, a novel navigation and visualization technique based on Google Maps Application Programming Interface (API) is being developed. This approach will allow us to combine a comprehensive mashup of the collected data, high speed visualization and a facilitated supplementation of the data by available knowledge from public Internet resources. The mashup tool provides novel navigation techniques to quickly and efficiently locate and zero-in on interesting plant-microbe associations. The web application will integrate data from different sources and explore experimental datasets that are based upon a common geographical and biological sample space. The technology helps users to generate ideas by identifying relationships and associations both within plants and microbes, and between plants and microbes. The tool provides a sophisticated and intuitive graphical user interface to interactively browse the data using different navigation parameters. It provides for a rich interactive user-experience for users to; 1) filter data by biological, geographical or categorical traits of the collected samples; 2) correlate and compare data based on the user selected parameters; and 3) identify interesting trends and patterns in the collected datasets. Particularly, the tool is used to explore a relationship between soil characteristics, plant genotypic and phenotypic characteristics, and microbial phenotypic, metagenomic and metaproteomic characteristics in field-based studies of *Populus* associated microbial communities. The web application is accessible from http://pmi.ornl.gov.

199

Plant-Microbe Interfaces: Functional Analysis of Phytochrome Signaling in *Populus*

Abhijit A. Karve[1]* (karveaa@ornl.gov), David J. Weston,[1] Sara S. Jawdy,[1] Lee E. Gunter,[1] Stan D. Wullschleger,[1] and **Mitchel J. Doktycz**[2]

[1]Environmental Sciences Division, and [2]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tenn.

http://PMI.ornl.gov

Project Goals: See goals for abstract 194.

In addition to providing energy for photosynthesis, light also modulates the spatial and temporal responses of plants to their surrounding environment. Plants have evolved multiple mechanisms to modulate their growth and development in response to the spectral quality of light. One such mechanism involves red (R) and far-red (FR) light perception by phytochromes (PHY), where a reduction in R:FR due to vegetative shading results in 'shade avoidance response' (SAR). The SAR is characterized by rapid elongation of stem internodes and upward reorientation of leaves. The role of PHYs in shade avoidance is widely studied in the model plant *Arabidopsis thaliana*. In *Arabidopsis*, phytochromes are encoded by PHYA, PHYB, PHYC, PHYD and PHYE; of these, PHYB plays an important role in mediating responses to plant-plant competition. Here we report initial characterization of the PHY gene family from *Populus trichocarpa*. Unlike *Arabidopsis*, the *Populus* genome encodes three PHY genes namely, PtPHYA, PtPHYB1 and PtPHYB2. In order to gain insight into the role of PtPHYs in light signaling, the transcripts of the three PtPHYs in different plant tissues were measured by quantitative RT-PCR. Consistent

with the light labile nature, PtPHYA transcript was the least abundant of all three PHYs studied. The expression of PtPHYB1was highest in female flowers and that of PtPHYB2 was highest in the phloem. In *Arabidopsis*, SAR is characterized by upregulation of key response genes such as *ATHB4, HFR1* and *PIF3*. *Populus* homologues of *ATHB4* and *HFR1* as well as PtPHYB1 and PtPHYB2 showed significant upregulation in response to low R:FR ratio. These results suggest that the mechanism of PHY signaling is partly conserved between *Arabidopsis* and *Populus*. The gene regulatory networks involved in SAR are being studied by microarray analysis on *Populus* exposed to lower R:FR light. Finally, phytochromes are protein kinases and are believed to affect the downstream responses by interacting with other proteins. In order to identify the PHY interacting proteins, *Populus*-specific PHYs were cloned as C-terminal and N-terminal green fluorescent protein (GFP) and hemagglutinin (HA) tagged constructs. The protein localization was then studied by expressing the GFP-fusion protein in sweet pea leaf mesophyll protoplasts. The candidate PtPHY interacting proteins will be identified by protein-protein interaction assays in leaf mesophyll protoplasts. This work will lead to a conceptual model of phytochrome-mediated responses to shade avoidance and to a more detailed understanding of light-induced signaling cascades in *Populus.*

200

Plant-Microbe Interfaces: Transcript and Protein Evidence for Novel Small Protein Genes in *Populus*

Xiaohan Yang[1]* (yangx@ornl.gov), Gregory B. Hurst,[2] Abhijit A. Karve,[1] Timothy J. Tschaplinski,[1] Sara Jawdy,[1] Patricia K. Lankford,[2] Manesh B. Shah,[4] Gerald A. Tuskan,[1] Lee E. Gunter,[1] Christa Pennacchio,[4] and **Mitchel J. Doktycz**[3]

[1]Environmental Sciences Division, [2]Chemical Sciences Division, and [3]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tenn. and [4]DOE Joint Genome Institute, Walnut Creek, Calif.

http://PMI.ornl.gov

Project Goals: See goals for abstract 194.

Small proteins less than 200 amino acids in length encoded in short open reading frames (sORF) have major functions. Most of the small proteins characterized so far play important roles in cell-to-cell signal transduction. Hundreds or even thousands of novel sORF genes may await discovery in each plant species. The goal of this research is to systematically study the functional genomics of small proteins in relation to signal transduction involved in plant-microbe interactions. We hypothesize that small mobile proteins mediate long-distance signal transduction via the phloem/xylem channels. A systems biology approach is being used to realize three objectives: 1) discovery of sORF candidate genes using bioinformatics, transcriptomics and proteomics; 2) molecular characterization of the candidate genes using full-length gene cloning, sub-cellular localization and movement, and protein-protein interactions; and 3) functional characterization of the candidate genes using signal transduction assays and regulation of gene expression.

Our transcriptomics and proteomics research revealed thousands of sORF genes expressed in *Populus* leaf tissue under normal and drought conditions. Our recent comparative genomics analysis identified some interesting small protein candidates that are potentially involved in signal transduction via the phloem/xylem channels. Proteomics data were also analyzed for the presence of these small proteins. Protein extracts of leaves were prepared, and shotgun proteomics measurements were performed using an LC-MS-MS approach. By comparing experimental tandem mass spectra against a combined database containing current *Populus* gene annotations plus novel small protein candidates, we obtained proof-of-principle confirmation that small proteins are amenable to detection using these protocols. To identify more sORF candidate genes, sequencing of transcriptome and proteome in *Populus* phloem/xylem sap, as well as root system, including rhizosphere microorganisms (e.g., inoculated by fungus *Laccaria*), will be carried out.

For molecular characterization, small protein genes are being cloned as C- or N-terminal GFP (green fluorescent protein) and HA (hemagglutinin) tag constructs. The recombinant protein will be used for in vivo localization assays. The sub-cellular localization of the proteins will be studied by expressing the GFP-tagged constructs in the leaf mesophyll protoplasts. The inter-cellular long-distance movement of small proteins in plants will be monitored by imaging of fluorescent and radiolabeled protein probes. To characterize the role of the small proteins in mediating the plant microbe interactions, candidate genes will be analyzed by making transgenic *Populus* plants and/or hairy roots with altered gene expression (i.e., over-/down-regulation). In addition, the role of the small protein candidates will be studied using protoplast-based signaling assays.

201

Plant-Microbe Interfaces: Deciphering Plant-Microbe Signaling with Integrated Networks

David J. Weston[1]* (westondj@ornl.gov), Andrey Gorin,[2] Yunfeng Yang,[3] Sara S. Jawdy,[1] Abhijit A. Karve,[1] Jennifer Morrell-Falvey,[3] Gerald A. Tuskan,[1] and **Mitchel J. Doktycz**[3]

[1]Environmental Sciences Division, [2]Computer Science and Mathematics, and [3]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tenn.

http://PMI.ornl.gov

Project Goals: See goals for abstract 194.

The mechanisms by which symbiotic fungi engage host plant systems and induce local and systemic resistance

to pathogens, and promote growth and other beneficial effects remains poorly understood. A goal of our project is to develop experimental and computational methods to discover the signaling cascades and subsequent biochemical and molecular mechanisms driving these mutually beneficial interactions between host plants and fungal symbionts. A systems-level approach will be used to address this goal by: i) constructing a coexpression network map to characterize host plant transcriptional profiles in response to environmental perturbations with and without fungal symbionts; ii) developing computational methods to map regulatory and metabolic pathways implicated in plant-microbe mutualism; and iii) integrating proteome, metabolome, and transcriptome profiles to refine understanding of crucial genes and molecular mechanisms involved in these interactions.

To address these tasks, we have developed a model system where *Arabidopsis thaliana* was grown alone or co-cultivated with the root colonizing mycorrhiza-like fungi *Piriformospora indica.* An abiotic stress coexpression network map was created for *A. thaliana* that consisted of subnetworks (modules) enriched with gene products contributing to signaling, drought, heat, salt, and UV-B perturbations. We hypothesized that these modules would demonstrate robustness to environmental perturbation when *A. thaliana* was co-cultivated with *P. indica,* since this fungus has been shown previously to enhance host plant yield under abiotic stress conditions. After confirming that *P. indica* hyphae were associated with host plant roots using microscopy, plants with and without *P. indica* were subjected to ambient (25 ˚C) and heat (38 ˚C) treatments. Regardless of treatment, co-cultivation with *P. indica* induced gene expression in the heat shock and ROS (reactive oxygen species) signaling subnetworks. Co-cultivated plants exposed to the heat treatment maintained other subnetworks similar to ambient temperature expression levels, whereas heat treated plants cultured in the absence of the fungal symbiont had significant expression level changes for all subnetworks relative to ambient temperature controls. Biochemical quantification of ROS confirmed the network-derived signaling result and suggests that *P. indica* manipulates this signaling pathway to buffer host plants against abiotic perturbations.

To further refine our ability to discover biochemical and molecular mechanisms of plant - fungal mutualisms, expression changes in over 200 currently known co-expression modules reflecting tightly coordinated pathways in *A. thaliana* between inoculated and control samples are being systematically investigated. Rearrangements in the co-expression modules in response to mutualistic interactions, specifically genes that are co-expressed in the subnetworks, are being mapped. These genes will be targeted for further analytical studies as potential candidates involved in interspecific signaling pathways. Finally, a Bayesian statistical framework, calibrated on several well-studied *A. thaliana* molecular mechanisms to integrate proteome, metabolome and transcriptome data, is being used to decipher novel molecular mechanisms corresponding to the co-expression subnetworks that are found to be important in response to interactions with *P. indica.* The overall experimental and computational approach will be extended to future plant/microbiome communities investigated in this project.

202

Plant-Microbe Interfaces: Characterization of Native Microbial Communities in the Roots and Rhizosphere of *Populus deltoides*

Christopher W. Schadt[1]* (schadtcw@ornl.gov), Hector F. Castro-Gonzales,[1] Neil Gottel,[1] Dale A. Pelletier,[1] Jennifer Morrell-Falvey,[1] David J. Weston,[2] Rytas Vilgalys,[3] Gerald A. Tuskan,[2] and **Mitchel J. Doktycz**[1]

[1]Biosciences Division and [2]Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, Tenn. and [3]Dept. of Biology, Duke University, Durham, N.C.

http://PMI.ornl.gov

Project Goals: See goals for abstract 194.

The beneficial associations between plants and microbes exemplify complex, multi-organism systems that are shaped by the participating organisms and the environmental forces acting upon them. These plant–microbe interactions can benefit plant health and biomass production by affecting nutrient uptake, influencing hormone signaling, effecting water and element cycling in the rhizosphere, or conferring resistance to pathogens. The rhizosphere of *Populus* species represents an ideal system for understanding the natural diversity of these associations, as well as the molecular details that result in function. *Populus*, and more generally the Salicaceae (willow family) to which they belong, are host to bacterial endophytes and to two prominent types of mycorrhizal fungi, arbuscular endomycorrhizae (AM) formed by Glomalean fungi, and ectomycorrhizae (EM) formed by Ascomycetes and Basidiomycetes.

Broad-based efforts to understand the natural diversity of microbial associates of *Populus deltoides, P. trichocarpa,* and their natural and artifical hybrids in both native and controlled habitats have been initiated. As a pilot study for this work, a population of *P. deltiodes* as it occurs along the Caney Fork River was sampled in the early fall of 2009. Two *P. deltoides* stands were sampled, representing both upland and bottomland ecotypes and soil conditions that commonly occur in this region. This study is also serving as a "testbed" for methods development that will be applied more broadly for similar studies in the coming growing season. These efforts include four related foci: 1) Community assessment based on rRNA genes using pyrosequencing and other methods to describe the variation within and between individual trees in these two environments (described here); 2) Efficient cultivation, typing and physiological characterization of representative associates (described in poster by Pelletier et al.) that can be used as models for further molecular interaction studies; 3) Methods development for localization and quantification of microbial associates within

diverse *Populus* tissue-types; and 4) Methods development for single-cell manipulations that will compliment targeted cultivation, as well as metagenomic and metaproteomic efforts (described in poster by Retterer et al).

As part of this study, we are conducting 454 based pyrosequencing to describe both bacterial and fungal root and rhizosphere associations of the Caney Fork populations. Rhizosphere populations were examined from field samples that were washed in buffered saline. Mycrorrhizal and bacterial tissue-associated populations are being examined on the same surface sterilized root samples. Extensive efforts to optimize both surface sterilization methods and subsequent DNA extraction showed that efficient sterilization could be achieved by combining hydrogen peroxide and sodium hypochlorite based washing followed by commercially available DNA extraction methods. Using these methods, approximately 50-100ng DNA per mg root tissue was extracted and was readily amplifiable with PCR based approaches. Existing methods targeting the V1-V2 region of the 16S rRNA genes have shown that root and rhizosphere associated communities are extremely diverse, comprising thousands of OTUs per sample. Similar to past studies we show that these communities are dominated by alphaproteobacteria. Comparisons of variation within individual tree samples, between trees from similar environments, and between ecotypes are in process pending the completion of final samples from these collections. Our initial efforts in developing efficient methods for fungal community assessment have focused on descriptions of the D1/D2 region. This region can be targeted with conserved primer sets, allowing alignments over the entire diversity of fungi, but is also variable enough to allow for robust assessment of populations to the family level or below. Initial tests with multiple barcoded primer variants show that efficient and unbiased assessment may be achievable with the developed methods.

203

Plant-Microbe Interfaces: The Role of Plant Genotype and Phenotype in Regulating the Symbiotic Microenvironment

Timothy J. Tschaplinski[1]* (tschaplinstj@ornl.gov), Udaya Kalluri,[1] Lee E. Gunter,[1] Sara Jawdy,[1] Gerald A. Tuskan,[1] Maud Hinchee,[2] Jesse Labbé,[3] Francis Martin,[3] and **Mitchel J. Doktycz**[4]

[1]Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, Tenn.; [2]ArborGen LLC, Summerville, S.C.; [3]INRA, Nancy, Champenoux, France; and [4]BioSciences Division, Oak Ridge National Laboratory, Oak Ridge, Tenn.

http://PMI.ornl.gov

Project Goals: See goals for abstract 194.

A number of signals broadly classified as phytohormones (auxin, cytokinin and ethylene) and secondary metabolites (flavonols, alkaloids and polyamines) have been known to affect plant-microbe associations, however, the molecular underpinnings of how the signals are transduced into plant level phenotypic changes (e.g. lateral root proliferation, induction of systemically-acquired resistance) is poorly understood. Genotypic variability in plant carbon and nitrogen metabolism greatly alters both the intracellular and extracellular metabolite profiles and thus determines the biochemical microenvironment in which microbial (fungal and bacterial) symbionts exist. Our previous biochemical analyses of pure species of black cottonwood (*Populus trichocarpa*), eastern cottonwood (*P. deltoides*), their interspecific hybrids, other *Populus* sp. and their transgenic and cisgenic mutants provide a diverse variety of clones with tailored differences in primary and secondary metabolism that can be exploited to probe plant-microbe interactions. We hypothesize that enhanced production of metabolites of primary C and N metabolism in *Populus* roots, including simple sugars, organic acids, and amino acids, promotes colonization by ectomyccorhizal fungi and endophytic bacteria. We also hypothesize that unique secondary metabolites produced by *Populus* roots function as selective agents that both promote and inhibit specific microbial species. Here we present the current status of *Populus* genetic resources that have been characterized or created through transgenesis to provide the basis for future studies on *Populus*-microbe associations.

Populus trichocarpa (T) and *P. deltoides* (D) differ in their profiles of secondary metabolites and these metabolic differences segregate in successive generations such that quantitative trait loci associated with metabolite production (mQTL) can be mapped. We have created dense genetic marker maps and mapped several hundred mQTL for both leaf and root metabolites in two pseudo-backcross (TDxD') pedigrees, including family 13 and family 52124. The identification of 38 definitive and 117 suggestive mQTL for root metabolites in family 13 included the location of an mQTL hotspot on linkage group X that regulates the production of several key secondary metabolites, including salicortin, salireposide, and several phenylpropane glycosides (vimalin-like). Five candidate genes were selected within the mQTL region and constructs were designed for their up- and down-regulated expression. These cisgenic transformants with putatively increased and decreased secondary metabolism are scheduled to be available for characterization early in 2010 and will be available to test the effects of altered metabolism on plant-microbe associations. Additionally, the F2 backcross progeny with extremely high (and low) production of secondary metabolites that led to the identification of 239 mQTL associated with 105 metabolites in family 52124 are available and can also be selected to determine the effects of altered metabolism on plant-microbe associations.

We have shown that a key difference in secondary metabolites among *Populus* species is the nature and concentration of hydroxycinnamate-quinate/shikimate esters that are present and are likely to affect microbial colonization. The role of such metabolites in colonization can be assessed by the selection of *Populus* species with diverse profiles of such metabolites. Additionally, we are generating *P. deltoides*

clones that have been down-regulated for all genes in the lignin biosynthetic pathway, including hydroxycinnamate-quinate transferase (HCT) and coumarate-3-hydroxylase (C3H), that all have direct and indirect effects on the concentrations of these metabolites, given that they serve as storage compounds for the up-stream lignin precursors that have inhibitory effects on microbes. In addition to these clones, we have determined the metabolic phenotype of a number of activation-tagged *P. tremula* x *alba* clones that have elevated or greatly depleted concentrations of these metabolites. Microbial species that are successful colonizers must be able to tolerate the free mono- and diphenolic acids and their quinate/shikimate esters. These *Populus* clones can serve as the background plant material to determine what classes of compounds promote or inhibit key plant-microbe associations.

We have selected the early colonization events of the *Populus-Laccaria* association, which is the model perennial tree-fungal association with both organisms having had their genomes sequenced and thus have available a broad array of genomic resources. *Laccaria bicolor* is an ectomycorrhizal fungus that routinely colonizes *Populus*. The metabolic signaling responses involved in the establishment of the association is being characterized by analyzing the time-course of metabolomic and transcriptomic (array-based) responses of both organisms reared under *in vitro* culture, contrasted with greenhouse pot culture with samples in both studies collected every 2 weeks up to 8 weeks in the in vitro study, and up to 12 weeks in the pot culture study. Additionally, the role that the poplar genotype has in the metabolic responses involved in the establishment of the plant-microbe association is being investigated by altering the *Populus* host, including *P. deltoides*, *P. trichocarpa*, and three *P. trichocarpa x deltoides* hybrids. Once putative signals are identified, their roles will be assessed by initiating assays of indirect contact between *Populus* roots and *Laccaria in vitro* (i.e., a cellophane membrane allows molecular cross-talk through diffusible metabolites, such as auxins, without physical contact) at early time points. Similarly, the effects of the direct contact of a limited number of putative signals at very early time points will also be studied. The metabolite signals and microarray responses will be correlated in co-expression networks to identify novel signaling pathways that regulate the *Populus-Laccaria* association.

Recent reports suggest that root growth induced in *Populus* interacting with *Laccaria* required polar auxin transport as well as auxin signaling through *Populus* auxin response regulator proteins. Many sequenced microbial genomes carry genes of auxin biosynthetic pathway, but it is unclear to what extent these signaling pathways are universal or specific in establishing a symbiotic relationship between a specific microbe and host genotype. We hypothesize that certain auxin response factor proteins belonging to the *Aux/IAA* and *ARF* families play a direct role in establishment and /or signal transduction post-establishment. We are testing this hypothesis by co-culturing specific microbial strains with PCR-confirmed *Populus* RNAi lines specific to genes from the *Aux/IAA* and *ARF* gene families. A micropropagation protocol established to generate whole plantlets from *Populus* shoot tips is being used to also test candidate endophytic and rhizosphere microbes identified from *Populus* field surveys and will be tracked by imaging and molecular profiling methods. The efforts to identify the molecular factors are being complemented with measurements of the levels and impact on altered levels of auxin in co-cultivation experiments. These plant materials will be harvested at various timepoints and characterized at the transcriptomic and proteomic levels in order to understand the mechanisms of hormone cross-talk relaying the plant level outcomes due to the microbial association. This presents a pipeline for micro-scale screening of where, when, and how microbes associate with the host plant.

204

Plant-Microbe Interfaces: Isolation and Characterization of Cultivable Members of the *Populus* Rhizosphere-Endosphere Community

Dale A. Pelletier[1]* (pelletierda@ornl.gov), Tse-Yuan Lu,[1] Se Yeon Kim,[1] Christopher W. Schadt,[1] Marilyn Kerley,[1] Timothy J. Tschaplinski,[2] David J. Weston,[2] Jennifer Morrell-Falvey,[1] Amy L. Schaefer,[3] E. Peter Greenberg,[3] Caroline S. Harwood,[3] Scott T. Retterer,[1] and **Mitchel J. Doktycz**[1]

[1]Biosciences Division and [2]Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, Tenn.; [3]Dept. of Microbiology, University of Washington, Seattle

http://PMI.ornl.gov

Project Goals: See goals for abstract 194.

As part of the newly initiated ORNL Plant-Microbe Interfaces Science Focus area, we are using broad-based methods to characterize the natural diversity of microbial associates of *Populus* and elucidate the molecular mechanisms by which these diverse organisms interact. The goal of the isolation and screening efforts for this project is to develop methods for efficient cultivation, typing, and physiological characterization of representative associates that can be used as model organisms for further molecular interaction studies. In a pilot study for this work, we sampled a population of *P. deltoides* as it occurs along the Caney Fork River in Tennessee in the early fall of 2009. Two *P. deltoides* stands were sampled, representing both upland and bottomland ecotypes and soil conditions that commonly occur in this region. From these samples, a number of diverse fungal and bacterial associates have been isolated from *Populus* rhizosphere and surface sterilized root tissues using broadly compatible media and direct plating methods. These isolates are being screened for phylogenetic identity with bacterial and fungal rDNA primers, and for traits of interest involved in plant microbe interactions using molecular and biochemical assays, including nitrogen fixation (*nifH*), indole-3-acetic acid synthesis pathways (*iaa*), salicylate degradation genes (*nahJ* and *nahW*), and fungal oxalate production (*oahA*).

The exchange of chemical signaling molecules through either direct contact or diffusion has been demonstrated in a few well studied plant-microbe model systems to effect establishment and maintenance of symbiotic interactions. These signals (e.g., phytohormones, antimicrobials, quorum sensing compounds) affect a wide range of phenotypic responses that can influence plant-microbe and microbe-microbe interactions, including production of antimicrobials, exopolysaccharides, exoenzymes, motility, and conjugation. Microbes isolated from the *Populus* rhizosphere and endosphere are being screened for the production of small signaling molecules by GC-MS metabolomics of culture supernatants from isolates grown in the presence and absence of plant-derived exudates. Isolates are also being screened for known and novel homoserine lactone-derived quorum sensing compounds.

These cultivation and screening efforts are being complemented with improved methods for resynthesizing microbial relationships with *Populus* in greenhouse and tissue culture based assays that are critical components of proving the function of these microbial associates. In addition to conventional plate-based isolation methods, efforts are focused on single-cell microbial isolation and characterization techniques and the development of relevant technologies that will ultimately facilitate the genetic characterization of a greater portion of the endophytic and rhizospheric communities. We anticipate that multiple new microbial species will be characterized and advanced toward further studies.

205

Plant-Microbe Interfaces: Application of Microfluidic Technologies for Microbial Isolation, Cultivation, Characterization, and Emulation of the Plant-Rhizosphere Microenvironment

Scott T. Retterer[1,2*] (rettererst@ornl.gov), Meena Kalyanaraman,[1] A. Nicole Edwards,[3] Jennifer L. Morrell-Falvey,[1,3] Tim McKnight,[4] Steve L. Allman,[1] James G. Elkins,[1] Gladys Alexandre,[3,5] Martin Keller,[1] and **Mitchel J. Doktycz**[1,2,3]

[1]Biosciences Division, [2]Center for Nanophase Materials Sciences, and [4]Measurement Science and Systems Engineering Division, Oak Ridge National Laboratory, Oak Ridge, Tenn.; [3]Graduate School of Genome Science and Technology, University of Tennessee-ORNL, Knoxville; [5]Dept. of Biochemistry, Cellular, and Molecular Biology, University of Tennessee, Knoxville

http://PMI.ornl.gov

Project Goals: See goals for abstract 194.

Isolation, cultivation and characterization of microbes from complex samples, derived from soil, rhizosphere and plant material, are essential for defining the plant microbiome. Technologies that allow the physical and chemical manipulation of fluid volumes that approach the single cell level are critical to such processes and offer a means to understand the full range of genetic and phenotypic variation within plant associated microbial and mycorrhizal communities. Microfluidic platforms, created using soft lithographic and advanced nanofabrication techniques, have been developed and their utility for physically isolating single cells in microbead emulsions and creating complex chemical gradients for the study of microbial taxis has been demonstrated.

Micro-scale alginate beads were created using a microfluidic multiphase droplet generation system. By combining a solution of model cells within the aqueous/alginate phase, the system was successfully used to encapsulate individual and small numbers of GFP-expressing *E.coli*. Both internal and external gelation processes were examined for their effects on cell viability and bead stability. The successful isolation of individual cells within the alginate microbeads and subsequent cultivation of these isolates demonstrates the potential of this system for the small scale clonal expansion of microbial cells captured from more complex mixtures of cells taken from natural samples. As this technology is developed, its implementation in the field and application to the characterization of field samples will play a critical role in assessing the genetic and phenotypic diversity present within the *Populus* microbiome.

Once microbial species are isolated via conventional or emerging cultivation techniques, characterization of their phenotypic response to specific molecular signals found within the plant rhizosphere may be correlated with fluctuations in their population. Such fluctuations may correspond to both developmental and phenotypic stages of plant growth. Understanding microbial taxis in response to components of plant exudate/xylem sap, characterized via mass spectrometry during such stages, may provide insight into what role cell motility and taxis plays on microbial recruitment and colonization. A microfluidic platform that allows exquisite spatial and temporal control of chemical gradients while allowing sustained imaging and tracking of individual bacteria has been developed. This system will be used to examine taxis of model microbes and natural isolates as they become available.

The Predictive Microbial Biology Consortium

206

Genemap-MS: Stable Isotope Assisted Metabolite Profiling of *Synechococcus sp. 7002*

Richard Baran* (RBaran@lbl.gov), Benjamin P. Bowen, Steven M. Yannone, and **Trent R. Northen**

ENIGMA SFA, Lawrence Berkeley National Laboratory, Berkeley, Calif.

http://www.calccme.org/

Project Goals: Development of a metabolomics-centric platform for comprehensive validation and expansion of genome annotations.

Computational homology-based annotations of sequenced genomes of microbes provide an overview of their metabolic capabilities. Inherent uncertainties in homology-based functional annotations, presence of a significant fraction of genes of unknown function in annotated genomes along with a large number of enzymatic activities without a known corresponding gene limit the extent of genome annotations. Comprehensive profiling of cellular metabolites offers an attractive opportunity for the validation and expansion of genome annotations since the presence of specific metabolites indicates the presence of related enzymatic activities.

The first essential prerequisite for the exploitation of metabolites as indicators of enzymatic activities is the ability to identify these metabolites in complex metabolite profiles. Liquid chromatography coupled to electrospray time-of-flight mass spectrometry (LC-ESI-TOFMS) provides chromatographic separation among metabolites in complex mixtures and mass spectra with high mass accuracy. Metabolites can be identified by comparing their accurate mass, retention time, and fragmentation (MS/MS) spectra against chemical standards. Complex metabolite profiles often contain numerous features which do not correspond to any available and analyzed chemical standard. These features can be analyzed by assigning putative empirical formulas based on accurate mass and isotopic profiles (applicable to small compounds - m/z < ~200) and inferring partial structural information from MS/MS spectra.

To expand the unambiguous assignment of empirical formulas to larger metabolites (up to m/z ~500), uniform labeling of *Synechococcus sp. 7002* cultures was performed with stable isotopes 13C and 15N. Characteristic shifts in masses of features in 13C- and 15N-labeled metabolite profiles discriminated background signals from metabolites originating from the metabolism of *Synechococcus sp. 7002*. Additionally, these shifts determined the number of carbons and nitrogens in specific metabolites thus discriminating between alternative candidate empirical formulas, which were determined from accurate mass of features in an unlabeled control dataset only. This approach facilitated the inference of a large number of putative compounds with validated empirical formulas. A fraction of these was assigned to specific metabolites based on the correspondence of their accurate mass and retention time to chemical standard.

Draft metabolic network of *Synechococcus sp. 7002* was reconstructed using Pathway Tools software (BioCyc). This reconstruction is based on genome annotation but also accounts for additional enzymatic activities or spontaneous reactions inferred from the topology of the metabolic network (gap filling according to reference metabolic pathways). Comparison of the draft network against the set of confirmed empirical formulas showed that many of the unique empirical formulas have no correspondence to any of the 805 metabolites in the draft network. Moreover, a number of the confirmed empirical formulas do not correspond to any metabolite in MetaCyc or KEGG databases.

Following the analysis of MS/MS spectra, the identity of a subset of empirical formulas without any correspondence in MetaCyc or KEGG could be assigned to dipeptides of glutamate (at the N-terminus) with one of multiple hydrophobic amino acids. An intermediate of an alternative biosynthetic pathway, with only one of the pathway's reactions assigned to a gene in the genome, was also identified using MS/MS spectra. Another interesting finding is what appears to be a dead-end metabolite according to KEGG database. This metabolite is present predominantly in the metabolite profile of the culture media, which is significantly less complex that the profile of the cell extract.

Metabolite profiling is an attractive approach for comprehensive interrogation of cellular metabolism. Presence of specific metabolites may serve as an indicator of the presence of specific enzymatic activities or metabolic pathways. Comprehensive metabolite profiling may thus serve as the first step in the validation of genome annotations and identification of candidate enzymatic activities or pathways missing from the genome annotation. Combination of metabolite profiling with genetic/environmental perturbations and transcriptomics/proteomics may further zoom in on specific genes related to specific enzymatic activities or metabolic pathways.

207

Applications of GeoChip for Analysis of Different Microbial Communities

Joy D. Van Nostrand[1,2]* Liyou Wu,[1] Patricia Waldron,[1] Ping Zhang,[1] Ye Deng,[1] Zhili He,[1,2] Weimin Wu,[3] Sue Carroll,[4] Chris Schadt,[4,2] Anthony Palumbo,[4] Dave Watson,[4] Craig Criddle,[3] Phil Jardine,[4] Terry C. Hazen,[5,2] and **Jizhong Z. Zhou**[1,2] (jzhou@ou.edu)

[1]University of Oklahoma, Norman; [2]VIMSS (Virtual Institute of Microbial Stress and Survival) http://vimss.lbl.gov/; [3]Stanford University, Stanford, Calif.; [4]Oak Ridge National Laboratory, Oak Ridge, Tenn.; and [5]Lawrence Berkeley National Laboratory, Berkeley, Calif.

Project Goals: As part of the GTL program, our research is focused on expanding and improving the GeoChip, a comprehensive functional gene array, and to use this array to detect and monitor microbial communities.

Microarray technology provides the opportunity to identify thousands of microbial genes or populations simultaneously. The objective of this study was to further develop and apply a comprehensive functional gene array (GeoChip) to detect and monitor microbial communities. GeoChip 2.0 is a functional gene microarray which allows for the simultaneous detection of >10,000 genes involved in the geochemical cycling of C, N, and S, metal reduction and resistance, and organic contaminant degradation. Based on GeoChip 2.0, a new generation, GeoChip 3.0 has been developed, which has several new features. First, GeoChip 3.0 covers more gene groups including antibiotic resistance, energy processing, and additional functional genes involved in C, N, P, and S cycling. Second, the homology of automatically retrieved sequences by key words is verified by HUMMER using seed sequences so that unrelated sequences are removed. Third, a universal standard has been implemented so that data normalization and comparison of different microbial communities can be conducted. Fourth, a genomic standard is used to quantitatively analyze gene abundance. In addition, GeoChip 3.0 includes phylogenic markers, such as gyrB. Finally, a software package has been developed to facilitate management of probe design, data analysis, and future updates. This version of GeoChip was a recipient of a 2009 R&D100 Award from *R&D Magazine*, which recognizes the 100 most innovative scientific and technical breakthroughs of the year. Additional expansion is currently underway and will include genes related to stress response, virulence factors, human-microbiome, and phage genes.

The GeoChip has been used to examine dynamic functional and structural changes in microbial communities from many different environments. Here, examples of studies utilizing the GeoChip to examine microbial communities at contaminated sites are presented. These studies illustrate the ability of the GeoChip to provide direct linkages between microbial genes/populations and ecosystem processes and functions. These three studies examined areas within the U.S. DOE's Field Research Center (FRC) in Oak Ridge, TN. (1) Microbial communities within a pilot-scale test system established for the biostimulation of U (VI) reduction in the subsurface by injection of ethanol were examined using GeoChip 3.0. Functional community dynamics were examined during a period of nitrate exposure. After exposure to nitrate the diversity and richness increased several fold but quickly returned to pre-nitrate levels. Detrended correspondence analysis (DCA) indicated a shift in the overall community structure after nitrate exposure but the community began to return to pre-exposure structure once nitrate was removed. The relative abundance of several nitrogen cycling genes showed an increase immediately after nitrate exposure, including ammonification, denitrification, and nitrogen fixation genes indicating a stimulation of these communities.

(2) In the second study from the FRC, analysis of groundwater monitoring wells along a contamination gradient using GeoChip 2.0 revealed less overlap between wells with different levels of U and NO_3 contamination. While diversity of nitrate-fixation genes decreased in NO_3 -contaminated wells, the diversity of metal reduction and resistance genes did not correlate with metal concentrations. Signal intensity did, however, increase in heavily contaminated wells, indicating a larger percentage of organisms with metal- related genes. Sulfate-reduction genes had greater diversity and greater signal intensity in more contaminated wells. Individual principle component analyses (PCA) of the gene diversity and geochemistry of the wells separated them in similar ways. CCA indicated that pH was an important variable that correlated with gene diversity in the lowest-contamination well, while NO_3 and U correlated with the most highly contaminated well. Overall, contaminant level appears to have significant effects on the functional gene diversity along the contaminant plume at the FRC.

(3) A third study is currently underway using GeoChip 3.0 to examine functional gene changes in a U (VI) contaminated area after introduction of a slow-release-substrate (SRS), designed to provide a long-term electron donor for U (VI) reduction. Preliminary results indicate a stimulation of microbial communities. These studies demonstrate the analytical power of the GeoChip in examining microbial communities. This is the first comprehensive microarray available for studying the functional and biogeochemical cycling potential of microbial communities.

* Presenting author

208

Pipeline for Large-Scale Purification and Identification of *Desulfovibrio vulgaris* Membrane Protein Complexes

Peter J. Walian[1]* (PJWalian@lbl.gov), Simon Allen,[2] Lucy Zeng,[1] Evelin Szakal,[2] Eric Johansen,[2] Haichuan Liu,[2] Steven C. Hall,[2] Susan J. Fisher,[1,2] Mary E. Singer,[1] Jil T. Geller,[1] Swan Lin,[1] Terry C. Hazen,[1] H. Ewa Witkowska,[2] **Mark D. Biggin**,[1] and Bing K. Jap[1]

[1]Lawrence Berkeley National Laboratory, Berkeley, Calif. and [2]University of California, San Francisco

http://pcap.lbl.gov

Project Goals: To develop and apply a pipeline for the high-throughput isolation and identification of *Desulfovibrio vulgaris Hildenborough* membrane protein complexes in cultures grown under standard conditions, and to characterize changes in these complexes brought about by environmentally relevant stressors.

As a component of the LBNL ENIGMA Program, an important task of the Protein Complex Analysis Project (PCAP) is to develop and apply methodologies for the identification of membrane protein complexes isolated from the sulfate reducing bacterium *Desulfovibrio vulgaris Hildenborough (D. vulgaris)*. Given its demonstrated ability to reduce heavy metals, *D. vulgaris* is projected to play an important role in the management of contaminated sites. To optimize the use of *D. vulgaris* at these sites it will be critical to know how environmental changes affect its performance. To gain insight into these questions, we have developed a pipeline to isolate and identify stable membrane protein complexes present in cultures grown under standard conditions, stationary phase conditions, and in the presence of environmental stressors. Through these efforts we hope to assemble the data needed to characterize stress-induced changes in the relative abundance, composition, and stoichiometry of *D. vulgaris* membrane protein complexes. This data, in turn, will be used to model its stress response pathways and optimize the bioremediation capabilities of this microbe.

Membrane protein complexes pose unique purification and analysis challenges. Largely due to the requirement for detergent solubilization, stable isolation of homogeneous intact membrane protein complexes typically requires separation conditions that are different from those used for water soluble proteins. For this task we have been employing a "tagless" strategy optimized for purifying membrane proteins and then identifying them by mass spectrometry (MS). As opposed to strategies employing affinity tags for the purification of target molecules, use of a tagless strategy will enable us to obtain global views of stress-induced changes involving membrane proteins in *D. vulgaris* cultures grown under a variety of conditions.

In the pipeline, *D. vulgaris* cell membranes isolated from large-scale (100 liter) cultures are first treated with a relatively mild detergent suited for the extraction of inner-membrane proteins. The residual membranes of this gram-negative bacterium are subsequently treated with a second more active detergent to solubilize proteins of the outer-membrane. Each membrane extract is then independently processed. To purify candidate complexes of the inner- and outer-membrane fractions, ion exchange (IEX) and molecular sieve chromatography are used. Fractions obtained from these procedures are further analyzed using SDS and blue native gel electrophoresis to isolate candidate complexes and obtain molecular weight estimates. To prepare samples suitable for MS analysis, whole lanes are cut from blue-native PAGE gels, placed horizontally along the stacking sections of denaturing gels and subjected to a second dimension of SDS PAGE. Potential complex subunits manifest themselves as bands or spots providing insight into the composition of the native complex. Spots removed from these gels are subjected to in-gel digestion and analysis by liquid chromatography electrospray ionization tandem mass spectrometry (LC-ESI-MS/MS) or liquid chromatography and matrix assisted laser desorption time of flight mass spectrometry (LC-MALDI-MS/MS). Protein identification is achieved by searching a custom *D. vulgaris* database using the Mascot or Protein Pilot search engine.

We are currently completing a baseline membrane protein complex dataset derived from *D. vulgaris* large-scale cultures grown under standard conditions (mid-log phase). In addition to providing a catalog of *D. vulgaris* membrane protein complexes, this data will serve as an essential reference for the detection and characterization of changes in the complexes of cultures subjected to the aforementioned stressors.

209

High Throughput Identification, Purification and Structural Characterization of Soluble Protein Complexes in *Desulfovibrio vulgaris*

Bong-Gyoon Han[1]* (BGHan@lbl.gov), Haichuan Liu[2]* (haichuan.liu@ucsf.edu), Ming Dong[1]* (MDong@lbl.gov), Maxim Shatsky,[1,3] Steven E. Brenner,[1,3] Pablo Arbelaez,[3] Jitendra Malik,[3] Dieter Typke,[1] Terry C. Hazen,[1] Jil T. Geller,[1] Harry J. Sterling,[3] Lee Yang,[1] Megan Choi,[1] Evelin D. Szakal,[2] Simon Allen,[2] Steven C. Hall,[2] Susan J. Fisher,[1,2] Evan R. Williams,[3] John-Marc Chandonia,[1] Jian Jin,[1] H. Ewa Witkowska,[2] Robert M. Glaeser,[1] and **Mark D. Biggin**[1]

[1]Lawrence Berkeley National Laboratory, Berkeley, Calif.; [2]University of California, San Francisco; and [3]University of California, Berkeley

Project Goals: One of the main goals of the PCAP component of the LBNL ENIGMA SFA is to develop high throughput pipelines to purify and identify protein complexes and structurally characterize them by EM. *Desulfovibrio vulgaris* was selected as a model bacterium because of its bioremediation potential in immobilizing

toxic heavy metals in soils. Our overall workflow consists of (i) a multidimensional separation of the soluble proteome using tagless strategy; (ii) identification of putative complexes by quantitative mass spectrometry (MS); (iii) further purification of selected complexes; and (iv) structural analysis by EM. In some cases, the stoichiometry of protein complexes is studied using novel mass spectrometry techniques that enable analysis of non-covalent assemblages (native MS). Our major focus is to increase the throughput of MS and EM single-particle analyses and to develop methods for an automated assignment of protein complex components. Currently we are implementing automated collection of EM data, particle boxing, and analysis of structural variation, and the engineering of new support-film technologies for better EM sample preparation to preserve quaternary structure in a conformationally homogeneous state. In the future, we will also take advantage of the SAXS expertise within ENIGMA and incorporate this technology into our structural characterization of complexes.

One of the main goals of the PCAP component of the LBNL ENIGMA SFA is to develop high throughput pipelines to purify and identify protein complexes and structurally characterize them by EM. *Desulfovibrio vulgaris* (*Dv*H) was selected as a model bacterium because of its bioremediation potential in immobilizing toxic heavy metals in soils at the DOE sites by rendering them practically insoluble upon conversion to the low red-ox state. Our overall workflow consists of (i) a multidimensional separation of the soluble proteome; (ii) identification of putative complexes by quantitative mass spectrometry (MS) followed by bioinformatics analysis; (iii) further purification of selected complexes; and (iv) structural analysis by EM. In some cases, the stoichiometry of protein complexes is studied using novel mass spectrometry techniques that enable analysis of non-covalent assemblages (native MS). In future, we will also take advantage of the SAXS expertise within ENIGMA and incorporate this technology into our structural characterization of complexes.

To identify protein complexes, we have introduced and established a tagless strategy, which is based on the premise that the great majority of stable protein complexes will survive intact separation through a series of orthogonal chromatographic methods. Under this scenario, co-migrating polypeptide components of protein complexes generate overlapping elution profiles at each stage of protein separation. Changes in the relative concentration of each polypeptide (i.e., their elution profiles across the final chromatographic step) are measured with the aid of isotopic dilution mass spectrometry (MS) and iTRAQ reagents (Dong et al., 2008). Various data analysis approaches are being developed to automate assignment of the identified polypeptides to putative complexes. To date, ~64% of the target protein complex separation space has been analyzed, resulting in the identification of over 900 polypeptides. The great majority of the polypeptides are engaged in intermolecular interactions, as evidenced by more than 70% demonstrating significantly higher elution volume (at least 2x) in size exclusion chromatography (SEC) than expected from the molecular weight of the polypeptide predicted from genome sequence. In addition to heteromeric complexes, the tagless strategy allows detection of homomers, which are not easy to recognize by other methods (*e.g.,* TAP). Overall, at least 45 heteromeric and over 550 homomeric complexes have been identified so far. To address the challenges posed by a co-elution of non-related polypeptides with the legitimate components of complexes, we are evaluating monitoring elution profiles at the two final stages of protein complex separation (hydrophobic interaction chromatography and SEC), as opposed to the single step (SEC) used so far. The resulting 2D polypeptide elution map is expected to provide higher resolution data and consequently to increase confidence in protein complex assignments.

To determine molecular structures, we selected 16 complexes identified by the tagless strategy with molecular weights 400 - 1,000 kDa and subjected them to single-particle EM analysis. Half of the complexes studied proved stable enough to produce high-quality 3-D reconstructions with a resolution of ~2 nm (Han et al., 2009). This success rate for obtaining structures is about 10 times greater than that of previous "proteomic" screens. We have found that there are a surprisingly large number of differences in the quaternary structures of complexes isolated from *Dv*H compared to those of homologous proteins from other microbes. These differences occur so frequently that structures determined for complexes in other micro-organisms are likely to be inadequate as templates for modeling the biochemical networks within a given microbe of interest. By extension, we suspect that it may also be the case that complexes change structure frequently under different physiological conditions and future work will address this possibility. Our major focus now is to increase the throughput of EM single particle structural analysis. This effort currently includes the implementation of automated data collection, particle boxing, and analysis of structural variation (Shatsky et al., 2009), and the engineering of new support-film technologies for better EM sample preparation to preserve quaternary structure in a conformationally homogeneous state.

References

1. Dong *et al.,* 2008, A "tagless" strategy for identification of stable protein complexes genome-wide by multidimensional orthogonal chromatographic separation and iTRAQ reagent tracking. *J Proteome Res.* 7:1836-49.
2. Han *et al.,*. 2009. Survey of large protein complexes in *D. vulgaris* reveals great structural diversity. *PNAS*. 106:16580-16585.
3. Shatsky *et* al., 2009. A method for the alignment of heterogeneous macromolecules from electron microscopy. *J. Struct. Biol.* 166: 67-68.

210

Protein Complex Analysis Project (PCAP): Large-Scale Identification of Protein-Protein Interactions in *Desulfovibrio vulgaris* Using Tandem-Affinity Purification

Swapnil Chhabra,[1] Gareth Butland,[1]* Dwayne Elias,[2] Sonia Reveco,[1] Veronica Fok,[1] Barbara Gold,[1] Thomas Juba,[2] John-Marc Chandonia,[1] Ewa Witkowska,[3] Terry Hazen,[1] Judy Wall,[2] Jay Keasling,[1,4] and **Mark Biggin**[1] (mdbiggin@lbl.gov)

[1]Lawrence Berkeley National Laboratory, Berkeley, Calif.; [2]University of Missouri, Columbia; [3]University of California, San Francisco; and [4]University of California, Berkeley

Project Goals: Develop a high-throughput platform for chromosomal modifications of microbes of interest to DOE. Identify protein-protein interactions in the model sulfate reducer *Desulfovibrio vulgaris* using tandem affinity purification. Develop computational tools and models to characterize sulfate-reducing bacteria.

Most cellular processes are mediated by multiple proteins interacting with each other in the form of multi-protein complexes and not by individual proteins acting in isolation. One of the major goals of LBNL ENIGMA SFA is to develop a comprehensive knowledgebase of protein complexes and protein-protein interactions (PPI) in microbes of interest to DOE. As part of the Protein Complex Analysis Project component of ENIGMA, *Desulfovibrio vulgaris,* a sulfate reducing bacterium (SRB) found to exist in several DOE waste sites, has been used as a model to screen for PPI using a high throughput tandem affinity purification (TAP) approach. The challenges in working with *D. vulgaris* are typical of organisms of interest to DOE. While likely the best characterized SRB, *D. vulgaris* is an obligate anaerobe and possessed very limited genetic tools. Our approach therefore required the development of a high throughput pipeline to enable the creation of a library of genetically engineered strains, which builds upon generic principles of recombination of non-replicating gene replacement constructs ("suicide" constructs). We have applied this pipeline to the creation of *D. vulgaris* strains bearing TAP-tagged alleles designed for tandem affinity purification of individually tagged bait proteins; however, the versatility of this approach enables its potential application for chromosomal modifications of the majority of microbes of interest to DOE.

Previously we reported the development of large-scale single- and double- crossover chromosomal integration platforms for generating TAP-tagged strains of *D. vulgaris*. Transformation of plasmids into *D. vulgaris* is inefficient and multiple strategies, including constructing a restriction endonuclease host strain mutant that enhanced transformation of stable plasmids into *D. vulgaris* have been explored. However the single greatest improvement in transformation and integration of suicide constructs has resulted from the transition from the TOPO-Gateway® scheme to the Sequence and Ligation Independent Cloning (SLIC) technique for suicide construct generation. Success rates for suicide construct generation improved from 50% to more than 80%, transformation and integration of constructs into *D. vulgaris* from 34% to 65% and ~60 % of these isolates were found to express a TAP-tagged fusion protein detected by IP-western. Currently, we have a library of over 700 suicide plasmid constructs which have been employed for the generation of over 300 tagged SRB strains of which ~200 baits have been analyzed by TAP to date.

Here we present details of our high throughput pipeline for strain engineering along with results of the subsequent TAP analysis of these engineered strains. The interactions presented cover a range of biological processes, including energy conservation (ATP Synthase, Ech hydrogenase), sulfate reduction (dissimilatory sulfite reductase, adenylylsulfate reductase) and protein secretion (YajC-HflCK complex), and include both novel and previously predicted interactions.

211

Systems Approach in a Multi-Organism Strategy to Understand Biomolecular Interactions in DOE-Relevant Organisms

Sung Ho Yoon[1]* (syoon@systemsbiology.org), Christopher Bare,[1] David Reiss,[1] Dan Tenenbaum,[1] Min Pan,[1] Joseph Slagel,[1] Sujung Lim,[2] June Burn,[2] John A. Leigh,[2] Murray Hackett,[3] Angeli Lal Menon,[4] Michael W.W. Adams,[4] Sunia A. Trauger,[5] Gary Siuzdak,[5] Steven M. Yannone,[6] Benjamin Bowen,[6] Stephen R. Holbrook,[6] John A. Tainer,[6] and **Nitin S. Baliga**[1] (nbaliga@systemsbiology.org)

[1]Institute for Systems Biology, Seattle, Wash.; [2]Dept. of Microbiology and [3]Dept. of Chemical Engineering, University of Washington, Seattle; [4]Dept. of Biochemistry and Molecular Biology, University of Georgia, Athens; [5]Center for Mass Spectrometry, Scripps Research Institute, La Jolla, Calif.; [6]Dept. of Molecular Biology, Lawrence Berkeley National Laboratory, Berkeley, Calif.

http://gaggle.systemsbiology.net/projects/doe-archaea/2007-04/
http://maggie.systemsbiology.net
http://baliga.systemsbiology.net

Project Goals: Bolster through high-end state-of-art systems approaches, developed specifically for the study of archaeal organisms, the comprehensive analysis of multi-protein complexes in DOE-relevant organisms.

Rational re-engineering of biology for the purpose of bioremediation, bioenergy or C-sequestration requires deep understanding of all functional interactions of relevant components within native cell (s). Many of these functional interactions are conserved across diverse species to different

degrees depending on their evolutionary distance. We are conducting integrative analysis of genomic architecture and composition, transcriptome and proteome structure/function, protein-protein and protein-DNA interactions and metabolic networks to find keystone complexes and specialized circuit architectures for important application-relevant genes within four archaeal organisms. These organisms have enormous potentials from the standpoint of H_2 production, N_2 fixation, and C-sequestration; they include an anaerobic thermophile (*Pyrococcus*), an acidophilic and aerobic thermophile (*Sulfolobus*); a hydrogenotrophic methanogen (*Methanococcus*), and a photoheterotrophic halophile *Halobacterium* NRC-1. A key aspect of our approach is to use the power of systems biology to delineate the process of nucleation, assembly, and turnover of key complexes. Here, we report comparative analysis of dynamically changing transcriptome structures of the four archaea with special emphasis on the conditional activation of unconventional transcriptional promoters within conserved genes and operons.

Note: Computational and experimental results from this study will be freely available upon publication at http://maggie.systemsbiology.net/. All of the software tools developed in this project have been made freely available at http://gaggle.systemsbiology.net/projects/doe-archaea/2007-04/.

This research was supported by U.S. Department of Energy, Award No. DE-FG02-07ER64327 and DG-FG02-08ER64685.

212

Metabolic Transformations and Chemical Differences

Ben Bowen* (BPBowen@lbl.gov), Richard Baran, Steve Yannone, John Tainer, and **Trent Northen**

ENIGMA SFA, Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, Calif.

http://www.calccme.org

Project Goals: Use autocorrelation for the analysis of metabolites using liquid chromatography coupled to mass spectrometry.

Defining the complete metabolic reconstruction for even a simple organism requires detailed knowledge of enzyme function, activity, and biochemistry. Uncharacterized enzymes and unknown metabolic pathways limit the utility of inferring metabolic capability based on genome sequencing and annotation. Homology-based protein function predictions are limited by evolutionary processes that result in conserved domains, and the complexity of biochemistry is so diverse that only a small portion has been defined. Therefore, improving genome annotations and furthering our knowledge of biochemistry is of tremendous importance to take advantage of the potential of genome sequencing.

The detection and analysis of metabolites using liquid chromatography coupled to mass spectrometry (LC/MS) has the potential to define metabolic pathways, understand the regulation of substrate utilization, measure biomass composition, and much more. The primary challenge associated with this analytical platform is the interpretation of the thousands of molecular features detected in a typical dataset. There are three hurdles which define this crux: 1) a large fraction of the detected features are associated with uncharacterized metabolites, 2) standard methods for the detection of known metabolites are not widespread, and 3) optimum methods have not been defined for feature identification from raw data.

Although the scale biochemistry is vast, the emergence of primary and secondary metabolites is due to a limited number of elementary chemical-differences relating compounds. In this work, we explore a finite number of chemical transformations, and examine chemical-difference-space to understand alterations in metabolism. This is accomplished using autocorrelation, an established method in data analysis with a very high sensitivity for detecting correlated relationships. Here, we have applied autocorrelation to interpret raw mass spectrometry data and then linked the global correlation spectra to chemical differences. This has enabled a global analysis of LC/MS data without relying on feature identification. This analysis was applied to *Sulfolobus sulfataricum*, an archaea that is of interest in creating biofuels, and early results indicate that discrete chemical differences can be measured and interpreted.

213

Towards Localization of Functionality in *Desulfovibrio vulgaris* by Electron Microscopy

David A. Ball,[1] Swapnil Chhabra,[1] Dwayne Elias,[2] Veronica Fok,[1] Jil T. Geller,[1] Amita Gorur,[1] **Terry C. Hazen**,[1] Danielle Jorgens,[1] Thomas Juba,[2] Ambrose Leung,[1] Jonathan Remis,[1] Mary E. Singer,[1] Andrew Tauscher,[1] **Judy Wall**,[2] **Manfred Auer**,[1] and **Kenneth H. Downing**[1]* (khdowning@lbl.gov)

[1]Lawrence Berkeley National Laboratory, Berkeley, Calif. and [2]University of Missouri, Columbia

http://PCAP.LBL.GOV

Project Goals: The twin goals of this project are 1. to develop an integrated set of high throughput pipelines to identify and characterize multi-protein complexes in a microbe more swiftly and comprehensively than currently possible and 2. to use these pipelines to elucidate and model the protein interaction networks regulating stress responses in *Desulfovibrio vulgaris* with the aim of understanding how this and similar microbes can be used in bioremediation of metal and radionuclides found in U.S. Department of Energy (DOE) contaminated sites.

Bacteria display a spatially and temporally defined three-dimensional organization of their macromolecular complex inventory that allows them to grow, divide, and respond to

environmental cues. As part of the PCAP component of the LBNL ENIGMA Program, our overall goal is to develop tools that can be used to define the localization of such complexes within the cell.

Desulfovibrio vulgaris Hildenborough (DvH) is an anaerobic sulfate-reducing bacterium (SRB). It is used as a model organism for the study of environmental bioremediation of heavy metal and radionuclide contamination. Its ability to efficiently reduce toxic heavy metals such as uranium and chromium is of particular interest to the DOE for use in high-risk metal contaminated sites, and it can also provide novel insights on the range of metabolic pathways available in microbes. Protein expression levels and subcellular localization of proteins and macromolecular complexes may change in response to various environmental factors such as exposure to the target toxins, and may also differ among bacterial cells of identical genetic origin within a given culture. We aim to use our localization technology to understand how protein abundance and spatial distribution relate to the cell's normal metabolism and how changes in these factors are involved in the cell's response to changes in the environment.

For some of the very large complexes that display a distinctive size and shape, structural approaches for identification will allow complex localization, for example by identifying single-particle-EM derived 3D structures in whole mount cryo-tomograms. A high-resolution single particle analysis of the highest molecular weight protein structures purified from DvH cells has been conducted by the PCAP EM project, characterizing structures of 16 individual molecular machines at a resolution compatible with electron tomography (Han et al., PNAS 2009). It was hoped that these 'templates' could be used to localize and monitor the differing numbers of single particle structures within the tomographic datasets of intact DvH cells at distinctive stages of its growth and under different stress conditions.

While our cryo-electron tomography studies of DvH have been quite productive in characterizing aspects of overall morphology, they have so far yielded less information about cytoplasmic composition than we had initially hoped. The resolution of the tomograms within the DvH cytoplasm is noticeably lower than the resolution obtained in tomographic datasets from cells of other microbes of similar size. It was not clear if this lack of information is due to a lower number of internal molecular machines, i.e. ribosomes, GroEL, RNA polymerase etc, within the DvH cells, or if the cytoplasm of DvH is for some unknown reason idiosyncratically more electron dense than in other cells, which would decrease the contrast and thus limit our ability to identify cytoplasmic complexes.

The fledgling technique of vitreous cryo-sectioning has been adopted to elucidate this problem. A number of various cell types, including DvH, yeast, *E. coli* and *Caulobacter* have all been successfully high-pressure frozen in specialized copper tubes, sectioned and imaged in the frozen-hydrated state. Variations in the density of the extracellular medium, such as with dextran that is often used as a cryo preservative, provide a mechanism of contrast matching to complement other measures of cell mass density. A number of tentative tomographic datasets have been recorded from the sections, and data comparison is ongoing.

For smaller proteins, where shape or size is no longer a unique identifier, we can apply tag-based labeling approaches to identify the precise location of the tagged protein. To this end we have explored the SNAP-tag labeling system for DvH. This approach overcomes problems of using GFP derived-fusion proteins in an obligate anaerobe. Labeled cells can be examined at the light microscope level, and can also be photoconverted to provide contrast for EM visualization at much higher resolution. Over the last year we have overcome several obstacles that had plagued us in earlier work, and we recently were able to fluorescently label 13 out of 25 strains. Successful labeling was judged by fluorescence microscopy, SDS gel electrophoresis and plate reader analysis. Using photoconversion of fluorescent signals, followed by resin-embedding, we have begun to map out a variety of proteins at higher resolution. Several of the fluorescent cell lines have been successfully photoconverted and imaged at the TEM level to identify high-precision subcellular location of the tagged proteins. Interestingly we found cell-to-cell differences in labeling signal strength, which we attribute to real differences in protein expression. We also found cell-to-cell differences in extracellular metal reduction activity.

Energy Dispersive X-ray Analysis (XEDS) has also been used to track the evolution of internal and external electron dense material, which becomes visible through the life cycle of an anaerobic DvH culture. Large internal elemental sulfur balls as well as both internal and external Iron sulfide bodies have been identified and characterized.

Ultimately, in addition to describing the distribution of protein complexes within the cell, by comparing cell-to-cell differences in abundance and subcellular localization with uneven extracellular metal reduction activity we hope to be able to gain insight into the distribution and hence function of candidate proteins, assumed to play a role in extracellular metal reduction and other aspects of the cell's redox chemistry.

214

Environmental Microbiology Core Research on Stress Response Pathways in Metal-Reducers ENIGMA:VIMSS:ESPP

Terry C. Hazen[1,2*], Gary Anderson,[1,2] Sharon Borglin,[1,2] Eoin Brodie,[1,2] Steve van Dien,[8] Matthew Fields,[1,7] Julian Fortney,[1,2] Jil Geller,[1,2] E. Hendrickson,[5] Kristina L. Hillesland,[1,6] Hoi-Ying Holman,[1,2] J. Leigh,[1,6] T. Lie,[1,6] Dominique Joyner,[1,2] Romy Chakraborty,[1,2] Dwayne Elias,[1,3] Aindrila Mukhopadhyay,[1,2] Christopher Schadt,[1,3] David Stahl,[1,6] Sergey Stolyar,[1,6] Chris Walker,[1,6] Judy Wall,[1,5] Zamin Yang,[1,3] Huei-che Yen,[1,5] Grant Zane,[1,5] and Jizhong Zhou[1,9]

[1]Virtual Institute of Microbial Stress and Survival; [2]Lawrence Berkeley National Laboratory, Berkeley, Calif.; [3]Oak Ridge National Laboratory, Oak Ridge, Tenn.; [4]Diversa, Inc.; [5]University of Missouri, Columbia; [6]University of Washington, Seattle; [7]Montana State University, Bozeman; [8]Genomatica, San Diego, Calif.; and [9]University of Oklahoma, Norman

Project Goals: The environmental microbiology core of the ESPP project is the source of environmental data and samples that determine the stressors that will be studied, provides the environments for growing the organisms to be tested, simulates stressed environments, and verifies the conceptual models to determine how these stress regulatory pathways control the biogeochemistry of contaminated sites.

Field Studies

Previous research specifically points toward SRB as environmentally relevant experimental systems for the study of heavy metal and radionuclide reduction, and our recent data has detected *Desulfovibrio* sequences at the FRC and Hanford 100H. To effectively immobilize heavy metals and radionuclides, it is important to understand the cellular responses to adverse factors observed at contaminated subsurface environments, such as mixed contaminants and the changing ratios of electron donors and acceptors. In a recent study, we focused on responses to Cr (VI). At Hanford 100H as part of our ERSP project we injected 40 lbs of HRC (polylactate) as a slow release electron donor in August 2004. Until March 2008 reducing conditions were maintained, along with undetectable levels of Cr (VI) (Hubbard et al, 2008; Faybishenko et al., 2008). During this time the environment was dominated by sulfate reducers and we were able to detect *Desulfovibrio vulgaris*-like organisms with direct fluorescent antibody. Last year we injected 10 lbs of HRC at the same site to determine if there is a 'memory' response and observed H_2S production after only 23 h. Once the community stabilizes we will begin push-pull stress tests in the field with NO_3 and monitor functional gene, community structure, and stress responses as compared with previously published models by our group with pure cultures. We are also isolating consortia and determining the dominant community structure to compare with our lab studies.

In order to characterize microbial community dynamics associated with Cr (VI) biostimulation at the Hanford 100-H area, both groundwater and sediment communities are being tracked via SSU rRNA gene sequences. Stainless steel-mesh columns packed with Hanford sediments are placed down-well and microbial communities are allowed to establish. Four wells are being monitored, and baseline communities have been analyzed for the sediment-associated populations. Predominant populations include: *Acetivibrio; Geobacter; Ruminococcus, Alkaliflexus, Thiohalomonas, Acidovorax*, *Aquaspirillum,* and *Herbaspirillum*. Groundwater sample analysis is in progress, and biomass loads appear to be lower for groundwater compared to sediments. Rarfaction curves indicate that coverage has not been saturated; therefore, pyrosequencing efforts are also underway. Once baseline groundwater and sediment populations are determined during stimulation with HRC, community dynamics will be tracked during an in situ nitrate perturbation. Methanogen enrichments have yielded slow-growing cultures. A hydorgenotrophic enrichment is dominated by *Methanocella paludicola* (90% similar), and an acetoclastic enrichment is predominated by *Methanosarcina barker*i (99% similar).

We are also using the GeoChip 3.0, a comprehensive functional gene array contains ~25,000 probes covering ~47,000 sequences for 292 gene families involved in the geochemical cycling of C, N, and S, metal reduction and resistance, and organic contaminant degradation was used to examine microbial communities both at Hanford and at Oak Ridge within a pilot-scale test system established for the biostimulation of U (VI) reduction in the subsurface by injection of ethanol. Sediment from eleven different sampling wells, representing two different treatment zones within this system, was evaluated. The results showed that different microbial communities were established in different wells and high gene overlap was observed from wells within the same treatment zone. Higher microbial functional gene number, diversity and abundance were observed within the active bioremediation zone. The microbial community structure was highly correlated with the hydraulic flow rate and geochemical conditions of the treatment zone, especially pH, manganese concentration and electron donor level. In a different study of the same system, functional community dynamics were examined during a period of oxidation by nitrate. After exposure to nitrate the diversity and richness increased several fold but quickly returned to pre-nitrate levels. The relative abundance of several nitrogen cycling genes showed an increase immediately after nitrate exposure, including ammonification, denitrification, and nitrogen fixation genes indicating a stimulation of these communities after nitrate exposure.

Desulfovibrio spp. and consortia genomes. *Desulfovibrio* FW1012B was isolated from contaminated groundwater during biostimulation for U (VI) reduction in Oak Ridge, TN. Genome sequencing of isolate *Desulfovibrio* FW1012B, tentatively *Desulfovibrio oakridgensis*, was completed in late spring of 2009 at the DOE's Joint Genome Institute at Lawrence Berkeley National Lab in conjunction with Los Alamos National Lab, Oak Ridge National Lab, and Montana State University. The G+C content of the chro-

mosome is 66.5%. Based on the draft, incomplete sequence, automated gene scanning software predicted a total of 3,737 protein-encoding genes, 3,190 of which match best to KEGG-annotated genes in Proteobacteria genomes. The closest organisms with fully sequenced genomes are *Desulfovibrio vulgaris* Miyazaki (dvm) at 740 hits, *Desulfovibrio desulfuricans* G20 (dde) at 479 hits, *Desulfovibrio vulgaris* DP4 (dvl) at 375 hits and *Syntrophobacter fumaroxidans* (sfu) at 196 hits. A further 333 predicted protein-encoding genes had no match which could signify novel and/or unique genes. Analysis of the data raises some interesting questions regarding the metabolism of the organism. Unlike most *Desulfovibrio* species, *D. oakridgensis* appears to have a fully intact glycolysis pathway, a pyruvate dehydrogenase, and a nearly intact citric acid cycle. α-ketoglutarate is made by an NADP+ dependent isocitrate dehydrogenase rather than NAD+ dependent which creates and lacks a putative malate dehydrogenase. The presence of nitrogenase supports our observations that the organism can assimilate atmospheric nitrogen, and the lack of an aldehyde dehydrogenase coincides with the inability to use ethanol as electron donor and carbon source. Still, other questions are left unanswered. We are also sequencing a number of Hanford isolates that show association with *Desulfovibrio* at Hanford: *Geobacter metallireducens, Pseudomonas stutzeri*, and *Desulfovibrio vulgaris*.

This work was part of the ENIGMA SFA and the Virtual Institute for Microbial Stress and Survival (http://VIMSS.lbl.gov) supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomics:GTL program through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U.S. Department of Energy.

215

Laboratory Models for the Study of Community Interaction, Functional Stability, and Survival

A. Arkin[1]* (aparkin@lbl.gov), E. Baidoo,[1] P. Dehal,[1] D. Elias,[2] M. Fields,[3] J. Geller,[1] T. Hazen,[1] Z. He,[8] K. Hillesland,[4] J. Keasling,[1] C. Keller, M. Keller,[5] L. Krumholz, B. Meyer, L. Miller,[6] J. Mosher,[5] A. Mukhopadhyay,[1] A. Palumbo,[5] T. Phelps,[5] M. Podar,[5] L. Rajeev, A. Redding,[1] C. Schadt,[5] D. Stahl,[4] S. Stolyar,[7] A. Venkateswaren, C. Walker,[4] J. Wall,[2] Z. Yang,[5] G. Zane,[2] A. Zhou,[8] and J. Zhou[8]

[1]Lawrence Berkeley National Laboratory, Berkeley, Calif.; [2]University of Missouri, Columbia; [3]Montana State University, Bozeman; [4]University of Washington, Seattle; [5]Oak Ridge National Laboratory, Oak Ridge, Tenn.; [6]Brookhaven National Laboratory, Brookhaven, N.Y.; [7]Pacific Northwest National Laboratory, Richland, Wash.; and [8]University of Oklahoma, Norman

Project Goals: Model microbial communities can bridge the experimental gap between the simplicity of monoculture and the complexity of open environmental systems. We are developing laboratory communities of increasing phylogenetic, functional, and spatial complexity in order to: 1) identify extended genomic and regulatory networks and assess their importance to processes in the field, 2) determine how community organization influences activity and resilience, 3) determine how evolution affects community structure and activity over extended periods of adaptation, as is relevant to extended nutrient stimulation of field sites and the consequences of previous field site manipulation on achieving remediation goals.

Initial studies examined a relatively simple two-species community in which a sulfate reducer (*Desulfovibrio vulgaris* Hildenborough) was paired with a hydrogen consuming methanogen (*Methanococcus maripaludis*) in suspended steady-state cultures. This two-member community represents a critical step in microbial food webs that control organic matter mineralization. Only when the two species are growing together (syntrophic growth) can organic substrates be oxidized in the absence of electron acceptors such as sulfate. Significant change in community activity was associated with transition from suspended to attached (biofilm) growth states, forming different exopolymers and producing more methane. Methane production was also influenced by the spatial distribution of the two populations, pointing to the importance of initial conditions of biofilm formation for community structure and activity. Although flux balance models were in generally good agreement with experimental data, growth rate and yield of the two-member community increased significantly with long-term maintenance in the lab, indicating that adaptive evolution must be considered in extended manipulation of field sites. In addition, *D. vulgaris* from independently evolving communities had varying competitive abilities when paired with coevolved or foreign *M. maripaludis*, suggesting that there are different adaptive mechanisms and that species interactions affect the course of evolution. Complementary transcriptional and proteome analyses of the two-member community showed that *Desulfovibrio* uses largely independent energy generation pathways during syntrophic growth and sulfate-respiration, and further indicated the importance of using model communities to identify key genetic and physiological processes not expressed in monoculture.

In separate studies, we examined *D. vulgaris* adaptation to salt stress, a common environmental stressor. The most significant improvements in salt tolerance occurred between 500 and 1000 generations in a long-term evolution experiment. Transcriptional and ongoing genome sequencing analyses indicate that the contributing mutations were fixed by 1000 generations, with little subsequent increase in salt tolerance. Ongoing studies are now characterizing the contribution specific mutants identified in the salt-evolved *Desulfovibrio*, and *Desulfovibrio* recovered from evolved two-member communities, to altered phenotypes.

We are also developing higher communities and supporting technologies to understand interactions of increasing complexity, using the two-species community as the foundation for a tri-culture by adding the acetate oxidizing *Geobacter*

sulfurreducens as well as tri-cultures of *Clostridium cellulolyticum*, *D. vulgaris*, and *G. sulfurreducens* established in steady state chemostats using cellobiose and different electron acceptors. These higher-complexity communities more closely capture the multiple physiological states of metal-reducing bacteria in the environment, where they are dependent upon associated fermenters for nutrients and reducing equivalents. Parallel studies using steady-state chemostats established from Cr (VII) contaminated groundwater from Hanford, WA simulate the ongoing lactate injection experiments at Hanford, and allow us to better define key populations and interactions that lead to emergent system level properties of nutrient amended subsurface environments. Together these laboratory efforts are designed to improve process control and predictive understanding of environmental microbial activities relevant to DOE priorities.

216
VIMSS Systems Biology Knowledge Base

Paramvir S. Dehal[1,2]* (PSDehal@lbl.gov), Dylan Chivian,[1,2,3] Adam D. Deutschbauer,[1,2] Jason Baumohl,[1,2] Marcin P. Joachimiak,[1,2] Keith Keller,[1,2] Morgan N. Price,[1,2] and **Adam P. Arkin**[1,2,3,4]

[1]Virtual Institute for Microbial Stress and Survival, http://vimss.lbl.gov/; [2]Lawrence Berkeley National Laboratory, Berkeley, Calif.; [3]DOE Joint BioEnergy Institute, Emeryville, Calif.; and [4]Dept. of Bioengineering, University of California, Berkeley

Project Goals: The VIMSS Systems Biology Core group seeks to generate and integrate multiple functional genomic data sets in order to create a comprehensive framework for understanding the biology of the sulfate reducing environmental microbe *Desulfovibrio vulgaris* Hildenborough.

Background: The VIMSS Systems Biology Core group seeks to generate and integrate multiple functional genomic data sets in order to create a comprehensive framework for understanding the biology of the sulfate reducing environmental microbe *Desulfovibrio vulgaris* Hildenborough. The group is responsible for high throughput experiments, data management, data integration, data analysis, and comparative and evolutionary genomic analysis of the data for the VIMSS project. We have expanded and extended our existing tools sets for comparative and functional genomics to deal with new data produced by the VIMSS ESPP2 members. The Systems Biology Core is developing methods to store and analyze diverse data sets including: microarrays, ChIP-chip arrays, tiling arrays, proteomics, metabolomics, metabolic flux, phylochips, metagenomics sequencing, genome sequencing, growth curves, phenotype arrays, bar-coded knock out strain collections and links to existing literature and web based resources. Our analysis has been incorporated into our comparative and functional genomics website MicrobesOnline (http://www.microbesonline.org) and made available to the wider research community. By taking advantage of data integration across diverse functional and comparative datasets, we have been able to pursue large research projects in evolutionary and systems biology studies.

Systems Biology Experiments: The Systems Biology Core is currently prioritizing the functional annotation of SRB genomes and the creation of a complete systems-level investigation into the physiology of DvH. To accomplish these tasks in the most robust and reproducible manner possible, we have created a robotic setup to automate most of the sample prep, growth curve fitness and phenotype assays and data collection. To functionally annotate SRB genomes, we are systematically generating large sequence-defined transposon libraries in *D. desulfuricans* G20 and DvH. These libraries are "bar-coded" with unique DNA tags which will enable the parallel monitoring of strain fitness in thousands of SRB mutants. By screening across hundreds of growth conditions and monitoring per gene fitness effects, we will be able to assign phenotypic outcomes to each gene.

A complete systems-level investigation into DvH physiology requires a complete parts list of all transcribed elements in the genome. Towards this aim, we are using high-density tiling microarrays and next-generation sequencing technologies to precisely define the transcriptome of DvH. By combining high density tiling arrays with the sequencing data, we will be able to define transcription starts, operon structure, terminator sequences, improve promoter motif predictions, and identify potential antisense transcripts and small RNA genes.

The methods and techniques we have developed are applicable over a broad range of microbes and we will have to capacity to reproduce several of these experiments in related species, allowing the results to be analyzed in the broader context of *Desulfovibrio* evolution.

Data Integration: Data management, integration and distribution are critical functions for all large projects. A primary goal of the Systems Biology Core is to capture all experimental data from the ESPP2 investigators, including relevant metadata, raw data and processed data, and to make these data sets available through intuitive queries. Our group has developed Experimental Information and Data Repository (http://vimss.lbl.gov/EIDR/) and the MicrobesOnline database to provide this functionality. Researches have access to datasets from biomass production, growth curves, image data, mass spec data, phenotype microarray data and transcriptomic, proteomic and metabolomic data. New functionality has been added for storage of information relating to mutant strains, transposon knockout libraries and protein complex data, in addition to new visualization for assessing existing data sets such as the phenotype microarrays.

RegTransBase and RegPrecise: We have built tools and resources for studying regulation in bacteria and archaea using comparative genomics approach. In addition to working on a high quality semi-manual regulon inference in a wide range of species we are building several on-line resources covering different aspects of regulation. RegTransBase, a database of regulatory interactions from literature

collected by a group of experts, currently includes 5100 annotated articles describing 12 thousand experiments. RegPrecise describes manually curated computational predictions of regulons in bacterial genomes done by comparative genomics. RegPredict is a set of highly integrated web tools for fast and accurate inference of regulons. All regulation-related resources are based on the MicrobesOnline data.

The MicrobesOnline Database: The MicrobesOnline database (http://www.microbesonline.org) currently holds over 1000 microbial genomes and will be updated semi-annually, providing an important comparative and functional genomics resource to the community. New functionality added this year includes the addition of fungal genomes and the framework for adding additional eukaryotic genomes, an updated user interface for the phylogenetic tree based genome browser that allows users to view their genes and genomes of interest within an evolutionary framework, improved tools to compare multiple microarray expression data across genes and genomes, phylogenetic profile searches using our high quality species tree, and addition of external microarray data from the Many Microbial Microarrays Database for bacteria and Yeast. Additionally we have begun adding metagenomic data to MicrobesOnline.

MicrobesOnline continues to provide an interface for genome annotation, which like all the tools reported here, is freely available to the scientific community. To keep up with the rapidly expanding set of sequenced genomes, we have begun to investigate methods for accelerating our annotation pipeline. In particular we have completed work FastHMM and FastBLAST, methods to speed up the most time consuming process of our analysis pipeline, homology searching through HMM alignments and all against all BLAST. These methods now enable us to deal with the many millions of gene sequences generated from metagenomics. And our FastTree program allows us to create phylogenetic trees for all gene families, even those with over 100,000 members, so that all genes can be studied within an evolutionary framework.

217

Environmental Stress Pathway Project: Study of *Desulfovibrio vulgaris* Hildenborough

Adam P. Arkin[1]* (aparkin@lbl.gov), Edward E. Baidoo,[1] Kelly Bender,[5] Peter I. Benke,[1] Adam Deutschbauer,[1] Matthew Fields,[3] Terry C. Hazen,[1] Zhili He,[4] Dominique C. Joyner,[1] Jay D. Keasling,[1] Kimberly Keller,[2] Eric G. Luning,[1] Aindrila Mukhopadhyay,[1] Lara Rajeev, Jayashree Ray, Judy D. Wall,[2] Grant Zane,[2] Aifen Zhou,[4] and Jizhong Zhou[4]

[1]Lawrence Berkeley National Laboratory, Berkeley, Calif.; [2]University of Missouri, Columbia; [3]Montana State University, Bozeman; [4]University of Oklahoma, Norman; and [5]Southern Illinois University, Carbondale

Project Goals: The anaerobic, sulfate reducing bacterium *D. vulgaris* Hildenborough, provides an avenue to examine important microbial metal reducing functions in the environment. A systematic understanding of the physiology of such organisms provides invaluable insights into key metabolic and regulatory networks mechanisms and their evolution. The Environmental Stress Pathway Project (ESPP) builds upon the genetic tools, systems biology and analytical methodologies developed for *D. vulgaris* to obtain an in-depth knowledge of the metabolic capabilities, stress response, adaptation and regulatory networks of physiological states and factors that impact its physiology with respect to its environment. Projects outlined in this poster include cells wide studies, transposon library analyses, targeted studies of regulatory factors and signal transduction systems.

The anaerobic, sulfate reducing bacterium *D. vulgaris* Hildenborough, provides an avenue to examine important microbial metal reducing functions in the environment. A systematic understanding of the physiology of such organisms provides invaluable insights into key metabolic and regulatory networks mechanisms and their evolution. The Environmental Stress Pathway Project (ESPP) builds upon the genetic tools, systems biology and analytical methodologies developed for *D. vulgaris* to obtain an in-depth knowledge of the metabolic capabilities, stress response, adaptation and regulatory networks of physiological states and factors that impact its physiology with respect to its environment. Projects outlined in this poster include cells wide studies, transposon library analyses, targeted studies of regulatory factors and signal transduction systems.

Functional genomics continues to provide a valuable strategy to gain a broad cell wide understanding of *D. vulgaris* physiology. Having applied these measurements to specific stress response assessment, studies are now being focused on understanding *D. vulgaris* biofilms. Both transcript and protein expression profiles demonstrated that biofilm cells have an altered flux of carbon and energy compared to planktonic cells, which may influence metal-interacting capacity and survivability. In addition planktonic and biofilm cells were also exposed to different concentrations of

dissolved oxygen to establish if exposure had any effects on cells and/or biofilm formation. Results from these studies provide insight to better control the growth of sulfate-reducing biofilms.

To improve the functional annotation of *D. vulgaris* and related genomes, large sequence-defined transposon libraries have been generated in both *D. vulgaris* and *D. desulfuricans* G20. These libraries are "barcoded" with unique DNA tags to enable the parallel monitoring of strain fitness in thousands of mutants. Several broad and targeted studies have been conducted using these libraries. Furthermore, a complete systems-level investigation into *D. vulgaris* physiology requires a complete parts list of all transcribed elements in the genome. To address this a high-density tiling microarray and next-generation sequencing technologies are being used to precisely define the transcriptome of *D. vulgaris*.

A key focus of ESPP is to elucidate the structure and evolution of molecular networks in *D. vulgaris*. Towards this aim, a systematic examination of the two component signal transduction pathways is being undertaken. Two component systems trigger responses to a variety of stress and environmental signals. Approximately 70 such systems are annotated in *D. vulgaris*. Using a library of tagged two component system proteins (Histidine Kinases and Response regulators) high throughput protein-DNA interaction strategies are being adopted to map the two component regulatory network. These studies are being complimented using gene deletion mutants in Histidine kinases. The *D. vulgaris* genome also encodes Crp-Fnr like genes that are known to serve as positive transcription factors and play an important role in response to environmental stresses. To examine the function, regulation, and possible networks of these regulators, individual knockout mutants in all four Crp-Fnr genes are being studied. Findings from these studies will be summarized.

218

The Metalloproteomes of Microorganisms are Largely Uncharacterized: A Component of the MAGGIE Project

Michael Thorgersen[1]* (mthorger@uga.edu), Aleksandar Cvetkovic,[1]* Angeli Lal Menon,[1] Farris L. Poole II,[1] Joseph Scott,[1] W. Andrew Lancaster,[1] Jeremy Praissman,[1] Sarat Shanmukh,[1] Ewa Kalisiak,[2] Sunia Trauger,[2] Gary Siuzdak,[2] Steven M. Yannone,[3] John A. Tainer,[3] and **Michael W.W. Adams**[1]

[1]Dept. of Biochemistry and Molecular Biology, University of Georgia, Athens; [2]Center for Mass Spectrometry, Scripps Research Institute, La Jolla, Calif.; and [3]Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, Calif.

Project Goals: The overall goal of the MAGGIE project is to provide robust GTL technologies and comprehensive characterization to efficiently couple gene sequence and genomic analyses with protein interactions and thereby elucidate functional relationships and pathways. The operational principle guiding MAGGIE objectives is that protein functional relationships involve interaction mosaics that self-assemble from independent protein pieces that are tuned by modifications and metabolites, including metals. The objective is therefore to comprehensively characterize protein complexes on a genome-wide basis, including metal-protein complexes, underlying microbial cell biology.

The overall goal of the MAGGIE project is to provide robust GTL technologies to efficiently couple gene sequence and genomic analyses with protein interactions and thereby elucidate functional relationships and pathways. This involves a comprehensive characterization on a genome-wide scale of protein complexes and protein-cofactor complexes. Metals have been found to play essential roles as cofactors in many enzymes and proteins involved in nearly all cellular processes. However, the true extent of the metalloproteome of any organism remains unknown. It is not possible to predict the types of metal that an organism uses or the number and/or types of metalloprotein encoded in its genome sequence because metal coordination sites are diverse and not easily recognized. In this study, we used Pyrococcus furiosus, a hyperthermophilic archaeon that grows optimally at 100°C, as the model organism to gain insight into metalloprotein diversity. The key questions were, what elements does the organism assimilate from its normal laboratory growth medium, can techniques be devised to identify novel metalloproteins, and how specific is metal incorporation into proteins? Conventional non-denaturing liquid chromatography and high-throughput tandem mass spectrometry were used to separate and identify proteins, respectively, and metals were identified by inductively coupled plasma mass spectrometry (ICP-MS). Statistical and algorithmic methods were used to identify potentially novel metalloproteins. A total of 345 metal peaks were identified after fractionation of the cytroplasmic extract of P. furiosus through two levels of column chromatography, 160 of which could not be ascribed to any known or predicted metalloprotein. The peaks observed included metals known to be utilized by P. furiosus (Fe, Ni, Co, and Zn) and metals that P. furiosus was not previously known to take up (U, Pb, Ge, Mo, Mn, and V). Similar chromatographic and metal analyses were performed using the cytoplasmic extracts of two other microorganisms, *Sulfolobus solfataricus* and *Escherichia coli* (grown on their conventional laboratory media). These revealed peaks of yet other types of unanticipated metals. Several of the unassigned metal peaks in the *P. furiosus* analyses were purified to give a single protein using conventional liquid chromatography fractionation. This led to the identification of multiple new types of Ni- and Mo-containing proteins, the properties of which will be presented. The presence of novel V, Mn, Pb, U, and Ge proteins will also be discussed in terms of whether their incorporation into proteins by the organism appears to be intentional or unintentional. These results indicate that metalloproteomes are much more extensive than previously recognized, and likely involve both biologically conventional

and unanticipated metals with implications for a complete understanding of cell biology.

This research was funded by the Department of Energy (DE-FG02-07ER64326) as part of the MAGGIE project.

219

Access to Shape and Assembly of Macromolecular Complexes in Pathways Using Small Angle X-Ray Scattering

Greg L. Hura,[1]* Michal Hammel,[1] Angeli L. Menon,[2] Robert Rambo,[3] Michael W.W. Adams,[2] and **John A. Tainer**[3,4] (jat@scripps.edu)

[1]Physical Bioscience Division, Lawrence Berkeley National Laboratory, Berkeley, Calif.; [2]Dept. of Biochemistry and Molecular Biology, University of Georgia, Athens; [3]Life Science Division, Lawrence Berkeley National Laboratory, Berkeley, Calif.; and [4]The Scripps Research Institute, La Jolla, Calif.

Project Goals: The operational principle guiding MAGGIE (Molecular Assemblies, Genes, and Genomics Integrated Efficiently) objectives can be succinctly stated: protein functional relationships involve interaction mosaics that self-assemble from independent macromolecular pieces that are tuned by modifications and metabolites. Several metrics and tools have been developed in MAGGIE which identify, capture, characterize and predict the effects of macromolecular assemblies. One of our developed tools is small angle X-ray scattering (SAXS). MAGGIE was the first to recognize and apply SAXS as a proteomic scale tool by coupling robotic fluid handling with the extreme brightness of synchrotron light.

The operational principle guiding MAGGIE (Molecular Assemblies, Genes, and Genomics Integrated Efficiently) objectives can be succinctly stated: protein functional relationships involve interaction mosaics that self-assemble from independent macromolecular pieces that are tuned by modifications and metabolites. Several metrics and tools have been developed in MAGGIE which identify, capture, characterize and predict the effects of macromolecular assemblies. One of our developed tools is small angle X-ray scattering (SAXS). MAGGIE was the first to recognize and apply SAXS as a proteomic scale tool by coupling robotic fluid handling with the extreme brightness of synchrotron light.

SAXS characterizes the shape and assembly of macromolecules in solution. Information both on routine cellular processes and cell state are often carried in the shape and assembly of macromolecules on the length scales measured by SAXS. As we have enabled data collection at a rate of 96 samples in 4 hours on practical concentrations and volumes for most macromolecules a new scale of structural characterization has been enabled. Facile access to shape and conformation of macromolecules in a large number of contexts which are encountered during cellular processes has had significant impact on our understanding of pathways and mechanisms for information transduction. Often macromolecules interact with several metabolites and specific combinations cause large length scale re-arrangements.

Pathways which maintain genomic stability, protein re-naturation and most recently components of energy generation have been targets of our approach. Our purification and data collection success rate has been significantly enhanced using proteins from our 3 target extremophile organisms. In addition to providing higher stability proteins and complexes we hope to learn the mechanisms these organisms utilize to accomplish these ubiquitous tasks under the challenging conditions in which they thrive. These may provide design principles for engineering pathways of interest which are of ultimate interest to GTL.

In addition to utilizing SAXS as a tool to characterize shape and assembly of pathway components, members may be identified which are tractable for other biophysical techniques. Multimerization state influences concentrations required for many techniques. Large flexible regions may prohibit crystallization. Specific alterations to sequence or post translational modifications may stabilize a conformation which enhances the likelihood of success by other techniques.

The developed infrastructure for SAXS is a unique resource for GTL and has already been utilized by hundreds of research labs around the country. Future target pathways and relevant components will be directed by the bioinformatics core of MAGGIE. In addition, the extension of SAXS to membrane proteins is currently being explored.

Molecular Interactions, Protein Complexes, and Structural Biology

220

The ENIGMA Project: Mapping Protein Assemblies and Modifications by Cellular Deconstruction and Mass Spectrometry in the Hyperthermophiles *Sulfolobus solfataricus*, *Pyrococcus furiosus*, and *Halobacterium NRC-1*

Robert Rambo,[1] Adam Barnebey,[1] Michael W. Adams,[2] Gary Siuzdak,[3] Sunia Trauger,[3] Nitin S. Baliga,[4] Stephen R. Holbrook,[1] Trent Northen,[1] Ben Bowen,[1] Greg Hura,[1] John A. Tainer,[1,5] and **Steven M. Yannone**[1]* (SMYannone@lbl.gov)

[1]Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, Calif.; [2]Dept. of Biochemistry and Molecular Biology, University of Georgia, Athens; [3]Center for Mass Spectrometry, Scripps Research Institute, La

Jolla, Calif.; [4]Institute for Systems Biology, Seattle, Wash.; and [5]Dept. of Biochemistry and Molecular Biology, Scripps Research Institute, La Jolla, Calif.

Project Goals: As part of the integrated ENIGMA project our goals are to develop generalized methodologies for identifying and isolating molecular machines and modified proteins from native biomass. We aim to identify and characterize both stable and transient macromolecular assemblies within the cell, the cell membrane, and outside of the cell using a rational cellular deconstruction approach. Proteins are identified by MS/MS from our cellular deconstruction fractions have inherent cellular locality, assembly mass, protein modification, and potential partner protein information associated with their identification. Our proteome-wide approach to identify macromolecular assemblies and protein modifications will ultimately lead to identification of metabolic modules suitable for transfer between microbes. We are developing generally applicable molecular and biophysical technologies for GTL to ultimately apply functional modules to confer specific metabolic capabilities to microbes and address DOE mission goals.

Dynamic protein-protein interactions are fundamental to most biological processes and essential for maintaining homeostasis within all living organisms. These interactions create dynamic and diverse functional networks essential to biological processes. Thus, a thorough understanding of these networks will be critical to engineering biological processes for DOE missions. The project was conceived, in part, as a response to the DOE GTL initiative to develop technologies to map the proteomes of model organisms. In this project we are exploiting unique characteristics of members of extremophilic Archaea to identify, isolate, and characterize multi-protein molecular machines. We have teamed expertise in mass spectrometry, systems biology, structural biology, biochemistry, and molecular biology to approach the challenges of mapping relatively simple proteomes.

As part of the ENIGMA project, we have developed methods for whole cell deconstruction of microbes that separates intact protein complexes under native conditions. The cellular deconstruction primarily fractionates the organism into four major classes; 1) membrane, 2) large mass (>800 kDa), 3) small mass (<800 kDa), and 4) extracellular. The final cellular partitions were resolved on SDS-PAGE and excised for high throughput MS/MS protein identification at the Scripps Center for Mass Spectrometry. We have achieved over 50 percent coverage of the predicted proteome and all identifications have inherent subcellular locality and co-fractionation information associated with them: 1) proteins identified in sedimenting fractions are by definition members of large mass complexes, 2) isolated membrane vesicles allow identification of membrane associated proteins, and 3) extracellular proteins contain extracellular membrane-bound and secreted proteins groups. We are refining analysis tools to identify co-variations of proteins across the fractionation scheme to build a system-wide protein interaction network. Ultimately, we aim to identify metabolic modules suitable to transfer specific metabolic processes between microbes to address specific DOE missions while developing generally applicable molecular and biophysical technologies for GTL.

221

High-Pressure Cryocooling of Protein Crystals: Applications to Understanding Pressure Effects on Proteins

Sol M. Gruner*[1,2] (smg26@cornell.edu), Chae Un Kim,[3] Nozomi Ando,[1] Buz Barstow,[1] Mark Tate,[1] and Yi-Fan Chen[3]

[1]Physics Dept., [2]Cornell High Energy Synchrotron Source (CHESS), and [3]MacCHESS (Macromolecular facility at the Cornell High Energy Synchrotron Source), Cornell University, Ithaca, New York

http://bigbro.biophys.cornell.edu

Project Goals: Understand effects of pressure on proteins.

A novel high-pressure cryocooling technique for preparation biological samples for x-ray analysis has been developed [1-3]. The method, high-pressure cryocooling, involves cooling samples to cryogenic temperatures (~ 100 K) in high-pressure Helium gas (up to 200 MPa). It bears both similarities and differences to high-pressure cooling methods that have been used to prepare samples for electron microscopy, and has been especially useful for cryocooling of macromolecular crystals for x-ray diffraction. Many different kinds of macromolecular crystals have been successfully high-pressure cryocooled and excellent crystal diffraction has been obtained with little or no penetrating cryoprotectants.

This new method has great potential for understanding pressure effects on proteins. As an example, high pressure cryocooling has been used to understand the structural basis for why the emission spectrum of the protein, Citrine, is pressure dependent [4,5]. The deformation of the Citrine chromophore is actuated by the differential motion of two clusters of atoms that compose the β-barrel scaffold of the molecule, resulting in a slight bending of the β-barrel. The high-pressure structures also reveal a perturbation of the hydrogen bonding network that stabilizes the excited state of the Citrine chromophore. The perturbation of this network is implicated in the reduction of fluorescence intensity of Citrine. The blue-shift of the Citrine fluorescence spectrum resulting from the bending of the β-barrel provides structural insight into the transient blue-shifting of isolated yellow fluorescent protein molecules under ambient conditions and suggests mechanisms to alter the time-dependent behavior of Citrine under ambient conditions. Finally, the Citrine example serves as a model for the way in which global pressure-induced structural perturbations affect the activity of proteins,a s well as a model for how these perturbations may be studied.

Supported by the Dept. of Energy (DE-FG02-97ER62443), the National Institutes of Health (RR-001646) and CHESS

(National Science Foundation and NIH-NIGMS through NSF DMR-0225180.)

References

1. Kim et al., Proc. Natl. Acad. Sci., USA 106 (2009) 4596.
2. Kim et al., J. Appl. Cryst., 41 (2008) 1.
3. Kim at al., Acta Cryst. D63 (2007) 653.
4. Barstow et al., Biophys. J., 97 (2009) 1719.
5. Barstow et al., Proc. Natl. Acad. Sci. USA, 105 (2008) 13362.

222

Opportunities for Structural Biology and Imaging at NSLS-II

Lisa M. Miller[1,2*] (lmiller@bnl.gov) and Wayne A. Hendrickson[1,3]

[1]NSLS-II Project and [2]National Synchrotron Light Source, Brookhaven National Laboratory, Upton, N.Y.; and [3]Dept. of Biochemistry and Molecular Biophysics, Columbia University, New York, N.Y.

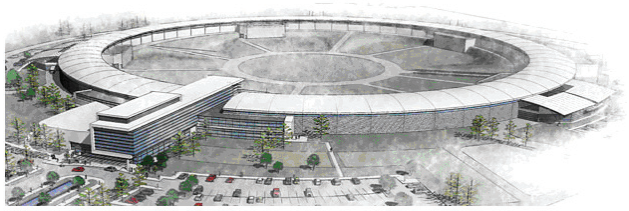
http://www.bnl.gov/nsls2/

Project Goals: In this poster, we will describe how the unique characteristics of NSLS-II can address scientific problems relevant to BER scientists by presenting a series of applications to bioenergy, carbon cycling and sequestration, and contaminant transport and cleanup in the environment. We will emphasize the wide range of techniques that will permit multiscale exploration: at the molecular level, to understand how genes determine biological structure and function; at the cellular level, to understand how molecular processes are coordinated to execute cell function; and at the level of microbial communities and higher organisms to understand how cells interact and respond to their environment. Finally, we are eager to solicit new applications of synchrotron science to research problems of interest to BER scientists.

The National Synchrotron Light Source-II (NSLS-II), which is now under construction at Brookhaven National Laboratory, will provide a broadband source of synchrotron photons from infrared light to x-rays with a brightness unsurpassed by any synchrotron facility worldwide. This new facility is scheduled to be operational in 2015 and will replace the existing NSLS.

The extreme brightness and coherence of NSLS-II will enable characterization techniques, such as nanoscale imaging, that are currently in their infancy or do not exist today. But importantly, NSLS-II will also take widely utilized methods, such as macromolecular crystallography (MX), small-angle x-ray scattering (SAXS), and x-ray absorption spectroscopy (XAS), and extend them to new regimes in time- and spatial-resolution that cannot be achieved today.

Recent workshops by BER have identified both biological challenges and technological needs that are important to the BER research community. For example, in May 2009, BER held the New Frontiers in Characterizing Biological Systems workshop to address the next generation challenges in genomics science and its connection to functional systems. The panel identified numerous knowledge gaps that inhibit the understanding of biological systems. These knowledge gaps are relevant to understanding research areas paramount to BER interest, including the generation and processing of biomass into chemical energy, climate change and the cycling of carbon and nutrients, and the transformation of natural and man-made contaminants in the environment.



The NSLS-II will be operational in 2015 and will provide a broadband source of photons with a brightness and coherence unsurpassed by any synchrotron worldwide. For BER science, it will enhance time-resolved structural studies of macromolecules and complexes, especially in more natural environmental settings. High throughput structure/function determination will be optimized to link genomic information to molecular events. And it will provide a wide range of nanoscale imaging capabilities with the possibility of multi-modality characterization of identical samples.

Synchrotron-based characterization tools are well-suited to fill the identified gaps. Synchrotron studies will generate basic understanding of biological processes, and not just for particular phenomena at a certain physical or temporal scale, but as linked pan-genomically across scales of investigation. With the high brightness and coherence of NSLS-II, structural studies of macromolecules and complexes will be possible in a time-resolved manner, especially in more natural environmental settings. Moreover, high throughput structure/function determination will be able to link genomic information to molecular events. NSLS-II will provide a wide range of nanoscale imaging capabilities, permitting multi-modality characterization of identical samples.

At the NSLS today, biological and environmental sciences users represent approximately 60% of the user community and more than 650 of the facility's annual publications. NSLS-II plans to follow in the footsteps of the current NSLS by providing a wide range of characterization techniques to the biological sciences community. In January 2008, a series of Scientific Strategic Planning workshops were held at the NSLS to identify a pathway forward to NSLS-II. An overarching conclusion from the Life and Environmental Sciences workshops was the desire within these communities to see increased interaction, collaboration, multi-technique integration, and cross-disciplinary approaches to doing science in the future. It was suggested that this mode of research can be achieved through a "Biol-

ogy Village" environment, which would include strategically locating beamlines for scientific interaction, having programmatic overlap through shared equipment, technology, and human resources, and playing an active role in the Joint Photon Sciences Institute (JPSI), an interdisciplinary facility at BNL that will facilitate R&D efforts.

In this poster, we will describe how the unique characteristics of NSLS-II can address scientific problems relevant to BER scientists by presenting a series of applications to bioenergy, carbon cycling and sequestration, and contaminant transport and cleanup in the environment. We will emphasize the wide range of techniques that will permit multiscale exploration: at the molecular level, to understand how genes determine biological structure and function; at the cellular level, to understand how molecular processes are coordinated to execute cell function; and at the level of microbial communities and higher organisms to understand how cells interact and respond to their environment. Finally, we are eager to solicit new applications of synchrotron science to research problems of interest to BER scientists.

Validation of Genome Sequence Annotation

223

Robotic Chemical Protein Synthesis for the Experimental Validation of the Functional Annotation of Microbial Genomes

Stephen Kent and Kalyaneswar Mandal* (kmandal@uchicag.edu)

Institute for Biophysical Dynamics, University of Chicago, Ill.

Project Goals: Robotic total chemical synthesis to make proteins and protein domains, for the validation of functional anotation of predicted open reading frames.

Modern total protein synthesis has evolved from the 'chemical ligation' methods introduced by the Kent laboratory [Kent SBH. Total chemical synthesis of proteins. *Chemical Society Reviews* 2009; 38: 338-51.]. Unprotected synthetic peptide segments, spanning the amino acid sequence of the mature polypeptide chain derived from a predicted open reading frame, are covalently joined to one another by chemo-selective reaction. Native chemical ligation, the thioester-mediated covalent bond-forming chemoselective reaction of unprotected peptides at a Cys residue, is the most robust and useful ligation chemistry developed to date. The synthetic protein is then used to experimentally validate the predicted biochemical function, and in selected cases to determine the Xray structure of the protein molecule (**Figure**).

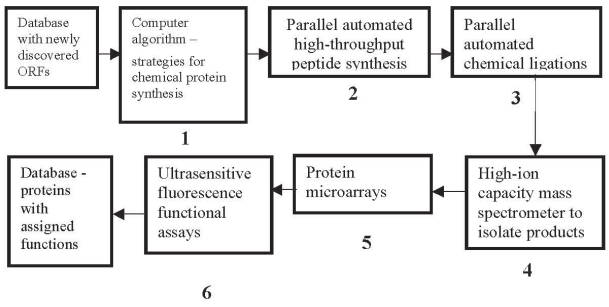


Figure 1. Modular high-throughput platform for fast and parallel total chemical synthesis, mass-spectrometric purification and single-molecule spectroscopic assay to annotate function for newly predicted proteins.

We have previosuly described the use of x,y,z robotics and laboratory automation and efficient Fmoc chemistry SPPS protocols for the simultaneous parallel synthesis of the key peptide-thioester building blocks needed for chemical protein synthesis. This made use of a recently reported novel resin linker[Blanco-Canosa JB, Dawson PE: An efficient Fmoc-SPPS approach for the generation of thioester peptide precursors for use in native chemical ligation. *Angew Chem Int Ed Engl.* 2008, 47:6851-5]. Typical data are shown in **Figure 2 (Top)**.

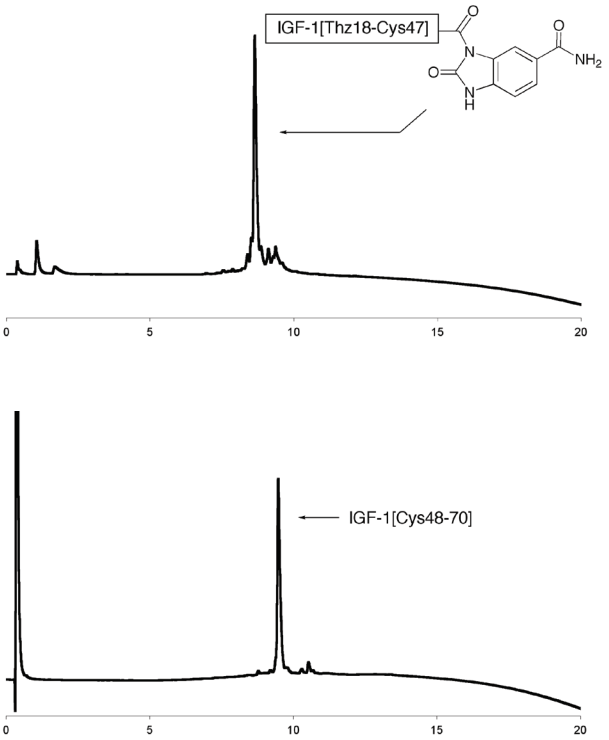


Figure 2. Automated robotic Fmoc SPPS preparation of peptide-thioesters. HPLC-electrospray MS of crude products are shown.

Ready preparation of peptide-thioesters enables the straightforward total chemical synthesis of proteins by native chemical ligation. Proof-of-concept total chemical

syntheses of predicted proteins from microbial and plant genomes will be presented.

224

Using Deep RNA Sequencing for the Structural Annotation of the *Laccaria bicolor* Mycorrhizal Transcriptome

Peter E. Larsen,[1]* Geetika Trivedi,[2] Avinash Sreedasyam,[2] Vincent Lu,[1] Gopi K. Podila,[2] and **Frank R. Collart**[1] (fcollart@anl.gov)

[1]Biosciences Division, Argonne National Laboratory, Lemont, Ill. and [2]Dept. of Biological Sciences, University of Alabama, Huntsville

http://www.bio.anl.gov/molecular_and_systems_biology/proteins.html

Project Goals: To facilitate the process for acquisition of function from complex environmental sequence data sets, we developed methods to utilize RNA-seq data to validate current gene model intron-exon boundary, correct errors in the structural annotation and extend the boundaries of the current gene models using assembly approaches.

Advances in sequences technology have enabled deep sequence interrogation of individual organisms as well as complex systems. This capability has led to an improved appreciation of the biological diversity associated with specific ecosystems and the complexity of the molecular systems involved in perception and response to external stimuli. Mapping these signaling pathways is challenging however in sequence data sets from environmental and/or metagenomic projects where uncharacterized organisms often represent a high proportion of the sequence data. To facilitate the process for acquisition of function from complex environmental sequence data sets, we are developing methods to utilize RNA-seq data to correct errors in the structural annotation and extend the boundaries of current gene models using assembly approaches. To validate the methods, we used a transcriptomic data set derived from the fungus *Laccaria bicolor* which develops a mycorrhizal symbiotic association with the roots of many tree species, in which the fungus provides nutrients to the tree in exchange for photosynthetically-derived sugars. This fungal-plant symbiosis is a widespread process of major ecological importance and knowledge of the molecular events and expressed protein sequences associated with the development of the mycorrhizal system is essential for our understanding of natural biological processes related to carbon sequestration, carbon management, sustainability and bioenergy.

We generated >30 million RNA-seq reads from Laccaria grown in culture. Our study used the 20614 gene "best model" set and 65-megabase Laccaria genomic DNA sequence from the publically available FTP site at the Joint Genome Institute. Our analysis focused on the subset of 1501 gene models that are differentially expressed in the mycorrhizal transcriptome and are expected to be important elements related to carbon metabolism, membrane permeability and transport, and intracellular signaling.

Our analysis of the intron-exon boundaries in current JGI best gene model set indicates the quality of *L. bicolor* structural annotation is enabling for homology-based comparison applications, but has severe limitations for experimental studies. For every intron-exon boundary in JGI Best Model set, we generated an 18-mer 'probe' sequence consisting of 9 bp up and down-stream of the intron-exon boundary. This intron-spanning sequence was used to search the set of RNA-seq reads. At least one read-containing probe was considered validation of gene model intron-exon boundary. Using these criteria, we were able to validate ~80% of the intron-exon boundaries within the gene model boundaries. This level of validation is notable in view of the complexity of the fungal genome (*L. bicolor* genes contain an average of 5.4 introns) and the annotation limitations arising from the relatively small number of sequenced fungal genomes. However, the combination of the error rate and intron density means that 42% of the current gene models contain intron/exon boundaries that do not map to the mRNA sequence data. Also, 58% of gene model 5' and/or 3' boundaries did not agree with the collected transcriptomic data. Inaccurate representations of the protein coding sequence are a consequence of these inconsistencies. Accurate coding sequences are essential for experimental approaches to characterize protein function and also to enhance the utility of tools that enable identification cellular localization signals and functional domains. Substantial changes to predicted UTRs also affect the ability to predict the regulatory mechanisms of mycorrhizae-specific genes. To improve the experimental utility of the gene model set, we developed algorithms that use the RNA-seq data to extend the boundaries of the current gene model set where appropriate, identify those intron-exon boundaries that can be validated by the transcriptomic data, and to generate novel intron-exon boundaries to bridge those regions of the gene models that are not supported by RNA-seq data. This extended and bridged contiguous expressed sequence was then aligned to the genome using a modified Smith-Waterman algorithm to recover gene model's structural annotation. Of the set of 1501 gene models, 1439 (96%) successfully generated modified gene models in which all error flags were successfully resolved and sequences aligned to genomic sequence. The remaining 4% (62 gene models) either had deviations from transcriptome data that could not be spanned or generated sequence that did not align to genomic sequence. We considered a gene model significantly changed if at least one of three criteria were meet: 1) an inconsistency in the original gene model was successfully bridged and aligned to scaffold, 2) the revised gene model contained a change in the total number of exons, and/or 3) we observed an absolute change in expressed gene size of more than 10%. Based on application of these criteria to the set of 1439 revised models, 974 (69%) of gene models required changes to match the transcriptomic consensus sequence. Additionally, for 465 (31%) of the models in the original best gene model set, we did not detect any inconsistencies and therefore have independently confirmed the previously published 'BestModel' annotation. Of those 62 gene models that could not be adequately

validated by the method proposed here, a number appear to have multiple isoforms in the expressed transcriptome data identifying them as genes of potential biological interest.

The outcome of this process is a set of high confidence gene models that can be reliably used for experimental characterization of protein function. This improved annotation process can be extended to other important gene families and will facilitate the process to identify the molecular mechanisms leading to the development of the mycorrhizal symbiosis and its implications in improving carbon sequestration by poplar.

The submitted manuscript has been created by UChicago Argonne, LLC, Operator of Argonne National Laboratory ("Argonne"). Argonne, a U.S. Department of Energy Office of Science laboratory, is operated under Contract No. DE-AC02-06CH11357.

225

Molecular Approaches for Elucidation of Sensory and Response Pathways in Cells

Sarah Zerbs,[1]* Sarah E. Giuliani,[1] Elizabeth Landorf,[1] Maureen McNulty,[1] William Studier,[2] and **Frank R. Collart**[1] (fcollart@anl.gov)

[1]Biosciences Division, Argonne National Laboratory, Lemont, Ill. and [2]Biology Division, Brookhaven National Laboratory, Upton, N.Y.

http://www.bio.anl.gov/molecular_and_systems_biology/proteins.html

Project Goals: This program addresses the hypothesis that cellular behavior can be modeled through an understanding of the biological interface with the environment and the cellular responses that originate from the cell/environment interaction. The long term objective of the program is to define cellular sensory and regulatory pathways that respond to environmental nutrients thereby facilitating a system-level model that predicts the cellular response to environmental conditions or changes.

Increased knowledge of protein function enhances our understanding of cellular functions and is ultimately required to model biological activities and systems. This program addresses the hypothesis that cellular behavior can be modeled through an understanding of the biological interface with the environment and the cellular responses that originate from the cell/environment interaction. The long term objective of the program is to define cellular sensory and regulatory pathways that respond to environmental nutrients thereby facilitating a system-level model that predicts the cellular response to environmental conditions or changes. The program uses a parallel strategy of technology development to improve capabilities for extraction of relevant biological information from the sequence data coupled to genome scale approaches for elucidation of protein function and cellular regulatory networks.

One aspect of this program will develop tools to bridge the gap between genomes and systems biology. Progress in sequencing technology has provided molecular validation of the diversity and complexity of environmental systems. However, sequencing capacity has far outpaced computational and experimental methods to fully utilize the genomic data. We are addressing this gap between DNA sequence and the ability to extract relevant biological information from the sequence data by the development of genome scale approaches for elucidation of protein function and cellular regulatory networks. These approaches utilize next generation sequencing technology and high throughput approaches to enable economical and efficient protein production and characterization.

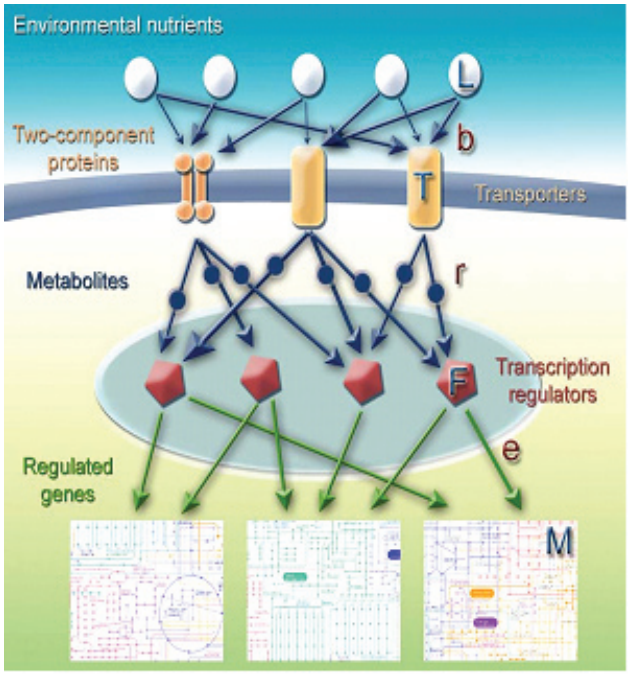


Fig 1. Illustration of experimental approach and application to systems modeling.

The capabilities to improve functional interrogation of sequences are coupled to *in vitro* methods for functional characterization of proteins involved in cellular sensory and response pathways. The functional screens will focus on key proteins that mediate communication between the cell and the environment such as transporters, two-component sensory systems, and membrane receptors (Fig. 1). This functional characterization will be linked to the cellular regulatory network by identification of the transcription factors whose activity is mediated by the environmental ligands or their metabolic derivatives. A coupling of the regulatory ligands with the DNA-binding regions of the transcription factors allows the association of metabolic pathways with the regulatory network. This genome scale process will determine the functional properties and potential of microbes and plants that are central to DOE missions. The functional assignments and ability to define specific sensory and regulatory pathways will increase the predictive capability of current models and support the development of

predictive systems-level models. This increased knowledge of the molecular components and control features of cellular sensory and response pathways is essential for our understanding of natural biological processes related to carbon management, sustainability and bioenergy.

The submitted manuscript has been created by UChicago Argonne, LLC, Operator of Argonne National Laboratory ("Argonne"). Argonne, a U.S. Department of Energy Office of Science laboratory, is operated under Contract No. DE-AC02-06CH11357. Brookhaven National Laboratory is operated under Contract No. DE-AC02-98CH10886.

226

Functional Linkage of ABC Transporter Profile with Metabolic Capability in *Rhodopseudomonas palustris*

Sarah E. Giuliani,[1]* Ashley M. Frank,[1] Catherine Seifert,[1] Lisa M. Miller,[2] Loren Hauser,[3] and **Frank R. Collart**[1] (fcollart@anl.gov)

[1]Biosciences Division, Argonne National Laboratory, Lemont, Ill.; [2]National Synchrotron Light Source, Brookhaven National Laboratory, Upton, N.Y.; and [3]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tenn.

http://www.bio.anl.gov/molecular_and_systems_biology/proteins.html

Project Goals: We suggest that the functional profile of the genome set of transporter proteins is predictive of metabolic capabilities and ecological niche of organisms. To test this hypothesis, we profiled the genome set of ABC transporters for *Rhodopseudomonas palustris* CGA009 to define the relationship between the transporter profile and metabolic capability for *R. palustris* CGA009.

Transporter proteins are an organism's primary interface with the environment. The expressed set of transporters mediates cellular metabolic capabilities and influences signal transduction pathways and regulatory networks. The role and impact of different transporters families differ in eukaryotic and prokaryotic organisms and the absolute number of transporters is dependent on the characteristics of the ecological niche. We suggest that the functional profile of the genome set of transporter proteins is predictive of metabolic capabilities and ecological niche of organisms. To test this hypothesis, we profiled the genome set of ABC transporters for *Rhodopseudomonas palustris* CGA009. In the *R. palustris*, ABC-type transporters represent approximately 45% of all transporters encoded in the genome. The ABC transporters family is widely distributed in soil organisms and can transport a variety of substrates such as metals, small ions, mono- and oligosaccharides, peptides, amino acids, iron-siderophores, polyamines, and vitamins.

An ABC transporter complex consists of a permease, ATPase, and a solute binding protein. The ligand specificity is determined by the solute binding protein which in some cases can utilize multiple membrane permeases. The genome of *R. palustris* CGA009 encodes approximately 117 ABC type transporters as determined by the number of encoded solute binding proteins. The functional properties of these transport proteins are largely unknown and less that 10% have specific functional assignments. The largest group of binding proteins is annotated as "branched-chain amino acid" binding protein. To improve the utility of the function annotation, we expressed and purified the set of binding proteins from *R. palustris* and are characterizing ligand-binding specificity using ligand libraries consisting of environmental and cellular metabolic compounds and high throughput binding screens, including fluorescence thermal shift, small angle x-ray scattering, x-ray absorption spectroscopy, circular dichroism spectroscopy, and infrared spectroscopy. To date, this process resulted in the assignment of specific binding ligands for approximately 60% of the purified and screened proteins. In most cases, the binding was observed for specific compound classes and was observed for only 1-3 compounds from the entire ligand library. For approximately 20% of the screened proteins, a specific binding ligand was not observed, which we attribute to the limited scope of the screening library relative to the complexity of compounds in the natural environment.

The impact of these studies is two-fold. First, our screening method generated specific functional annotations for an important group of uncharacterized or incorrectly annotated transporter proteins. For example, six proteins encoded by genes annotated as branched chain amino acid binding proteins were demonstrated to bind various aromatic compounds derived from lignin degradation. Analysis of the flanking genomic regions reveals the co-localization of these transporter genes with metabolic genes associated with utilization of the transported compounds. Similar functional insight was obtained for previously uncharacterized proteins associated with the transport of fatty acids, dicarboxylic acids, oligopeptides, metals, and additional small molecule compounds. This functional insight can be used to improve the annotation of related organisms and provides a route to evaluate the evolution of the important and diverse group of transporter proteins.

Second, the results of this study also provide important biological insight for the metabolic capabilities and environment fitness of this organism. The profile and number of transport proteins specific for aromatic compounds is consistent with ecological and laboratory studies which demonstrate the capabilities of this organism for the utilization of plant degradation products such as lignin-derived aromatic compounds.

One of these binding proteins, RPA1385, showed high affinity and selectivity for vanadate, which is a catalytic component of a nitrogenase protein complex. *R. palustris* is a nitrogen fixing bacteria and has been shown to utilize a vanadium nitrogenase (V-nitrogenase) as a metabolic alternative when molybdenum is limited in the environment. Prior to this research, the cyanobacterium *Anabaena variabilis* (which also contains a V-nitrogenase) was the only organism know to contain a defined high-affinity vanadate

transport system. In *R. palustris*, genes RPA1381-1386 are annotated as components of a vanadate nitrogen fixation system based on homology to other similar proteins. However, in *R. palustris*, homology search approaches failed to identify the high-affinity vanadate transport system. Our ligand mapping approach identifies the RPA1385 protein as the vanadate SBP gene for this ABC transport system. This finding not only identifies a key component of the vanadate nitrogenase fixation pathway for this organism, but may also confirm a proposed hypothesis that the presence of this system in *R. palustris* suggests vanadate transport systems have evolved at least twice from dissimilar ancestral genes.

The functional assignments in conjunction with gene expression profiles and transcription factor DNA binding sites enable the identification of the cellular regulatory and metabolic components that enable the use of lignin degradation products for cell growth. This approach is being applied to other sequenced strains of *R. palustris* to provide evolutionary insight for the number and substrate specificity of this family of ABC type transporters. These capabilities will enable the identification and characterization of metabolic and regulatory pathways that are associated with a specific environmental niche.

The submitted manuscript has been created by UChicago Argonne, LLC, Operator of Argonne National Laboratory ("Argonne"). Argonne, a U.S. Department of Energy Office of Science laboratory, is operated under Contract No. DE-AC02-06CH11357. Brookhaven National Laboratory is operated under Contract No. DE-AC02-98CH10886.

227

Phylogenomics-Guided Validation of Function for Conserved Unknown Genes

Valérie de Crécy-Lagard[1]* (vcrecy@ufl.edu), Basma El Yacoubi,[1] Marc Bailly,[1] Ian K. Blaby,[1] Anne Pribat,[2] Aurora Lara-Núñez,[2] and Andrew D. Hanson[2]

[1]Dept. of Microbiology and Cell Science and [2]Dept. of Horticultural Sciences, University of Florida, Gainesville

Project Goals: Our overall goal is to establish an innovative integrative approach to predict and experimentally verify the in-vivo function of genes that lack homologs of known function ('unknown' gene families) and that are highly conserved among prokaryotes and plants. By implementing this approach we will predict, and experimentally validate for a chosen subset, the function of ~1500 unknown genes.

Identifying the function of every gene in all sequenced organisms is a central challenge of the post-genomic era. We are submerged in genomic, transcriptomic, and proteomic data but the functions of about half (range 20 to 60%) of the genes in any given organism are still unknown. Our goal is to predict and experimentally verify the *in-vivo* function of proteins that lack homologs of known function ('unknown' protein families) and that are highly conserved between prokaryotes and plants. Our approach combines the extensive post-genomic resources of the plant field with the use of comparative genomic tools made possible by the availability of thousands of sequenced microbial genomes. This is an integrative approach to predict gene function whose early phase is computer-assisted, and whose later phases incorporate intellectual input from expert plant and microbial biochemists. It allows bridging of the gap between automated homology-driven annotations and the classical gene discovery efforts driven mainly by experimentalists. Our goal is to predict and experimentally validated the function of 15 "unknown" protein families". We have already validated predictions for seven families (in orange in Table1 and we present the other eight most mature predictions that currently being tested (in yellow in Table 1). Two examples of this second list will be presented in more detail to emphasize the synergistic aspects of plant-microbe comparative genomics.

COG0799 proteins occur in plants, in nearly all bacteria, and in animals and fungi. Plants have two isoforms, one apparently chloroplastic, the other mitochondrial. The archetypal member of the family is the plant Iojap protein; *iojap* mutants of maize lack functional chloroplast ribosomes. In bacteria, COG0799 genes cluster strongly with the NAD synthesis gene *nadD* (nicotinate mononucleotide adenylyltransferase) and sometimes the two genes are fused. COG0799 genes also cluster with genes encoding the ribosomal biogenesis protein ObgE and ribosomal proteins L21 and L27, making a connection with the ribosome lesion in the maize *iojap* mutant. Furthermore, transcriptomic data from *Arabidopsis* show co-expression of *iojap* with various chloroplast ribosome protein genes. NadD mediates a reaction in the *de novo* synthesis of NAD and potentially in salvage of nicotinamide mononucleotide (NMN). We therefore predict that Iojap catalyzes a process in ribosome biogenesis that releases NMN from NAD, and that the NMN is recycled by NadD. Possible Iojap reactions include a NAD-dependent DNA ligase-like reaction or an ADP-ribosyltransferase.

Case	Hypothesis	TAIR ID	COG, gene name	Subsystem in SEED	Experimental verification status	PubMed ID
1	Pterin carbinolamine dehydratase with role in Moco metabolism	At1g29810 At5g51110	COG2154, phhB	Pterin_carbinolamine_ dehydratase	Validated in in 7 eukaryotes and 8 prokaryotes	18245455
2	t6A biosynthesis	At5g60590	COG0009, YrdC	YrdC-YciO	Validated in Yeast, Archaea and two bacteria;	19287007
3	PTPS family protein replacing the FolB step in folate synthesis	-	COG0720	Experimental-PTPS	Validated in 1 eukaryote and 8 prokaryotes	19395485, 18805734
4	Metal chaperone-Zinc homeostasis	At1g15730, At1g26520, At1g80480	COG0523	COG0523	Validated in several bacteria	19822009
5	Folate-dependent Fe/S cluster synthesis or repair protein	At4g12130 At1g60990	COG0354, ygfZ	YgfZ-Fe-S	Validated in *E. coli*, *Haloferax volcanii*, *Arabidopsis*, *Leishmania*, yeast, mouse	Submitted
6	Alternative route for 5-formyltetrahydrofolate disposal	At2g20830	COG3643	Experimental_Histidine_ Degradation	Verified in 5 prokaryotes	Manuscript in prep
7	t6A biosynthesis	At2g45270, At4g22720	COG0533, YgjD	YrdC-YciO	Validated in yeast	Manuscript in prep
8	NAD-dependent nucleic acid AMP ligase	At3g12930, At1g67620	COG0799, alr4169	Iojap	In progress *E.coli*	
9	5-Formyltetrahydrofolate cycloligase paralog	At1g76730	COG0212	5-FCL-like_protein	Predicted role in thiamine recycling	
10	Hydroxyproline-galacatosyl hydrolase	At5g12950, At5g12960	COG3533, SAV1144	COG3533	In progress in *X. campestris*	
11	m6A in small rRNA	At4g28830	COG2263	rRNA_modification_Archaea	Mutant analysis in *H. volcanii* in progress	
12	Choline transporter	NiaP homolog At1g13050	MFS superfamily	Choline transport and metabolism	In progress *R. solanacearum* and *B. xenovorans*	
13	Ribosome assembly/translation termination	At1g09150	COG2016	rRNA_modification_Archaea	In progress in yeast and *H. volcanii*	
14	Phytol phosphate kinase	At1g78620	COG1836, alr1612	COG1836	In progress in Synechocystis	
15	Pyridoxal phosphate enzyme in amino acid metabolism, most likely in the Glu-Pro area	At4g26860, At1g11930	COG0325, yggS	PROSC	In progress in *E. coli*	

Table 1. Status of most advanced fifteen families

COG3533 genes are found in all plants and occur sporadically in plant pathogens (bacteria and fungi) and in human pathogens. The corresponding proteins are similar to glycosyl hydrolase but the specific substrates are not known. The *Arabidopsis* COG3533 genes (At5g12950 and At5g12960) are expressed highly in pollen. Bacterial COG3533 genes are physically clustered with genes for hydroxyproline degradation, arabinose catabolism, and peptidases. We therefore propose that COG3533 proteins are glycosylhydrolases that cleave the hydroxyproline-linked galactosides found in plant cell wall proteins or in collagen. Such a hydrolase has been predicted to have a role in pollen growth and would allow plant pathogens to utilize plant cell wall components as carbon sources.

228

Functional Annotation of Putative Enzymes in *Methanosarcina acetivorans*

Ethel Apolinario,[1] Libuse Brachova,[3] Yihong Chen,[2] Zvi Kelman,[2] Zhuo Li,[2] Basil J. Nikolau,[3] Lucas Showman,[3] Kevin Sowers,[1] and **John Orban**[2]* (orban@umbi.umd.edu)

[1]Center of Marine Biotechnology, University of Maryland Biotechnology Institute, Baltimore; [2]Center for Advanced Research in Biotechnology, University of Maryland Biotechnology Institute, Rockville; [3]W. M. Keck Metabolomics Research Laboratory, Iowa State University, Ames

Project Goals: The goal of the project is to develop rapid experimental approaches for accurate annotation of putative enzymatic functions. Targets of interest range from those with tentatively assigned function to hypotheticals.

Methane-producing organisms provide an efficient and cost-effective biofuel which is self-harvesting and can be distributed readily using infrastructure that is already in place. As with other genomes, however, accurate functional

annotation in methanogens lags significantly behind the large body of sequence data, representing a sizable gap in our understanding of biology in these organisms. We are using the methanogenic archaeon, *Methanosarcina acetivorans* (MA), as a model system for developing experimental tools for rapid and reliable annotation and validation of function. The target genes are putative enzymes in MA with detectable *in vivo* expression.

Our experimental approach utilizes a combination of methods for rapid function assignment. NMR spectroscopy is used to screen for putative substrates, products, or their structural analogs. Where possible, we have followed up on function assignments by checking to see if the MA gene can complement the corresponding *E. coli* knockout. We have used this approach to both validate and correct functional assignments in MA target genes, as will be illustrated with examples. Further, insights into the functional annotation of "hypotheticals" are being obtained by integrating mass spectrometry based metabolite profiles of gene knockouts with NMR-based approaches and these will also be discussed.

229

Robust Prediction of Protein Localization Via Integration of Multiple Data Types

Margaret Romine[1]* (Margie.romine@pnl.gov), Lee Ann McCue,[1] Gretta Serres,[2] Tatiana Karpinets,[3] Mustafa Syed,[3] Sam Purvine,[1] Michael Lueze,[3] Guruprasad Kora,[3] Denise Schmoyer,[3] Ed Uberbacher,[3] Jim Fredrickson,[1] and Mary Lipton[1]

[1]Pacific Northwest National Laboratory, Richland, Wash.; [2]Marine Biological Laboratory, Woods Hole, Mass.; and [3]Oak Ridge National Laboratory, Oak Ridge, Tenn.

Project Goals: The primary research emphasis will be on associations between autotrophic and heterotrophic microorganisms with the additional objective of obtaining a predictive understanding of how interactions impart stability and resistance to stress, environmental fitness, and functional efficiency.

Genome annotations play a central role in omics-based characterization of cellular behavior and consequently it is important that they are as accurate and functionally descriptive as possible. Currently, domain content is the primary type of functional evidence used for automated functional annotation of protein-coding genes (CDSs) deduced from genome sequences. While domain content can sometimes suggest a precise function or at least provide a general functional categorization (e.g. TonB-dependent receptor), they are more often only useful for establishing that proteins having the same domain (s) are somehow functionally related. Protein localization prediction is a form of evidence that is generally under-utilized in automated annotation pipelines but has the potential to provide very useful clues regarding CDS function. As part of our efforts to manually improve the annotation of the currently available *Shewanella* genome sequences, we developed a strategy for more accurately predicting subcellular protein localization through integration of proteome data, the output of several different localization prediction tools, ortholog analysis, and domain analyses.

At the outset of this exercise we recognized that one of the major limitations of tools that computationally predict protein localization is that an accurate gene model is needed. Since many of the commonly used tools search for the occurrence of characteristic N-terminal targeting peptides, they will be unable to detect secretion substrates that are encoded by genes with start codon inaccuracies, gene sequencing mistakes, or genomic mutations that result in displacement or loss of sequences that encode the N-terminal targeting peptide. In order to address issues with the accuracy of the gene models we first mined available MS-MS proteome data from 12 sequenced shewanellae genomes for partial tryptic peptides that could be mapped to the mature termini of proteins deduced from the original or subsequently adjusted gene models. These analyses included searches for peptides that map to N-termini produced by cellular proteolytic processing by signal peptidase I, methionine aminopeptidase, or proline aminopeptidase. We identified such peptides for 1290 proteins (~30% of the total predicted) in the extensively studied model organism *S. oneidensis* MR-1 and between 299 and 661 proteins (~10% of the total predicted) in 11 other shewanellae for which proteome data was available but was derived from only a single sample. The positions of mobile elements (insertion elements, MITES, phage, and other integrative elements) were mapped to facilitate detection of gene fragments encoding targeting peptides that were displaced by gene interruption. This analysis resulted in an increase in pseudogene count from 735 to 1499. Ortholog tables comprised of proteins from all 19 strains were constructed so that we could compare, within each ortholog group, the output of several localization and domain predictors with the expectation that inconsistencies in predictions would most often arise due to errors in either the gene model or predicted ortholog grouping. Ortholog groups with inconsistencies in predicted domain content, function, or location prediction or for which members were missing in a genome were then manually evaluated for inaccuracies in gene models or ortholog grouping. These analyses lead to the addition of 769 new genes and removal of 1554 genes from the gene models of these 19 shewanellae. Taking into account only changes made to the gene models of intact genes, we adjusted the start position positions in 2466 genes thereby achieving a greater consistency in predictions of localization or domain content within each ortholog group.

Since Gram negative bacteria like *Shewanella* sp. have a complex cell envelope consisting of inner and outer membranes that are separated by a periplasmic space, they employ specialized systems to mediate translocation of proteins across one or both membranes, insertion of proteins into one membrane or the other, or to tether them to one side of a membrane. The sorting signals recognized by these systems differ from one another and thus no single algorithm is optimal for predicting the subcellular locations of all proteins. This need to apply more complex

logic for predicting protein location became evident when we discovered that representatives of all six specialized protein translocation systems (T1SS-T6SS) known to occur in gram negative bacteria were present in at least one sequenced *Shewanella*. We developed a series of rules to identify substrates of specialized secretion systems as either bioinformatics tools were not available to identify their substrates or their predictions were not particularly robust. For example, combining domain information and proteomics data for the NiFe hydrogenase orthologs allowed us to identify these proteins as substrates of the twin arginine translocation (TAT) system. Substrates of the TAT secretion system are expected to include proteins that possess metallic redox active centers and therefore all proteins having such domains, including the NiFe hydrogenases, were carefully evaluated for the presence of N-terminal targeting peptide recognized by this secretion pathway. In *Shewanella*, the NiFe hydrogenases have an unusually long targeting peptide that was validated by proteome analysis (68 amino acids) but routinely missed by both TatP and Tatfind algorithms. The identification of outer membrane proteins was also not very accurate using a single computational tool. The Bomp beta barrel prediction tool, for example, inconsistently detected outer membrane proteins within ortholog groups even after gene model adjustment. Therefore, we supplemented these analyses by searching for a C-terminal outer membrane targeting consensus motif. Since it is known that some outer membrane proteins do not encode this domain at the C-terminus (e.g. OmpA family proteins, secretins) we also used location-informative domains to assist in identification of outer membrane proteins. Other systems, such as the type II secretion system (T2SS) that translocate periplasmic proteins across the outer membrane have no universally recognized targeting motif, but are instead believed to be recognize targeting signals that are species-specific. In *Shewanella* it is known that at least three lipoproteins are substrates of this system. A comparative analysis of these lipoproteins with other proteins deduced from the genome sequence revealed a putative targeting motif similar to those described for extracellular proteins in other bacteria, providing us a means to expand the number of predicted T2SS substrates in this Genus.

We estimate that approximately 40% of the predicted proteome for each strain of *Shewanella* is translocated out of the cytoplasm. These extracytoplasmic proteins play a central role in modulating the interactions of members of this genus with their external environments and in generating the energy and accessing the nutrient necessary to support growth and metabolism. As part of PNNL's new Foundational Science Focus area on Biological Systems Interactions we intend to employ this general strategy to identify secreted proteins in new model organisms and microbial communities to facilitate future studies directed at developing a broader understanding of microbial interactions.

submitted post-press

Genome-Scale Phylogenetic Function Annotation of Large and Diverse Protein Families

Barbara E. Engelhardt,[1,4] Michael I. Jordan,[2] Susanna Repo,[3] and Steven E. Brenner[3] (brenner@compbio.berkeley.edu)

[1]EECS Dept., [2]Dept. of Statistics, and [3]Plant and Microbial Biology Dept., University of California, Berkeley; and [4]Computer Science Dept., University of Chicago, Ill.

Project Goals: The goal of the project is to enhance the algorithms and statistical models of SIFTER, our protein function prediction method. We will also extend SIFTER's applicability by including additional sources of function evidence. With these improvements, SIFTER will become applicable to a broader range of protein families, including large, and functionally diverse families, and to work on genome-wide scale. In addition, we will adapt SIFTER on metagenomic data.

It is now easier to discover thousands of protein sequences in a new microbial genome than it is to biochemically characterize the specific activity of a single protein of unknown function. Through metagenomic analysis, next-generation sequencing heralds unprecedented opportunities for understanding the environmental microbiota. A single experiment alone, the Global Ocean Sampling study, more than doubled the number of known protein sequence entries. However, despite this large body of new sequence information, functional annotation remains a major challenge. Molecular functions of proteins in the novel genomes continue to be discovered, in large part by homology to those experimentally characterized in model organisms.

Typically, protein function annotation involves finding homologs of a protein sequence, followed by database queries and computational techniques to predict function from the annotated homologs. These methods rely on the principle that proteins from a common ancestor may share a similar function. However, most protein families have sets of proteins with different functions and therefore traditional bioinformatics approaches are unable to reliably assign the appropriate function to unannotated proteins. Currently, protein function databases have a large proportion of erroneously annotated proteins, where the incorrect annotations were either derived using an imprecise computational technique or inferred using another incorrect annotation[1-4].

We have proposed integrating available functional data using the evolutionary relationships of a protein family, and we implemented this method in the program SIFTER (Statistical Inference of Function Through Evolutionary Relationships). The SIFTER methodology uses a statistical graphical model to compute the probabilities of molecular functions for unannotated proteins. Currently, SIFTER takes as input a reconciled phylogeny and a set of annota-

tions for some of the proteins in the protein family. We incorporate known information about function by computing the probability of each of the candidate functions for the proteins in the tree with available functional evidence from the GOA database. The candidate molecular functions are represented as a boolean vector, where initially the probability associated with each candidate function is a function of the set of annotations for that protein and their corresponding evidence types (e.g., experimental, electronic). From this reconciled phylogeny with sparse observations, SIFTER computes the posterior probability of each molecular function for all proteins in the family using a simple statistical model of protein function evolution.

We tested the performance of SIFTER on three different protein families: AMP/adenosine deaminases, sulfotransferases and Nudix hydrolases with cross-validation experiments. SIFTER's performance was compared with three other function prediction algorithms: BLAST, GOtcha and Orthostrapper, and SIFTER was shown to outperform the other methods. In addition, on a genome-wide scale we used SIFTER to annotate the experimentally characterized proteins from *Schizosaccharomyces pombe*, based on the annotations from 26 other fungal genomes. The newest version of SIFTER implements a faster method for calculating the posterior probabilities, and this improvement, together with a more general evolutionary model make SIFTER applicable on large and functionally diverse protein families and on genome-scale function annotation.

The development of SIFTER is an ongoing project and a new version of the program is now available (manuscript under review). We are currently testing SIFTER for metagenomic sequences with the acid mine drainage datasets from Jill Banfield. In the near future, we are planning to expand our analysis to other metagenomic datasets, such as the termite gut datasets from the JGI. We also use SIFTER to annotate enzymes from chlorite dismutase and perchlorate reductase families, in order to identify species that are capable of perchlorate reduction. Furthermore, we are validating SIFTER predictions experimentally using the large and extremely diverse Nudix family of hydrolases as a test bed.

This project has been funded with DOE grant number BER KP 110201.

References

1. Brenner SE 1999 *Trends Genet*. **15** 132-3
2. Galperin MY and Koonin EV 1998 *In Silico Biol*. **1** 55-67
3. Jones CE, Brown AL and Baumann U 2007 *BMC Bioinformatics*. **8** 170
4. Schnoes AM, Brown SD, Dodevski I and Babbit PC 2009 *PLoS Comput Biol*. **5** e1000605

Computing for Systems Biology

230

Standards in Genomic Sciences: Launch of a Standards Compliant Open-Access Journal for the 'Omics Community

G. M. Garrity[1,5]* (garrity@msu.edu), N. Kyrpides,[2,5] D. Field,[3,5] P. Sterk,[3,5] H.-P. Klenk,[4,5] and the Editorial and Advisory Boards of Standards in Genomic Sciences

[1]Michigan State University, East Lansing; [2]DOE Joint Genome Institute, Walnut Creek, Calif.; [3]NERC Center for Ecology and Hydrology, Oxford, United Kingdom; [4]DSMZ – German Collection of Microorganisms and Cell Cultures GmbH, Braunschweig, Germany; [5]Genomic Standards Consortium, Seattle, Wash.

Project Goals: The goal of DOE funding was to underwrite a pre-launch meeting of the Editorial and Advisory Boards of Standards in Genomic Sciences. The meeting was held at Michigan State University on March 12–14, 2009.

Standards in Genomic Sciences (SIGS) is an open-access e-journal that was created to promote the data standardization efforts of the Genomic Standards Consortium (GSC). The GSC was founded in 2005 by an international community of like-minded scientists to work toward improving the descriptions of our rapidly growing collection of genomes and metagenomes [1,2]. In the absence of metadata standards, the difficulty of exchanging and integrating genomic data into analytical models and public knowledgebases increases while the overall value of each subsequent sequence diminishes. This is problematic because the ease and cost of producing sequence data have dropped sharply while the cost of annotation and documentation has increased.

Membership in the GSC consists of biologists, bioinformaticians, and computer scientists, with representatives from the National Center for Biological Information (NCBI), European Molecular Biology Laboratory (EMBL), National Institute of Genetics Japan (NIG), J. Craig Venter Institute (JCVI), DOE Joint Genome Institute (JGI), European Bioinformatics Institute (EBI), Sanger Institute, and a number of other international research organizations involved in cross-cutting research. As a first step toward accomplishing organizational goals, the GSC published the "Minimum Information about a Genome Sequence" (MIGS) specification, which describes the core information that should be reported with each new genome or metagenome [3]. The GSC has led the development of a richer set of descriptors within GCDML (Genomic Contextual Markup Language), an XML variant for mark-up and transport of genomic and metagenomic data and the Genomic Rosetta Stone, a proposed resolver for mapping genome identifiers across databases [4,5]. The GSC also participates in initiatives on data

standardization, including Biosharing and the sequencing finishing standards recently described by Chain et al. [6,7]

The rational for SIGS is to provide a venue for publication of highly structured, standardized publications of genome and metagenome sequences in accordance with MIGS and to report on other efforts that promote data standardization and data sharing [8]. Whereas peer-reviewed publications of genomes were once commonplace in a number of journals, the current trend is for many general and discipline-specific publications to eschew such papers, leading to a loss of contextual information that is critical for analyzing and interpreting genome sequence data [9]. SIGS aims to counter this trend and to provide concise, standardized descriptions of the sequencing and annotation methods along with biological information about the source organism, which we refer to as short genome reports. To that end, and with the generous support from the Michigan State University Foundation to fund the editorial office and the U.S. Department of Energy Office of Science, Biological and Environmental Research Program to convene the first meeting of the editorial and advisory boards, an open-access publication was launched to help meet those needs.

Publication of SIGS began in July 2009. At the end of October, the journal had published 28 short genome reports on bacterial and archaeal species, many of which were derived from the ***G**enomic **E**ncyclopedia of **B**acteria and Archaea* collaboration between the DOE Joint Genome Institute (JGI) and the German Collection of Microorganisms and Cell Cultures (DSMZ). We anticipate publishing at least another 14 short genome reports before the close of 2009, along with approximately five to six additional papers, bringing the total to approximately 60 published articles in the first volume.

Growth of the journal, to date, has been largely organic, through the journal website, search engines and forward linking of SIGS DOIs on the websites of other journals that have been cited in SIGS articles. Published articles have been downloaded > 4,750 times. Readership of SIGS is worldwide, with visitors to the site coming from over 60 countries during November, with the preponderance of visitors coming from North America and western Europe. Approximately half the daily visitors are new to the site, and the bulk of that traffic appears to be directed to the site either by search engines or by direct linking from other sites. We anticipate that traffic will continue to grow as additional content is published and SIGS becomes accepted in the major literature indices (PubMed Central, PubMed, ISI Web of Science). Our goal is to engage with the GTL community to solicit feedback and discuss additional unmet publishing needs.

References

1. Field D., Hughes J. Cataloguing our Current Genome Collection. Microbiology 2005; 151: 1016-1019. PubMed doi:10.1099/mic.0.27914-0.
2. Field D., Garrity G., Morrison N., Sterk P., Selengut J., Thomson N., Tatusova T. Meeting Report: eGenomics: Cataloguing Our Complete Genome Collection I. Comp Funct Genomics 2006; 6: 357-362. doi:10.1002/cfg.493.
3. Field D., Garrity G., Gray T., Morrison N., Selengut J., Sterk P., Tatusova T., Thomson N., Allen M.J., Angiuoli S.V., et al. The Minimum Information about a Genome Sequence (MIGS) Specification. Nat Biotechnol 2008; 26: 541-547. PubMed doi:10.1038/nbt1360.
4. Kottmann R., Gray T., Murphy S., Kagan L., Kravitz S., Lombardot T., Field D., Glockner F.O. A Standard MIGS/MIMS Compliant XML Schema: Toward the Development of the Genomic Contextual Data Markup Language (GCDML). OMICS 2008; 12: 115-121. PubMed doi:10.1089/omi.2008.0A10.
5. Van Brabant B., Gray T., Verslyppe B., Kyrpides N., Dietrich K., Glockner F. O., Cole J., Farris R., Schriml L. M., De Vos P., et al. Laying the Foundation for a Genomic Rosetta Stone: Creating Information Hubs Through the Use of Consensus Identifiers. OMICS 2008; 12: 123-127. PubMed doi:10.1089/omi.2008.0020.
6. Chain P. S., Grafham D.V., Fulton R. S., Fitzgerald M. .G, Hostetler J., Muzny D., Ali J., Birren B., Bruce D. C., Buhay C., et al. Genomics. Genome Project Standards in a New Era of Sequencing. Science 2009; 326: 236-237. PubMed doi:10.1126/science.1180614.
7. Field D., Sansone S. A., Collis A., Booth T., Dukes P., Gregurick S. K., Kennedy K., Kolar P., Kolker E., Maxon M., et al. Megascience. 'Omics Data Sharing. Science 2009; 326: 234-236. PubMed doi:10.1126/science.1180598.
8. Garrity, G. M., Field D., Kyrpides N., Hirschman L., Sansone S. A., Angiuoli S., Cole J. R., Glockner F.O., Kolker E., Kowalchuk G., et al. Toward a Standards-Compliant Genomic and Metagenomic Publication Record. OMICS 2008; 12: 157-160. PubMed doi:10.1089/omi.2008.A2B2.
9. Liolios K., Mavormatis K., Tavernarakis N., Kyrpides N. The Genomes On Line Database (GOLD) in 2007: Status of Genomic and Metagenomic Projects and Their Associated Metadata. Nucleic Acids Res 2008; 36: D475-D479. PubMed doi:10.1093/nar/gkm884.

231

NamesforLife Semantic Resolution Services for the Life Sciences

Charles T. Parker,[1] Dorothea Rohlfs,[1] Sarah Wigley,[1] Nicole Osier,[1] Catherine Lyons,[1] and George M. Garrity[1,2,*] (garrity@namesforlife.com)

[1]NamesforLife, LLC, East Lansing, Michigan and [2]Michigan State University, East Lansing

Project Goals: The overall objective of the Phase II study is to develop and deploy a set of convenient, easy to use semantic services that provide end-users with on-demand access to key information. This ensures that they can accurately interpret the meaning of any bacterial or archeal name when encountered in digital content.

Within the Genomes-to-Life Roadmap, the DOE recognizes that a significant barrier to effective communication in the life sciences is a lack of standardized semantics that accurately describe data objects and persistently express knowledge change over time. As research methods and biological concepts evolve, certainty about correct interpre-

tation of older data and published results decreases because both become overloaded with synonymous (multiple terms for a single concept) and polysemous terms (single terms with multiple meanings). Ambiguity arising from rapidly evolving terminology is a common and chronic problem in science and technology. N4L services are being developed to address this problem. The core of N4L consists of a data model, an XML schema, and an expertly managed vocabulary that is interlinked with Digital Object Identifiers (DOIs) to form a transparent semantic resolution service that disambiguates terminologies, makes them actionable, and provides direct links back to key literature and data resources.

Objectives - The overall objective of the current Phase II study is to deploy a set of convenient, easy to use semantic services that allow end-users to accurately interpret the meaning of a biological name or other dynamic term encountered in digital content, on demand and without having to query external resources or to leave the material they are reading or searching. The service can be used by database owners, publishers, or other information providers, to semantically enable their offerings; making them readily discoverable by their clients, even when the definition of a name or term has changed.

Curatorial Efforts - We significantly extended the scope of our data curation and built a framework for distributing and enhancing N4L information services to different categories of users. The target vocabulary consists of the validly published names of Bacteria and Archaea, which provides a rich and complex set of interrelated terms and interlinked resources that have a high value to the GTL community. At the end of 4Q 2009, there were 13043 validly published names (of which 3022 are synonyms of varying complexity) corresponding to 12630 taxonomic concepts and 8976 biological entities. Currently, new validly published names appear in the literature at a rate of 3.9 names/day. This number is however significantly lower than the number of names that have no standing in the literature (14.9 names/day) that appear in INSDC records and the GenBank taxonomy. Trivial names appearing on INSDC records add further confusion to the process and occur at a rate more than five-fold higher than validly published names.

The NamesforLife data have undergone further refinement to improve their accuracy. All names, taxon and exemplar records have been asserted by literature, corresponding to 9474 references, including 277 involving judicial opinions that affect the legitimacy and valid publication of 433 names. This has significance to the GTL program as some genomes are currently posted under rejected names (e.g. Sinorhizobium medicae). We have also verified all of the strains, culture collection deposits, and 16S rRNA sequences used in taxonomic assertions based on a review of the published record. This addresses a growing problem that has arisen from more than a decade of automated data harvesting, coupled with transitive data closure, leading to numerous systematic errors that are being routinely re-propagated.

N4L BrowserTool – The N4L BrowserTool provides a means of wide-spread distribution of our semantic resolution services to end-users of scientific and technical literature, published in digital form and distributed via the web. The tool is currently distributed as a Firefox extension and provides on-the-fly client-side mark-up of bacterial and archaeal names with links to NamesforLife information objects that can be actuated on demand. Alpha testing of the BrowserTool ran from May – December 2009. Large-scale beta-testing is scheduled for January-February 2010 with a product release in March 2010.

N4L Autotagger – The N4L Autotagger provides publishers and other content providers with a method for enabling and enhancing content during composition. This results in articles that contain persistent links to N4L information objects and allow readers to view such content in any browser. Collaborative work is underway with the Society for General Microbiology to enhance and enable content appearing in the International Journal of Systematic and Evolutionary Microbiology.

N4L Contextual Index – The BrowserTool and AutoTagger are designed to recognize bona-fide nomenclatural events in pre-composition XML and HTML, thus allowing for high-fidelity harvesting of new/modified names and associated references from the taxonomic literature automatically. These tools can also capture the metadata for each source, thus allowing us to track all such events. This information is being used to create a contextual index that enhances the value of N4L tools as each successive use can be placed into a variety of larger contexts and used for a variety of purposes, ranging from resource planning to plotting research trends at both a fine-grained (taxon specific) and global level. In addition to the scientific literature, we are building the necessary infrastructure to permit the use of our tools to uncover prior art in areas of interest to the DOE (e.g. bioenergy/biobased feedstocks/genomics) in the U.S. and EPO patent literature.

This research is supported by the Office of Biological and Environmental Research of the U.S. Department of Energy under Phase II STTR Award DE-FG02-07ER86321 A001

232

Numerical Optimization Algorithms and Software for Systems Biology: An Integrated Model of Macromolecular Synthesis and Metabolism of *Escherichia coli*

Ines Thiele[1]* (ithiele@hi.is), R.M.T. Fleming,[1] A. Bordbar,[2] R. Que,[2] and B.O. Palsson[2]

[1]Center for Systems Biology, University of Iceland, Reykjavik, Iceland and [2]Bioengineering Dept., University of California, San Diego, La Jolla

Project Goals: This project aims to reconstruct a genome-scale model of metabolism and macromolecular synthesis and to develop algorithms capable of solving the resulting large, stiff and ill-scaled matrices. This project combines

state of the art reconstruction and constraint-based modeling analysis tools with high-end linear optimization solvers and convex flux balance analysis. The incorporation of thermodynamic information in addition to environmental constraints will allow an accurate assessment of feasible steady states. While we will prototype the reconstruction and algorithm developments with *Escherichia coli*, we will employ the resulting networks to determine thermodynamically favorable pathways for hydrogen production by *Thermotoga maritima*.

Systems biology is a rapidly growing discipline. It is widely believed to have a broad transformative potential on both basic and applied studies in the life sciences. In particular, biochemical network reconstructions are playing a key role as they provide a framework for investigation of the mechanisms underlying the genotype-phenotype relationship. The constraint-based reconstruction and analysis approach was applied to reconstruct the transcriptional and translational (tr/tr) machinery of *Escherichia coli*. This reconstruction, denoted 'Expression-matrix' (E-matrix), represents stoichiometrically all known proteins and RNA species involved in the macromolecular synthesis machinery. It accounts for all biochemical transformations to produce active, functional proteins, tRNAs, and rRNAs known to be involved in *E. coli*'s tr/tr machinery. An initial study investigated basic properties of the E-matrix, including its capability to produce ribosomes, which was found to be in good agreement with experimental data from literature. Furthermore, quantitative gene expression data could be integrated with, and analyzed in the context of, the resulting constraint-based model. Adding mathematically derived constraints to couple certain reactions in the model allowed the quantitative representation of the size of steady state protein and RNA pools.

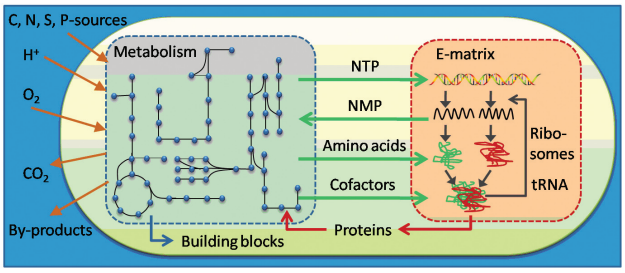


Figure 1: Functional synergy between the metabolic network and the macro-molecular synthesis network in *E. coli*.

The E-matrix was integrated with the genome-scale *E. coli* metabolic model and extended the transcriptional and translational reactions to encompass genes encoding all the respective metabolic enzymes. The resulting Metabolite-Expression-matrix (ME-matrix), exceeds the predictive capacity of the metabolic model and it can, for example, be used to predict the biomass yield since it represents the production of almost 2,000 proteins. *E. coli*'s ME-matrix is the first of its kind and represents a milestone in systems biology as it demonstrates how to quantitatively integrate 'omics'-datasets into a network context, and thus, to study the mechanistic principles underlying the genotype-phenotype relationship. We will show some possible applications which include protein engineering, interpretation of adaptive evolution, and minimal genome design. An integration of the ME-matrix with remaining cellular processes, such as regulation, signaling, and replication, will be a next step to complete the first whole-cell model.

Building on this reconstruction effort we now started to construct the ME-matrix for *Thermotoga maritima* based on published data. Furthermore, significant advances have been made in incorporating thermodynamic constraints with metabolic networks, as shown in the accompanying poster "Numerical Optimization Algorithms and Software for Systems Biology". This work sets the stage for the goal of thermodynamically favorable pathways for hydrogen production by *Thermotoga maritima*.

233

The Ribosomal Database Project: Tools and Sequences for rRNA Analysis

J.R. Cole* (colej@msu.edu), Q. Wang, B. Chai, J. Fish, E. Cardenas, R.J. Farris, D.M. McGarrell, G.M. Garrity, and **J.M. Tiedje**

Michigan State University, East Lansing

Project Goals: The Ribosomal Database Project (RDP; http://rdp.cme.msu.edu) offers aligned and annotated rRNA sequence data and analysis services to the research community. These services help researchers with the discovery and characterization of microbes important to bioenergy production, biogeochemical cycles, including those affecting carbon and climate, and bioremediation.

Updated monthly, the RDP maintained 1,281,097 aligned and annotated quality-controlled rRNA sequences as of December 2009 (Release 10.17). The *my*RDP features have grown to support a total of over 2,500 active researchers using their *my*RDP accounts to analyze over 4,000,000 pre-publication sequences in 38,708 sequence groups, while the RDP Pyrosequencing Pipeline has been used by over 700 researchers to analyze next-generation sequences.

New NCBI/EMBL Short Read Archive Submission Tool: Because it is very difficult for researchers to submit their next-generation rRNA sequence data to the three INSDC databases (GenBank, EMBL, and DDBJ), RDP developed a combination of web and downloadable programs, the *my*RDP SRA PrepKit, to allow users to prepare and edit their submissions. This package provides an effective solution to the difficult and confusing process involved in preparing metadata documents that are required for submission to the GenBank Short Read Archive (SRA) or EMBL European Read Archive (ERA), the two databases for reads generated from ultra-high-throughput sequencing technologies. It can be applied on sequence data generated from 454 (GS 20, FLX and Titanium), Illumina/Solexa, ABI SOLiD and Helicos platforms. It transforms the

preparation of six separate XML document types required for each submission into a clear flow of tasks implemented in easy-to-understand forms for collecting metadata about the study, samples, experiments, analyses, and sequencing runs. A set of suggested attributes in the data forms assist researchers in providing metadata conforming to the MIMS Minimal Information about a Metagenome Sequence specification, and the upcoming MEINS Minimal Information about an Environmental Sequence specifications (Field et al., 2008, *Nat. Biotechnol.* 26:541; http://gensc.org/gc_wiki/index.php/MIENS). The user can save unfinished work for later sessions and copy individual components to new submissions to avoid repetitive entry of shared data. In addition, a provided Java Web Start program creates a Fastq file from sequence reads. The *my*RDP Submission Web Start program makes it easy to perform the tasks needed to finalize your submission. A help page outlining the workflow is also provided. (The USDA provided additional funding for the *my*RDP SRA PrepKit.)

RDP Pyrosequencing Pipeline: This toolkit has been used by 777 researchers (unique e-mail addresses) to analyze their next-generation sequence data. This pipeline offers a collection of tools that automate the data processing and simplify the computationally intensive analysis of large sequencing libraries. A number of new functions have been developed for the pipeline, including a new distance matrix tool that generates distance matrices in two popular formats used by third-party tools such as Mothur (Schloss, 2009, *Appl. Environ. Microbiol.* 75:7537). All the tools now accept compressed files to reduce the upload time of large amounts of sequence data. The Initial Process, Aligner, and Clustering tools have been enhanced to return graphical summary files that provide a visual representation of sequence quality and diversity. (The USDA and NIEHS provided additional support for the RDP Pyrosequencing Pipeline.)

Other RDP tools have been used, on average, in **18,633 analysis sessions per month** by an average of **5,634 researchers** (unique IPs). These include the **RDP Classifier**, which is also available as an open-source package through SourceForge and has been **downloaded 729 times**, the online Infernal secondary-structure based aligner (Nawrocki, 2009, *Bioinformatics* 25:1335) trained by RDP on representative bacterial and archaeal alignments, the **RDP Sequence Match** program for finding nearest neighbors, the **RDP Library Compare** program for determining differentially represented taxa between two environmental libraries, the **RDP Probe Match** program for determining taxonomic coverage of primers and probes, the **RDP Tree Builder** for rapid phylogenetic tree construction, and the **RDP Hierarchy Browsers** that provide entry to the RDP sequences in taxonomic order, by publication, or by completed genome (many genomes contain multiple rRNA operons). A **new RDP Multi-Classifier** is being provided as a command-line tool to accommodate the growing need for taxonomy-based analyses of large numbers of sequences in multiple samples. This tool combines the functions of both RDP Classifier and Library Compare, and thus provides a convenient solution for researchers to use as standalone tools or to be integrated into their own analysis workflow.

RDP Web Services have been expanded to provide interfaces for the RDP Classifier, Sequence Match, Probe Match and *my*RDP tools. There are, on average, **198,632 SOAP requests** received per month. Usage examples are provided in Java and Ruby. Researchers can incorporate these web services into their own analysis pipelines to make use of these popular RDP tools.

This research is supported by the Office of Science (BER), U.S. Department of Energy under Grant No. DE-FG02-99ER62848.

References

1. Cole, J. R., Q. Wang, E. Cardenas, J. Fish, B. Chai, R.J. Farris, A.S. Kulam-Syed-Mohideen, D.M. McGarrell, T. Marsh, G.M. Garrity, and J.M. Tiedje. 2009. The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res.* 37 (Database issue): D141-D145; doi:10.1093/nar/gkn879.
2. Wang, Q, G.M. Garrity, J.M. Tiedje, and J.R. Cole. 2007. Naïve Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. *Appl. Environ. Microbiol.* 73:5261-7; doi:10.1128/AEM.00062-07.

234

Identifying Proteins from Microbial Communities

William R. Cannon[1]* (William.Cannon@pnl.gov), Mitchell Rawlins,[1] Gaurav Kulkarni,[2] Andy Wu,[2] Ananth Kalyanaraman,[2] Douglas Baxter,[1] Mary Lipton,[1] and Steven Callister[1]

[1]Pacific Northwest National Laboratory, Richland, Wash. and [2]Washington State University, Pullman, Wash.

Project Goals: See below.

The lack of reliable genome sequences currently limits the effectiveness of proteomics studies of microbial communities because of the difficulty in identifying peptides. Characterizing the proteomics of microbial communities requires (1) the computational interpretation and integration of high-throughput experimental data, (2) the leveraging of existing sources of knowledge from multiple domains, and (3) searching for solutions that meet criteria on multiple levels in a large search space. Our goal is to develop novel methods needed to describe the proteins and metagenomic functional processes occurring within unsequenced microbial communities being investigated as part of DOE's missions in carbon sequestration, bioremediation and bioenergy research.

Student Presentation

235

Identifying the Mediators of Environmental Changes Through Integration of Steady State and Time-Course Gene Expression Profiles in *Shewanella oneidensis* MR-1

Qasim K. Beg,[1] Mattia Zampieri,[2,5] Sara Baldwin[2]* (baldwin2@bu.edu), Niels Klitgord,[2] Margrethe H. Serres,[4] Claudio Altafini,[5] and **Daniel Segrè**[1,2,3]

[1]Dept. of Biomedical Engineering, [2]Bioinformatics Program, and [3]Dept. of Biology, Boston University, Boston, Mass.; [4]Josephine Bay Paul Center, Marine Biological Laboratory, Woods Hole, Mass.; and [5]International School for Advanced Studies, Trieste, Italy

Project Goals: In this study, we combine mRNA microarray and metabolite measurements with statistical inference and dynamic flux balance analysis to study the transcriptional response of *S. oneidensis* as it passes through exponential, stationary, and transition phases. By measuring time-dependent mRNA expression levels during batch growth of *S. oneidensis* MR-1 under two radically different nutrient compositions, we obtain detailed snapshots of the regulatory strategies used by this bacterium to cope with gradually decreasing nutrient availability.

The dynamics of transcriptional regulation in microbial growth is an environment-dependent process. This dynamics is strongly controlled by two main factors: the wiring of the underlying regulatory network, and the time-dependent array of environmental stimuli. Understanding the interplay between these two factors is a fundamental challenge in systems biology, particularly relevant for the study of microbial systems, often adapted to rapidly changing environments. Certain genes may be activated as a response to the lack of a specific nutrient, and therefore display a strong dependence on environmental conditions; others may be more generally associated with growth rate, or growth phase requirements, and could therefore show similar behavior across different media. We address these questions in the environmental microbe *Shewanella oneidensis* MR-1, whose versatile respiratory functions make it a key player in environmental and bioenergy applications.

In this study, we combine mRNA microarray and metabolite measurements with statistical inference and dynamic flux balance analysis to study the transcriptional response of *S. oneidensis* as it passes through exponential, stationary, and transition phases. By measuring time-dependent mRNA expression levels during batch growth of *S. oneidensis* MR-1 under two radically different nutrient compositions (minimal lactate medium and LB medium), we obtain detailed snapshots of the regulatory strategies used by this bacterium to cope with gradually decreasing nutrient availability. In addition to traditional clustering, which provides a first indication of major regulatory trends and transcription factors activities, we implement a new approach for Dynamic Detection of Transcriptional Triggers (D2T2). This new method allows us to infer a putative topology of transcriptional dependencies, with special emphasis on the nodes at which external stimuli are expected to affect the internal dynamics. In parallel, we address the question of how to compare transcriptional profiles across different time-course experiments. Our growth derivative mapping (GDM) method makes it possible to relate with each other points that correspond to the same relative growth rate in different media. This mapping allows us to discriminate between genes that display an environment-independent behavior, and genes whose transcription seems to be tuned by specific environmental factors.

Several observed transcript time-courses raise interesting biological questions. For example, we observe a coupling between nitrogen-related genes and the glycogen biosynthesis/degradation pathway. To help rationalize the observed patterns, we measure extracellular metabolites and show how transcription and metabolism can be interpreted in the context of a dynamic flux balance analysis model.

236

Computational Design of Microbial Cross-Feeding Induced by Synthetic Growth Media

Niels Klitgord[1]* (niels@bu.edu) and **Daniel Segrè**[1,2] (dsegre@bu.edu)

[1]Program in Bioinformatics and [2]Dept. of Biology and Dept. of Biomedical Engineering, Boston University, Boston, Mass.

http://prelude.bu.edu

Project Goals: We seek to develop algorithm for engineering novel microbe-microbe interactions. Our method, based on stoichiometric genome-scale models of metabolism, is aimed at identifying environment that induce cross-feeding interactions. We envisage that such a "synthetic ecology" approach will be relevant for environmental and bioenergy applications.

Microbial ecosystems are ubiquitous on our planet, and play a major role in the global balance of the biosphere, as well as in the ongoing efforts for establishing renewable bioenergy sources. Since most microbe-microbe and microbe-environment interactions are likely mediated by metabolic intermediates, understanding the flow of metabolism between microbes constitutes a fundamental unsolved challenge. Here, towards addressing this challenge, we show how stoichiometric genome-scale models of metabolism can be extended to the ecosystem level, helping identify, understand and engineer interactions between pairs of microbial species. Specifically, we propose a novel suite of algorithms that can identify artificial environments predicted to induce mutualistic interactions between two given microbial species, by efficiently searching for growth media that sustain growth of two species only when simultaneously present. Our strategy is based on two major steps: *First*, we implement a procedure for automatically joining together the

stoichiometric models for two species, embedding them into a common environment. *Second*, we search the space of possible nutrient combinations for media that could not sustain growth of each species alone, but allow growth of both species simultaneously.

We validated our approach using three organism pairs of increasing complexity. The first is a simple toy model, in which one can arbitrarily pre-define expected mutualistic interactions. The second is a special case of the naturally occurring interactions between methanogenic archaea and hydrogen-producing microorganisms, which was recently analyzed in detail using flux balance models. The third is an experimentally engineered synthetic biological system of two yeast strains that can grow only in the presence of each other, because each of them is unable to synthesize a specific essential metabolite. In addition to recapitulating these known interactions, we will use our approach to generate new experimentally testable predictions of environments that induce interactions between pairs of environmentally relevant microorganisms, including *Shewanella oneidensis*. Selected predictions will be tested experimentally. We envisage that these algorithms will make it possible to engineer novel metabolism-based interactions between pairs of microbial species, helping develop a new computationally-driven synthetic ecology discipline

237

Multi-Scale Spatially Distributed Simulations of Microbial Ecosystems

William J. Riehl*[1] (briehl@bu.edu), Niels Klitgord,[1] Christopher J. Marx,[2] Nathaniel C. Cady,[3] and **Daniel Segrè**[1,4] (dsegre@bu.edu)

[1]Graduate Program in Bioinformatics, Boston University, Boston, Mass.; [2]Dept. of Organismic and Evolutionary Biology, Harvard University, Cambridge, Mass.; [3]College of Nanoscale Science and Engineering, University at Albany, N.Y.; and [4]Depts. of Biology and Biomedical Engineering, Boston University, Boston, Mass.

http://prelude.bu.edu

Project Goals: The goals of this project are: (1) to extend current genome-scale models to include spatio-temporal dynamics, (2) to allow more realistic simulations of microbial growth for individual species and ecosystems; and (3) to enable open source development of models for the study of renewable bioenergy sources, bioremediation challenges and ecosystem balance.

Genome-scale models of microbial metabolism represent the most advanced synthesis of genomic information, biochemical knowledge, and computational efficiency relevant for developing a predictive, quantitative understanding of microbial ecosystems. These models are becoming increasingly relevant for use in a number of endeavors, such as bioenergy production, bioremediation, and carbon and nitrogen cycling in the biosphere. As automated annotation pipelines, network gap-filling algorithms, and high throughput experimental methods improve, we will gradually approach the capacity to model virtually any sequenced microbe using this approach. Yet, some of the most fundamental properties of natural microbial ecosystems crucially depend on aspects that are well beyond the stoichiometries of individual biochemical species. These include contact- or metabolite-mediated interactions between different microbes, dynamical changes of the environment, spatial structure of the underlying geography and evolutionary competition between distinct subpopulations.

We present the early stage development of a broadly applicable and user-friendly platform for modeling these interactions by performing spatially distributed time-dependent flux balance based simulations of microbial ecosystems. We use a modified version of dynamic flux balance analysis (dFBA) to implement the dynamics of the system. By taking advantage of the computational efficiency involved in flux balance model calculations, we implement a spatially structured lattice of interacting metabolic subsystems. These subsystems represent a level of detail that is intermediate between a fine-grained single-cell modeling approach, and a broad global population modeling approach, and performs akin to a cellular automaton.

This platform has been developed with the capacity to bridge multiple spatial and temporal scales, making it possible to observe long term dynamics of microbial populations growing in a given environmental setting, based on constant updates of local nutrient availabilities and exchanges, and ultimately determined by the activity of individual metabolic reactions present in each microbial species. Thus, it can be used as a platform for modeling the spatial and temporal growth of a single bacterial species in a Petri dish, biofilm formation on complex substrate morphologies, seasonality of microbial communities in a specific geographical setting, or the growth and diffusion of a microbe that has been genetically engineered toward bioremediation in a contaminated body of water.

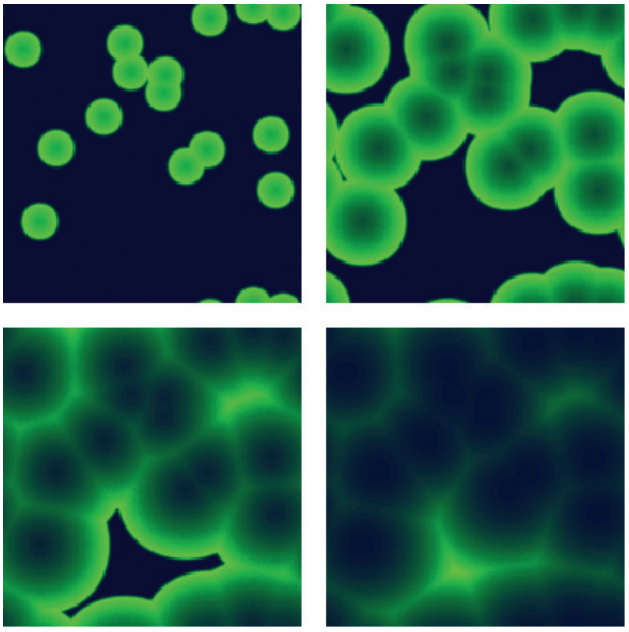


Figure 1. A sample run of the modeling platform, showing the effects of a model of *E. coli* colonies growing and merging together on a 2D surface with limited nutrient.

We present a prototype of our platform, which uses the open-source GNU Linear Programming Kit (GLPK) for performing the dFBA calculations, and a Java-based language (Processing) for coordinating the simulations and rapid visualization. We have applied this prototype to the analysis of several different examples, including the growth of a single species in a 2-Dimensional environment (see Figure 1) and syntrophic growth of microbial species. In future work, computer simulations will be integrated with experiments, allowing us to (i) calibrate the simulation parameters towards faithful representation of microbial growth patterns, and (ii) perform pilot studies on microbial ecosystem dynamics.

238

Ground and Transition State Binding Calculations to Improve Cytochrome $P450_{BM3}$ Reactivity and Specificity

Costas D. Maranas,[1] George A. Khoury[1]* (khoury@psu.edu), Michael J. Janik,[1] Patrick C. Cirino,[1] and Ping Lin[2]

[1]Dept. of Chemical Engineering and [2]Materials Simulation Center, Pennsylvania State University, University Park

http://maranas.che.psu.edu

Project Goals: The main goal of this work is to assess the impact of molecular interactions between an enzyme and its substrate at the ground and transition states on reactivity and substrate specificity. The identified trends are currently used to inform a method for systematically re-designing Cytochrome $P450_{BM3}$ to hydroxylate ethane.

In this work, we introduce the combined use of ground state and transition state calculations to understand how specific mutations present in engineered variants of cytochrome $P450_{BM3}$ confer improved reactivity. The cytochrome $P450_{BM3}$ monooxygenase has been the target of extensive directed evolution by other groups. The fatty acid hydroxylase is functionally expressed at high levels in *E. coli* and has been engineered to convert small alkanes to their corresponding alcohols, with an emphasis in biofuel production. We first identified and calculated the ground and transition state structures for the rate-limiting step using quantum mechanical methods. Next, we computationally assessed the effects of 14 different experimentally isolated mutations in P450 mutant 535-h (3 mutations lie in the active site) on interactions with the ground and transition state structures with a newly developed computational saturation mutagenesis procedure. The general trend found was that some mutations are important for improving substrate binding, while other mutations in different positions are important for improving transition state stabilization. We find that calculations at both ground and transition state appear to be important for rational enzymatic design. In the design phase, we systematically chose design positions based on sequence, structure, and energetic factors, and customized the Iterative Protein Redesign and Optimization (IPRO) framework to identify the energetically optimal mutations with the ground and transition states. We report on the general trends from the optimal designs predicted by IPRO.

Student Presentation

239

Improving Metabolic Models Using Synthetic Lethality Data and Generating Genome-Scale Isotope Mapping Models for Flux Elucidation

Patrick F. Suthers[1]* (suthers@engr.psu.edu), Alireza Zomorrodi,[1] Prabhasa Ravikirthi,[2]* and **Costas D. Maranas**[1]

[1]Dept. of Chemical Engineering, and [2]Dept. of Cell and Developmental Biology, Pennsylvania State University, University Park

http://maranas.che.psu.edu

Project Goals: The project goal of the research described here is twofold: First to improve the quality of genome-scale metabolic models by making use of gene essentiality and synthetic lethality experimental data. The second goal is to combine existing metabolic reconstructions with information from atom transitions to generate genome-scale isotope mapping models.

A pair of non-essential genes is referred to as synthetic lethal if the simultaneous deletion of both genes is lethal but the single gene deletions are not. One can generalize the concept of synthetic lethality to reactions or extend it further

by considering gene/reaction groups of increasing size where only the simultaneous elimination of all genes/reactions is lethal. Previous studies have demonstrated the utility of synthetic lethal predictions for the curation of genome-scale metabolic models. We recently used synthetic lethality information to identify twenty-one model improvements for the genome-scale model of *Escherichia coli*, *i*AF1260. In this talk, we discuss the systematic identification of synthetic lethal gene combinations for the most recent genome-scale metabolic model of yeast, (i.e., *i*MM904) for a variety of different growth medium conditions. By contrasting the *in silico* lethality predictions with *in vivo* observations we identified/corrected many missing regulatory mechanisms in yeast. The incorporation of the altered regulatory mechanisms into the genome-scale metabolic model led to a substantial increase in the accuracy of the *in silico* gene essentiality predictions. Overall, this study demonstrates the utility of synthetic lethality information for correcting genome-scale metabolic models.

Metabolic flux analysis (MFA) has so far been restricted to lumped networks lacking many important pathways, partly due to the difficulty in automatically generating isotope mapping matrices for genome-scale networks. Here we describe a procedure for the largely automated generation of atom mappings for genome-scale metabolic reconstructions. The developed procedure uses a compound matching algorithm based on the graph theoretical concept of pattern recognition along with relevant reaction heuristics to automatically generate genome-scale atom mappings which trace the path of atoms from reactants to products for every reaction in any given reconstruction. When applied to the *i*AF1260 metabolic reconstruction of *Escherichia coli*, the genome-scale *i*sotope *m*apping model *im*PR90068 is obtained. The model maps 90,068 non-hydrogen atoms, contains 1.37 x 10^{157} distinct isotope forms and accounts for all 2,077 reactions present in *i*AF1260 (the previous largest mapping model included 238 reactions). The expanded scope of *im*PR90068 allows for tracking of labeled atoms through pathways such as cofactor and prosthetic group biosynthesis and histidine metabolism. We also discuss how using an elementary metabolite unit (EMU) representation of *im*PR90068 significantly reduces the number of variables during MFA.

240

Computational Pathway Identification and Strain Optimization for Chemical and Biofuel Production

Sridhar Ranganathan[1]* (sur152@psu.edu), Patrick F. Suthers,[2] and **Costas D. Maranas**[2]

[1]Huck Institutes of the Life Sciences and [2]Dept. of Chemical Engineering, Pennsylvania State University, University Park

http://maranas.che.psu.edu

Project Goals: The main goal of this work is to develop new methods to discern novel pathways for chemical and biofuel production and to elucidate strain engineering strategies that will ensure production at desired target levels.

We present an integrated computational base to support pathway identification and strain optimization with an emphasis on biofuel production. An efficient graph-based algorithm is presented for the exhaustive identification of all pathways enabling the production of a targeted biofuel molecule. The algorithm is based on a min-path formulation. It searches over a database of biotransformations that spans reactions from KEGG, Metacyc, BRENDA and other resources with an emphasis on C4+ alcohols. The identified pathways are then integrated into the genome-scale model of the production host (*e.g.*, *Escherichia coli*). We describe the application of the OptForce computational framework to pinpoint engineering modifications (knock-outs/up/down) that are required for the targeted biofuel overproduction. This is accomplished by classifying reactions (and combinations thereof) in the metabolic model depending upon whether their flux values must increase, decrease or become equal to zero to meet the pre-specified overproduction target. A "force set" can then be extracted that contains a sufficient and non-redundant set of reactions that need to be directly changed to meet the production requirements. We apply the integrated framework for the production of 1-butanol, isobutanol, and other alcohols in *E. coli* using the most recent *in silico E. coli* model, *i*AF1260. We also examine the production of succinate in *E. coli*. The proposed computational workflow not only recapitulates existing pathways and engineering strategies but also reveals novel and non-intuitive ones that boost production by using and performing coordinated changes on sometimes distant pathways.

241

COBRA Toolbox 2.0: *In Silico* Systems Biology Suite

Jan Schellenberger[1]* (jschelle@ucsd.edu), Richard Que,[2] **Andrei Osterman**,[3] **Bernhard O. Palsson**,[2] and **Karsten Zengler**[2]

[1]Bioinformatics Program, and [2]Dept. of Bioengineering, University of California, San Diego; and [3]Burnham Institute for Medical Research, La Jolla, Calif.

http://systemsbiology.ucsd.edu/Downloads/Cobra_Toolbox

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

grated" genome-scale model, 3) to confirm and validate the ability of the integrated model to predict processing of various environmental signals.

With the advent of whole genome sequencing in the late 1990s, it became possible to build genome scale metabolic models. Since then, this field has undergone a renaissance in terms of 1) size and scope of reconstructions, 2) number of reconstructions and 3) number of analysis tools. The first version of the COBRA (Constraint Based Reconstruction and Analysis) toolbox was published in 2007 to combine many of these emerging methods into one easy to use package. We present version 2.0 here.

The COBRA toolbox is a set of Matlab scripts. Constraint Based models are loaded from various sources into a COBRA specific data structure. The user can then manipulate these models by using the command line or simple scripts. Methods can be chained to create simple data pipelines. The scope of COBRA falls under 8 categories as shown in Figure 1. New to version 2.0 are methods for gap filling, C13 analysis, visualization and thermodynamics. Also new in version 2.0 is a test case suite which gives examples of use of the different methods and expected results.

The objective of the COBRA toolbox is to abstract away details of implementation of constraint based methods. For the end user this reduces development time, cuts down on bugs and makes the code easier to share with other research groups. For the *Thermotoga* project, the COBRA toolbox is used to refine the model, analyze high throughput data, and visualize the results.

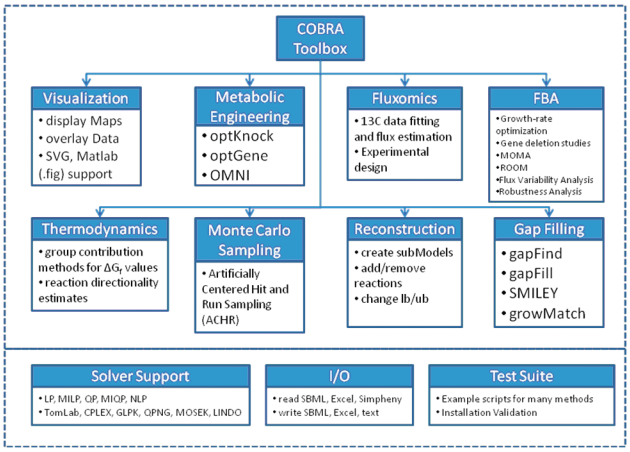


Figure 1: Features of the COBRA toolbox: Top) Scripts are available for methods in eight areas of metabolic systems biology. Bottom) Linear and Quadratic solvers are implemented through a simple yet flexible API in a vendor independent fashion. A set of test scripts are present to validate proper installation as well as demonstrate examples of use.

Reference

1. Becker, S.A., Feist, A.M., Mo, M. L., Hannum, G., Palsson, B.Ø., Herrgard, M.J. Quantitative prediction of cellular metabolism with constraint-based models: The COBRA Toolbox., *Nat. Protocols*, 2, 727-738 (2007).

Student Presentation

242

Numerical Optimization Algorithms and Software for Systems Biology: Optimality Principles in Nonequilibrium Biochemical Networks

Ronan M.T. Fleming[1]* (ronan.mt.fleming@gmail.com), Chris Maes,[2] Ines Thiele,[3] **Bernhard Ø. Palsson**,[4] **Yinyu Ye**,[5] and **Michael A. Saunders**[5]

[1]Science Institute and Center for Systems Biology and [2]Center for Systems Biology, University of Iceland, Iceland; [2]Institute for Computational and Mathematical Engineering, Stanford University, Stanford, Calif.; [3]Dept. of Bioengineering, University of California, San Diego; and [4]Dept. of Management Science and Engineering, Stanford University, Stanford, Calif.

http://www.hi.is/~rfleming

Project Goals: Simultaneous prediction of metabolic fluxes and concentrations in *Eschericia coli*.

We derive a new optimization problem on a steady-state non-equilibrium network of biochemical reactions, with the property that mass conservation, energy conservation, the second law of thermodynamics and the proportionality of reaction rate to reactant concentration, all hold at the problem solution. These nonlinear, non-convex constraints are enforced without recourse to linearization or any other form of approximation. This method provides the first computationally tractable method for enforcing thermodynamic, energy, and mass-conservation constraints, at genome scale. Moreover, the formalism has a clear thermodynamic interpretation and suggests a new optimality principle for non-equilibrium biochemical networks. This method may be used for simultaneously predicting reaction rate (flux) and metabolite concentrations in genome-scale biochemical networks. In particular, we demonstrate its utility for simultaneous integration of metabolomic and fluxomic data in *Escherichia coli*, in order to predict unmeasured concentrations and fluxes.

Communication and Ethical, Legal, and Societal Issues

243

Bacterial Toxicity of Engineered Metal and Metal Oxide Nanoparticles

A.K. Suresh[1]* (sureshak@ornl.gov), D.A. Pelletier,[1] K. Giri,[1] S.D. Brown,[1] B. Gu,[2] W. Wang,[2] D.P. Allison,[1,3] D.C. Joy,[3,4] M. Ji-Won,[1] T.J. Phelps,[1] and **M.J. Doktycz**[1]

[1]Biosciences Division, [2]Environmental Sciences Division, and [4]Center for Nanophase Materials Sciences, Oak Ridge National Laboratory, Oak Ridge, Tenn.; and [3]University of Tennessee, Knoxville

Project Goals: (1) Synthesis and characterization of metal and metal oxide nanoparticles. Nanoparticles containing Au, Pd, and ZrO_2 will be prepared by wet-chemical precipitation techniques. The size and composition of these structures will be varied and characterized by techniques such as dynamic light scattering, XRD, AFM, SEM and high-resolution TEM. Well-defined nanoparticle structures are essential for drawing accurate conclusions on how physical and chemical characteristics influence microbial response. (2) Assessment of microbial growth and viability from metal and metal oxide nanoparticle exposure. The goal of this aim will be to assess how the size, composition, and concentration of engineered nanoparticles impact microbial growth and the microbe's molecular response to these exposures. We will focus on *Escherichia coli* K12, *Bacillus subtilis*, and the metal-reducing species *Shewanella oneidensis* MR1 as initial model organisms.

Nanomaterials are of tremendous interest to pursuits in biology, medicine, electronics, catalysis and energy storage because of their unique size and shape dependent properties. Characteristics such as high surface to volume ratios, quantum confinements and the ability to selectively mediate chemical transformations make them unique from their bulk counterparts. Besides size, composition, surface coat and surface charges are properties that affect nanomaterials performance and may affect their fate and transport in the environment. The transformation of such nanoparticle catalysts in the environment is likely to influenced through interactions with bacteria. Nanoparticle production, nanoparticle toxicity, nanoparticle binding and incorporation with bacteria have all been observed. However, basic knowledge that would allow prediction of the probable interaction between an engineered nanoparticle and bacteria is lacking. Our efforts seek to quantify and characterize interactions between engineered metal and metal oxide nanoparticles on selected microbial species. Initial efforts are focused on the effects of cerium oxide, zinc oxide and silver nanoparticles on the growth, viability, structural changes and genetic response of *E. coli* and *B. subtilis*. Well-characterized CeO_2, ZnO and Ag nanoparticles have been prepared and presented to bacterial cells in a dose dependent manner. Apart from basic routine techniques such as disc diffusion tests, minimum inhibitory concentrations, viability assays, and colony forming units, advanced imaging techniques like TEM and AFM are also being used to evaluate the binding and fate of the nanoparticles on the bacterial cell. From our studies it is evident that most nanoparticles cause toxicity to the bacteria by interacting with them and causing perforations and aberrations which is direct interaction in some (Ag) and reactive oxygen species mediated in others (ZnO). The results of these studies will provide a basis for understanding how nanoparticle size and composition influence their interactions with microorganisms, and how microorganisms may alter the fate and transformation of engineered nanoparticles in the environment.

Student Presentation

244

Mechanisms for Transnational Coordination and Harmonization of Nanotechnology Governance

Gary E. Marchant (gary.marchant@asu.edu), Kenneth W. Abbott, Douglas J. Sylvester, Michaelyn Gulley,* Blair Moses, Vanessa Lancaster, and Keenan Whanamaker

Center for the Study of Law, Science, and Technology and the Sandra Day O'Connor College of Law, Arizona State University, Tempe

Project Goals: This project will analyze and explore the potential for transnational coordination and harmonization of regulatory and other oversight mechanisms for nanotechnology, focusing specifically on nanotechnology applications relating to bioenergy. The specific goals are to: (1) create and maintain a publicly available online database of proposed and enacted regulatory requirements and programs specific to nanotechnology at the international, national, and sub-national levels; (2) analyze proposed and enacted national and sub-national regulations for nanotechnology, including the consistency (or lack thereof) of regulatory requirements in different jurisdictions, using nanotechnology-related applications of bioenergy as the regulatory focus; and (3) prepare case studies of nine different transnational precedents or models for the regulation or oversight of various technologies, including an analysis of the strengths, weaknesses, and lessons of each model.

Nanotechnologies are a rapidly developing set of emerging technologies being pursued by industry and governments

around the world. While these technologies promise many benefits, they will also inevitably create some risks, and regulatory agencies in numerous countries are now considering regulatory oversight approaches for nanotechnology. This project is investigating models and approaches for coordinating or harmonizing international regulation of nanotechnology. The first step in the project is to create a publicly-accessible online database of transnational, national, and sub-national regulatory activities specific to nanotechnology. This part of the project will go "live" as a publicly-available and searchable online database in July 2009. The second part of the project is to identify and analyze nine different regulatory models for transitional oversight of nanotechnologies.

245

Linking Ethical, Legal, and Societal Implications (ELSI) Analyses to Nanoscale Science and Technology

Amy K. Wolfe[1]* (wolfeak@ornl.gov), Stephanie Wang,[2] David J. Bjornstad,[1] W. Christopher Lenhardt,[1] and Mitchel Doktycz[1]

[1]Oak Ridge National Laboratory, Oak Ridge, Tenn. and [2]Grinnell College, Grinnell, Iowa

Project Goals: Oak Ridge National Laboratory's Ethical, Legal, and Social Issues (ELSI) Scientific Focus Area (SFA) is a DOE-SC focused research activity that analyzes systematically the dynamic and evolving societal implications of emerging bioenergy and nanotechnology S&T.

One tenet of Oak Ridge National Laboratory's (ORNL's) Ethical, Legal, and Societal Implications (ELSI) research associated with DOE's nanoscale science user centers is that analyses should add value to the nanoscale science research and development (R&D) community. By "add value," we mean providing information and insights that are useful in framing R&D questions, managing the flow of information and intellectual property within and outside of the DOE nanoscale science user centers, and aiding scientists who engage in forums with other audiences. By "adding value," we do *not* aim to contribute to or alter any specific experiment or research; advocate for or against any technology, facility, or population; or in any way "manage" or affect perceptions and opinions.

In what ways might ELSI analyses be salient to the nanoscale science community? While the answer to this question is not immediately obvious, we suggest that the following elements are key: (1) strong, explicit links to nanoscale science and technology (S&T); (2) attention to nano product- and application-type; and (3) consideration of phase along the R&D life cycle (e.g., research vs. deployment vs. decommissioning). Before embarking on investigations whose design specifies these elements, we conducted an analysis to learn the extent to which existing nano-related ELSI publications do so. This poster presents results from that screening analysis and suggests potential implications with regard to nanoscale science and, by extension, genomics science communities.

We screened 68 nano-ELSI publications. There was considerable variation in the primary issues addressed in these documents, with the highest concentrations on ELSI, perception, equity, and governance issues. Most authors emphasized various portions of the research through deployment portion of the R&D life cycle, with very few addressing disposal, decommissioning, or the full life cycle. Perhaps most significantly, the vast majority of publications did not link their analyses either to particular nanoscale materials or processes or to specific nano-products or applications. Thus, there is a large gap in the nano-ELSI literature that our investigations will start to fill.

246

Implications of Alternative Intellectual Property Rights Management Approaches

David J. Bjornstad* (bjornstaddj@ornl.gov)

Oak Ridge National Laboratory, Oak Ridge, Tenn

Project Goals: This project is part of Oak Ridge National Laboratory's (ORNL's) Ethical, Legal, and Social Issues (ELSI) Scientific Focus Area (SFA), which seeks to estab¬lish a unique ELSI resource for the U.S. Department of Energy (DOE) and its Office of Science (SC).

DOE maintains a vigorous technology transfer program as a key element of its intellectual property (IP) activities. As part of the management strategy for administering its Bioenergy Research Centers, DOE has issued a set of Principles to guide IP practices for new findings emanating from the Centers. The Bioenergy Research Centers, in turn, responded with Management Plans for implementing the Principles. The Principles and Plans are innovative, far-reaching and represent a departure from past practice. If successful, DOE might consider them for other similar partnered research ventures. For these reasons, and because they are "zero- based," that is, instituted from the initiation of the Bioenergy Research Centers Program, the Plans offer a unique opportunity to document how they are implemented, how the incentives they embody influence research partners and potential licensees, and the extent to which they could be adjusted for application to other situations. This poster presents findings from our investigation of the implications of evolving and differing implementation of the Principles and Plans for the conduct of science at the three Bioenergy Research Centers and the flow of information and IP from the Centers.

247

CGE Approach to Estimating Employment, Income, and Revenue Impacts of Biofuel Mandates in Pacific Northwest Regions

Olga V. Livingston* (olga.livingston@pnl.gov), **David M. Anderson** (dma@pnl.gov), and Ryan S. Butner (Ryan.Butner@pnl.gov)

Pacific Northwest National Laboratory, Richland, Wash.

Project Goals: Quantifying economic and social impacts of biofuel mandates.

Biofuel mandates in the Pacific Northwest are expected to have a significant influence on agriculture with direct impacts on land use. One of the primary economic factors driving land allocation among crops is expected to be land rent. There are multiple economic forces at play that jointly determine returns on land. With prices and demand for various land classes being endogenous to the model, Computable General Equilibrium (CGE) has been chosen as the methodology to analyze effects of biofuel mandates in the Pacific Northwest.

The larger project of quantifying economic and social impacts of biofuel mandates consists of two essential parts. The first part utilizes CGE model to determine demand and price changes for crops, which would result from the increased demand for biofuel due to the mandated mixing ratio. This model also accounts for the changes in underlying production functions as a result of boost in yields driven by latest genomic research. The second part uses the results of regional CGE analysis as an input to determine patterns of the land use due to changes in the returns on land, based on various allocation scenarios.

This discussion covers the first part of the analysis outlined above. Primary study areas include the following economic regions in Idaho, Oregon, and Washington: The Palouse, Middle Columbia, Umatilla/Walla Walla, Willamette Valley, and the Rogue Valley. Calibration of Social Accounting Matrix (SAM) tables is developed from the IMPLAN database (MIG 2009). The model assumes producers maximize profits with output being sold either to the domestic (regional) market or on the export market. Analogously, domestic supply comes from both regional production and imports. Households also consume a mix of goods from both sources. Producers are modeled by a combination of Constant Elasticity of Substitution and Leontief production functions. Results of the GTL research impact the production side of the model the form of possible scenarios with increased crop yields. Each scenario enters as a modified production function at the level of intermediate inputs. The solution for each scenario is a unique set of equilibrium prices that determines the direction and magnitude of changes in employment, income, taxes and overall welfare.

248

Meeting the Demand for Biofuels: Impact on Land Use and Carbon Mitigation

Madhu Khanna* (khanna1@illinois.edu), Atul Jain, Hayri Onal, Haixiao Huang, Seungmo Kang, Xiaoguang Chen, and Matthew Erickson

University of Illinois, Urbana-Champaign

Project Goals: This integrated, interdisciplinary research project investigates the impact of meeting the biofuel mandates on land use, crop production, social welfare and the environment in the U.S. over a 15 year horizon. To this end, we conduct a variety of research activities including: (1) examination of the optimal allocation of existing cropland for feedstock production, the mix of feedstocks that should be produced, and the spatial pattern of land use in the U.S. to meet specified levels of renewable fuel production under a variety of policy scenarios; (2) determination of the productivity in terms of yield and greenhouse gas mitigation benefits for each type of feedstock both in the form of soil carbon sequestration and displacement of carbon emissions from gasoline; and (3) identification of the optimal size and location of biorefineries and the transportation network that is consistent with regional feedstock production patterns and the location of demand for ethanol.

This integrated, interdisciplinary research project investigates the impact of meeting the biofuel mandates on land use, crop production, social welfare and the environment in the U.S. over a 15 year horizon. To this end, we conduct a variety of research activities including: (1) examination of the optimal allocation of existing cropland for feedstock production, the mix of feedstocks that should be produced, and the spatial pattern of land use in the U.S. to meet specified levels of renewable fuel production under a variety of policy scenarios; (2) determination of the productivity in terms of yield and greenhouse gas mitigation benefits for each type of feedstock both in the form of soil carbon sequestration and displacement of carbon emissions from gasoline; and (3) identification of the optimal size and location of biorefineries and the transportation network that is consistent with regional feedstock production patterns and the location of demand for ethanol.

Wide-scale economic analysis of biofuel production is heavily dependent on development of crop growth models capable of accurately simulating cellulosic feedstock growth at a wide scale. We have developed a continental-United States scale biophysical crop growth model based on the ALMANAC model to produce fine resolution (0.1˚ grid scale) estimates of growth for various cultivars of switchgrass and *Miscanthus ×giganteus*, along with corn and soybeans for comparison. We integrate the spatially variable simulated yields of energy crops from this biophysical model with a dynamic, multi-market Biofuel and Environmental Policy Analysis Model (BEPAM) which includes markets for fuel, biofuel, food/feed crops and livestock for the period

2007-2022. We consider biofuels produced not only from corn but also from corn stover, wheat straw, switchgrass and *miscanthus* as well as the potential for imports of sugarcane ethanol from Brazil. Spatial heterogeneity in yields, costs of production and land availability is incorporated by using Crop Reporting Districts (CRD) as our decision making unit for economic analysis. A rolling horizon model is used in which crop producers form price expectations based on lagged prices and make crop choice and land use decisions for the next ten years. Food and fuel prices are endogenously determined annually and used to update price expectations and land use choices for decision-making. We incorporate temporal changes in yields and costs of production of perennial energy crops, switchgrass and *miscanthus*. Econometrically estimated acreage and yield functions are used to update land availability and yields in response to changes in crop prices and time trends. Life cycle analysis is used to estimate the greenhouse gas intensity of alternative fuels and the emissions due to changes in cropping patterns at the CRD level. Using the BEPAM model, we examine the impact of various biofuel policies such as biofuel mandates and subsidies and carbon taxes on U.S. agricultural land use, carbon mitigation and social welfare.

Finally, we have developed a mathematical modeling framework (specifically a linear mixed integer program) to identify the optimal transportation network and the location and capacity of biorefineries for a given spatial supply pattern of different bioenergy crops generated by the BEPAM model. The transportation/facility location model includes the transportation cost of biomass delivered from farms to refineries, the transportation cost of biofuel delivered from biorefineries to demand centers and the processing cost at biorefineries for a multi-year planning horizon. The model is applied to examine potential refinery location and transportation modes and networks at a county level within the Midwest given demand for ethanol and livestock feed throughout the U.S.

249

The Biofuels Revolution: Understanding the Social, Cultural, and Economic Impacts of Biofuels Development on Rural Communities

Theresa Selfa[1]* (tselfa@ksu.edu), Richard Goe,[1] Laszlo Kulcsar,[1] Gerad Middendorf,[1] and Carmen Bain[2]

[1]Dept. of Sociology, Kansas State University, Manhattan; and [2]Dept. of Sociology, Iowa State University, Ames

Project Goals: The goal of this project is to provide a better understanding of the socio-economic costs and benefits of biofuels development for rural communities in the Midwest, and to examine how different environmental and socio-economic factors contribute to the outcomes.

A new wave of economic growth is currently sweeping across rural communities in the Midwest region of the U.S., fueled by the construction and expansion of ethanol biorefineries and the expansion of biofuel crop production. While the expansion of the biofuels industry promises to bring jobs and economic vitality to rural communities, it is also creating dilemmas for farmers and rural communities in weighing the benefits of income growth and job growth against safety risks, increased pollution, and the potential of overextending water supplies. Presently, there is little empirical knowledge about the social and economic impacts of biofuels development on rural communities. This research is intended to help fill these lacunae through an in-depth analysis of the social and economic impacts of ethanol biorefinery industry on six rural communities in the Midwestern states of Kansas and Iowa. The goal of this project is to provide a better understanding of the socio-economic and cultural implications of biofuels development for rural communities, and to contribute to more informed policy development regarding bioenergy.

Research Questions:

1. To understand how the growth of biofuel production has affected and will affect Midwestern farmers and rural communities in terms of economic, demographic, and socio-cultural impacts.
2. To determine how state agencies, groundwater management districts, local governments and policy makers evaluate or manage bioenergy development in relation to competing demands for economic growth, diminishing water resources, and social considerations.
3. To determine the factors that influence the water management practices of agricultural producers in Kansas and Iowa (e.g. geographic setting, water management institutions, competing water-use demands as well as producers' attitudes, beliefs, and values) and how these influences relate to bioenergy feedstock production and biofuel processing.
4. To determine the relative importance of social-cultural, environmental and/or economic factors in the promotion of biofuels development and expansion in rural communities.

Research Methodology

We have analyzed data gathered for the first three case study communities. The comprehensive methodology includes: demographic analysis; in-depth key informant interviews, three focus groups with farmers, ethanol plant workers, and community leaders; a general population opinion survey of community residents; and a content analysis of local newspapers and print media. In the winter of 2010 we will be completing data collection for our final three case studies.

Preliminary Findings

Data from community level surveys, individual and focus group interviews in three case study communities in Iowa and Kansas in the Midwestern region of the United States are utilized to explore community perceptions about the biofuels industry. Results show that community members believe that ethanol plants have brought modest economic benefits to their community. Increased traffic and water competition were two areas of concern identified by

residents with respect to local ethanol plant, while other environmental impacts were not strongly identified. Widespread concerns were expressed about the future viability of the industry and the devastating impacts that future declines in the ethanol industry would have on communities. This research highlights the social vulnerabilities that communities where ethanol plants are located are experiencing.

For Additional Information:
Project information and research findings will be available at: http://www.ksu.edu/sasw/kpc/biofuels/project_doe.htm

250

Land Use Impacts of Second Generation Biofuel Mandates in the Pacific Northwest

J.S. LaRiviere* (jlarivie@ucsd.edu), D.M. Anderson, and V.L. Bailey

Pacific Northwest National Laboratory, Richland, Wash.

Project Goals: Biofuel mandates can have a significant effect on agricultural land use decisions (Bento and Landry 2009). This paper exploits a unique data set to analyze the impacts of changes in state and national biofuel mandates on agricultural land use decisions in the Pacific Northwest. This paper examines the effect of mandates on cellulosic biofuels, such as those derived from switchgrasses, on a farmer's discrete choice between type of crop and on the decision to convert "marginal" land into active farmland. Changes in land use decisions are found to be heavily dependent on three factors: yields in switchgrass production, variation in transportation costs to cellulosic biofuel processing plants, and changing biofuel prices due to different government mandates.

Biofuel mandates can have a significant effect on agricultural land use decisions (Bento and Landry 2009). Mandates influence prices, which in turn influences the relative profitability of different crop and land use decisions. While changes in relative prices drive land use decisions, idiosyncratic agricultural characteristics of individual parcels of land, such as access to irrigation, composition of soil nutrients, and tillage history, are critical to each individual farmer's crop decisions. The uniformity of land's agricultural characteristics across regions necessitates accounting for them in any study of the regional impacts of state or national biofuel mandates. This paper exploits a unique data set consisting of various agricultural characteristics of land in the inland Pacific Northwest to analyze the effect of mandates on second generation cellulosic biofuels, such as those derived from switchgrasses, on agricultural land use decisions.

Economic factors determine the relative profitability of farmers land use decisions. Most generally, crop yields depend on costly inputs, such as water. Lack of access to irrigation restricts the set of feasible crops and alters farming practices over crops that remain economically feasible, such as affecting rotations and tillage practices. Similarly, soil attributes affect the amount of fertilizer input needed in crop production. Further, past crop and tillage decisions influence current soil quality, and therefore influence land-use decisions. Biofuel mandates affect the supply of biofuel feedstocks by increasing the price which farmers receive for feedstocks in the marketplace. Additionally, the distance to biofuel processing plants influences the costs for biofuel feedstock production. Because agricultural characteristics like soil quality and access to water affect the relative profitability of second generation biofuel feedstocks, they must be accounted for in assessing the regional land use impacts of biofuel mandates.

The effect of second generation biofuel mandates on land-use decisions is subject to large amounts of uncertainty because feedstocks, like switchgrasses, have not been commercially produced and mandates have yet to influence prices. As a result this study simulates land use changes which would result from a range of potential outcomes informed by scientific study of switchgrass yields. We use a unique data set to account for regional agricultural attributes present in the inland Pacific Northwest. This paper examines on a farmer's discrete choice between type of crop and on the decision to covert "marginal" land into active farmland. Changes in land use decisions are found to be heavily dependent on three factors: yields in switchgrass production, variation in transportation costs to cellulosic biofuel processing plants, and changing biofuel prices due to different government mandates.

Reference

1. A Bento and J Landry, 2009. "Efficiency Effects of Increased Biofuel Mandates." *Working Paper*, Department of Applied Economics and Management, Cornell University.

251

Science Literacy Training for Public Radio Journalists

Bari Scott* (Bariscot@aol.com)

SoundVision Productions, Berkeley, Calif.

Project Goals: The purpose of the weeklong workshops is to teach public radio reporters, producers, editors and news directors about science, science journalism, and the creative use of radio and online technologies so that they can produce clear, accurate and engaging science stories.

To help journalists explain accelerating scientific advances and their profound social implications, SoundVision is offering another series of Science Literacy Training workshops. Our goal is to shrink the widening knowledge gap between the scientific community and the general public by increasing the number and quality of science stories on radio and the web. During the project's weeklong workshops experts teach public radio reporters about science, science journalism, and the creative use of radio and online technologies to produce clear, accurate and engaging science stories.

This project is timely. Broadcasters' growing commitment to report on science is reflected in our changing applicant pool—most of our applicants now work as science, environmental, health or technology reporters. As journalists become acutely aware of the need to tell science stories in a changing media landscape, more and more of them are looking to the Science Literacy Training project to teach them how. As a result, the challenge and scope of this project has had to change on every level. To remain relevant, we've had to evolve from teaching radio science reporting to mid-career radio journalists with similar skills to teaching a group of public radio and online journalists with diverse skills and specialties to communicate science on a variety of broadcast and multi-platforms.

We've outgrown our original project. Because the demand for this training will only increase, we must decide how the project should expand to meet it.

SoundVision selects up to twelve journalists to attend each workshop. As more people seek admission, it's becoming harder to choose from the growing pool of impressive candidates that includes journalists from all areas of public radio. Our applicants work on national programs including *The World*, *Studio 360, Living on Earth, Radio Lab,* and *Marketplace*. They come from NPR, statewide networks and major market and rural stations in the United States. We're also attracting a growing number of international applicants.

The variety in our workshop participants is striking. We've had a large influx of young Internet-savvy news people applying to the program, some of whom work exclusively in their stations' online departments, and we've broadened the curriculum to include more Internet reporting and production techniques. More of our candidates have advanced science and medical degrees than ever before, and we're attracting journalists who are new to radio but have extensive print experience; although they may have impressive journalism skills and scientific expertise, they may still be learning how to write for radio, use a microphone or handle recording equipment in the field.

SoundVision's challenge is to fill the gaps in each journalist's knowledge and provide all participants with an educational experience tailored to their individual requirements. To do that, we build the workshops around core content that everyone studies while designing additional courses to meet their individual needs and levels of expertise. We determine these needs through extensive pre-workshop questionnaires.

Major stations around the country are eager to host the workshops, and a number of universities, including Princeton, Santa Clara University, the University of Texas at Austin and University of California, Berkeley, have asked to hold the workshops and offered to help us identify and secure presenters from their own and other campuses. A workshop was held in the San Francisco Bay Area in April 2009 and another is scheduled there in April 2010.

Curriculum: Workshop leaders train participants in several key areas: science, including basic biology, chemistry, physics, statistics and nanotechnology, with a strong focus on energy-related issues; science journalism, including how to interview scientists, how to handle controversy and the standards and ethics of science reporting; radio craft, from story telling to audio and production tools, and web techniques and podcasting. Participants also go on field trips, attend informal gatherings with scientists and learn creative ways to deal with the unique limitations and advantages of radio and web production. Session leaders also teach participants how scientific methods differ from journalistic practices and show them how to explore new research and fact check stories on tight deadlines. In addition, online librarians show journalists how to make the best use of the Internet for research and identify reliable sources online. Finally, each participant's work is critiqued by expert radio producers and science editors.

To prepare participants to make the most of their experience while building a sense of community and excitement, SoundVision has also created a Google Group which is used in the weeks before the conference to send them weekly homework assignments and links to valuable articles and other resources. We also mail participants books and magazine articles to read. The homework and reading are designed to give all participants a baseline foundation in science, the scientific method and science journalism before they arrive.

The project also includes a web site that provides online resources, "tip sheets," and transcripts and selected audio from the training sessions. (For sample Tip Sheets, see www.scienceliteracyproject.org.) In response to requests, SoundVision schedules follow-up teleconferences and possibly a webinar with participants.

Evaluation: The Science Literacy Training workshops are evaluated by Rockman *ET AL*, a well-established San Francisco firm with expertise in assessing media projects and the impact of training on journalistic practice. Pre-workshop evaluations help us tailor presentations to participants' needs. During the workshops, we conduct daily evaluations in which we ask participants to list the key scientific concepts they have just learned to gauge their retention of the material. Approximately two months after the workshop, Rockman also conducts follow-up evaluation interviews with all the participants to see how much they've retained.

Rockman reported that the April 2009 workshops had achieved the right balance of sessions on science, science journalism and radio techniques. In their enthusiasm everyone wanted more time for Q&A with the science presenters. While they appreciated the depth and breadth of the material, some felt the schedule was packed too tightly because they wanted more time to digest the information and reflect on the experience and what they were learning. Their only criticism of scientists who made powerpoint presentations was that some of them read those presentations verbatim. Overall, Rockman concluded that the workshops had had a lasting impact on participants.

We've learned that the workshops have a ripple effect. Many radio reporters return to their stations and give their own workshops using some of our handouts. They also produce

stories using the tools and techniques they acquired in the workshops, and a large number keep in touch with each other, sharing editorial advice and support. Most important, at the end of the Science Literacy Training workshops, participants feel confident that they can handle complex scientific stories well.

An assistant news director who took the workshop ten years ago wrote recently, "I feel it has helped immeasurably with my reporting, my interactions with scientists, and my ability to weigh scientific information, studies and reports."

Another participant called the workshop a "transformative experience" while a third said, "I am a better reporter because of what I learned at that workshop…I use what I learned there almost every day."

The Department of Energy's funding has helped make this important work possible. As one applicant wrote, "science pervades every aspect of our lives and there is now a need like never before for people to better understand the world around them. The work of the Science Literacy Project, in bringing together and training radio science journalists from around the country, is an invaluable service."

We thank DOE for having the foresight to support this project.

252

Biological and Environmental Research Information System: A Multifaceted Approach to DOE Systems Biology Research Communication

Anita J. Alton, Shirley H. Andrews, Jennifer L. Bownas, Christopher E. Caldwell, Kris S. Christen, Holly L. Haun, Sheryl A. Martin, Marissa D. Mills, Judy M. Wyrick, and **Betty K. Mansfield*** (mansfieldbk@ornl.gov)

Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tenn.

Project Goals: Provide programmatic information via printed and online materials to help build the critical multidisciplinary community needed to advance systems biology research for DOE energy and environmental missions. The Biological and Environmental Research Information System group works with Genomic Science program managers to help develop and communicate key scientific and technical concepts to the research community and the public. Ideas are welcome to extend and improve communications and program integration and thus represent Genomic Science program research more comprehensively.

Concerted communication is key to progress in cutting-edge science and public accountability. For the past 20 years, our group, the Biological and Environmental Research Information System (formerly Genome Management Information System), has focused on presenting all facets of genomics research for the Department of Energy's (DOE) Office of Science. The materials we produce have helped ensure that scientists can participate in and reap the bounty of the genome revolution, that new generations of students can be trained in genomics and systems biology, and that the public can make informed decisions regarding genetics issues. Our goals focus on three objectives: (1) facilitate Genomic Science planning, research, and communication; (2) respond to outreach and information exchange needs of related DOE projects; and (3) inform a broader audience about DOE genomics research projects, progress, and significance to science and society.

This past year, our scope was extended to all programs within the Office of Biological and Environmental Research (BER), which conducts frontier research in climate, subsurface biogeochemistry, and genome science within the Office of Science. These programs explore scientific complexity at scales requiring contributions from teams of interdisciplinary scientists, thereby necessitating an unprecedented integrative approach both to the science and to science communication strategies. Because each scientific discipline has multiple perspectives and languages, effective communication to help foster achievement and translation of scientific discovery into appropriate DOE mission areas is critical to BER's success. We work with DOE staff and the research community to produce and disseminate information in various formats: technical reports, roadmaps, websites, brochures, databases, technical compilations, presentations, exhibits for scientific meetings, text, graphics, and posters. When appropriate, we also work with DOE grantees and members of the extended DOE BER community, especially the BioEnergy Research Centers and the Joint Genome Institute, to help increase their reach and visibility.

For BER's Biological Systems Science Division (BSSD), our recent Genomic Science program accomplishments include research plans and reports produced with the scientific community:

- *New Frontiers in Characterizing Biological Systems* (October 2009)
- *Systems Biology Knowledgebase for a New Era in Biology* (March 2009)
- *Sustainability of Biofuels: Future Research Opportunities* (March 2009)
- *Genomics:GTL 2008 Strategic Plan* (February 2009)
- *Carbon Cycling and Biosequestration: Integrating Biology and Climate Through Systems Science* (December 2008)

BER BSSD booklets and brochures include:

- *DOE Genomic Science Program Overview* (November 2009)
- *DOE Systems Biology Knowledgebase: Community-Driven Cyber Infrastructure for Sharing and Integrating Data and Analytical Tools* (November 2009)
- *Bioenergy Research Centers: An Overview of the Science* (a revision, July 2009)

- *Executive Summary – Systems Biology Knowledgebase for a New Era in Biology* (April 2009)
- *Report Overview – Carbon Cycling and Biosequestration: Integrating Biology and Climate Through Systems Science* (January 2009)

Other recently produced BER BSSD materials include the abstracts books for the *Fourth Annual U.S. DOE Joint Genome Institute User Meeting* (March 2009) and the *DOE Genomics:GTL Awardee Workshop VII and USDA-DOE Plant Feedstock Genomics for Bioenergy Awardee Workshop* (February 2009), and an exhibit created for the DOE Office of Science Genomic Science Program (October 2009).

BER BSSD works in progress include this abstracts book and several brochures: *Executive Summary – Sustainability of Biofuels: Future Research Opportunities; Overview – DOE BER Biological Systems Science Division;* and *Overview – DOE BER Joint Genome Institute*.

We also continuously update and enhance numerous websites including the Genomic Science website (*genomicscience.energy.gov*) and public image gallery (*genomics.energy.gov/gallery/*). A major redesign of the Genomic Science website is under way. The updated site will streamline content and design, while improving navigation and increasing functionality and accessibility. A Flickr image gallery will be among the new features.

BER-wide projects completed and in progress include the creation, marketing, and dissemination of:

- *Overview – DOE Office of Science Biological and Environmental Research* (July 2009)
- DOE BER poster
- Public BER Research Highlights database (*public.ornl.gov/hgmis/bernews/*)

Related works in progress in BER's Climate and Environmental Sciences Division include:

- *Overview – DOE BER Climate and Environmental Sciences Division*
- *Overview – DOE BER Environmental Molecular Sciences Laboratory*
- *Overview – DOE BER ARM Climate Research Facility*

The Biological and Environmental Research Information System is supported by the U.S. Department of Energy Office of Biological and Environmental Research in the DOE Office of Science.

Appendix 1: Participants

As of January 11, 2010

Michael Adams
University of Georgia, Athens
adams@bmb.uga.edu

Paul Adams
Lawrence Berkeley National Laboratory
pdadams@lbl.gov

Muktak Aklujkar
University of Massachusetts, Amherst
muktak@microbio.umass.edu

Simon Allen
University of California, Los Angeles
simon.allen@ucsf.edu

Eric Alm
Massachusetts Institute of Technology
ejalm@mit.edu

Daniel Amador-Noguez
Princeton University
damador@princeton.edu

Gennady Ananyev
Rutgers University
ananyev@princeton.edu

Gordon Anderson
Pacific Northwest National Laboratory
gordon@pnl.gov

Adam Arkin
Lawrence Berkeley National Laboratory
aparkin@lbl.gov

Scott Baker
Pacific Northwest National Laboratory
scott.baker@pnl.gov

Nitin Baliga
Institute for Systems Biology
nbaliga@systemsbiology.org

Chris Bare
Institute for Systems Biology
bmorrow@systemsbiology.org

David Baumler
University of Wisconsin, Madison
dbaumler@wisc.edu

Qasim Beg
Boston University
qasimbeg@bu.edu

Alexander Beliaev
Pacific Northwest National Laboratory
alex.beliaev@pnl.gov

Harry Beller
Lawrence Berkeley National Laboratory
hrbeller@lbl.gov

Kelly Bender
Southern Illinois University
bender@micro.siu.edu

Rajiv Bharadwaj
Joint BioEnergy Institute
rbharadwaj@lbl.gov

Mark Biggin
Lawrence Berkeley National Laboratory
mdbiggin@lbl.gov

David Bjornstad
Oak Ridge National Laboratory
bjornstaddj@ornl.gov

Paul Blum
University of Nebraska
pblum1@unl.edu

Harvey Bolton, Jr.
Pacific Northwest National Laboratory
harvey.bolton@pnl.gov

Yannick Bomble
National Renewable Energy Laboratory
yannick.bomble@nrel.gov

James Brainard
National Renewable Energy Laboratory
james.brainard@nrel.gov

Ian Brasg
Princeton University
ibrasg@princeton.edu

Mya Breitbart
University of South Florida
mya@marine.usf.edu

Thomas Brettin
Oak Ridge National Laboratory
brettints@ornl.gov

Eoin Brodie
Lawrence Berkeley National Laboratory
elbrodie@lbl.gov

Luke Burow
Stanford University
lburow@stanford.edu

Gareth Butland
Lawrence Berkeley National Laboratory
gpbutland@lbl.gov

William Cannon
Pacific Northwest National Laboratory
bill@pnl.gov

Peijian Cao
University of California, Davis
Joint BioEnergy Institute
pjcao@ucdavis.edu

Kevin Carpenter
Lawrence Livermore National Laboratory
carpenter37@llnl.gov

Michael Casler
Agricultural Research Service
U.S. Department of Agriculture
Great Lakes Bioenergy Research Center
mdcasler@wisc.edu

Jamie Cate
Lawrence Berkeley National Laboratory
jcate@lbl.gov

Romy Chakraborty
Lawrence Berkeley National Laboratory
rchakraborty@lbl.gov

John-Marc Chandonia
Lawrence Berkeley National Laboratory
jmchandonia@lbl.gov

Aarthi Chandrasekaran
Joint BioEnergy Institute
achandrasekaran@lbl.gov

Roger Chang
University of California, San Diego
rlchang@ucsd.edu

Christopher Chang
National Renewable Energy Laboratory
christopher.chang@nrel.gov

Yihong Chen
University of Maryland Biotechnology Institute
yihongchen2004@yahoo.com

Yi-Ru Chen
University of Minnesota
chenx137@umn.edu

Swapnil Chhabra
Lawrence Berkeley National Laboratory
Joint BioEnergy Institute
srchhabra@lbl.gov

Dylan Chivian
Lawrence Berkeley National Laboratory
Joint BioEnergy Institute
dcchivian@lbl.gov

Byung K. Cho
University of California, San Diego
bcho@ucsd.edu

Katy Christiansen
Joint BioEnergy Institute
kmchristiansen@lbl.gov

Liqiang Chu
University of Notre Dame
chuliqiang@nd.edu

George Church
Harvard Medical School
gmc@harvard.edu

Dean Cole
U.S. Department of Energy
dean.cole@science.doe.gov

James Cole
Michigan State University
colej@msu.edu

Frank Collart
Argonne National Laboratory
fcollart@anl.gov

Robert Cottingham
Oak Ridge National Laboratory
cottinghamrw@ornl.gov

Michael Crowley
National Renewable Energy Laboratory
michael.crowley@nrel.gov

Alan Darvill
Bionergy Science Center
University of Georgia, Athens
adarvill@ccrc.uga.edu

Mark Davis
National Renewable Energy Laboratory
mark.davis@nrel.gov

Brian Davison
Oak Ridge National Laboratory
davisonbh@ornl.gov

Valerie de Crecy-Lagard
University of Florida
vcrecy@ufl.edu

Kristen DeAngelis
Lawrence Berkeley National Laboratory
kdeangelis@lbl.gov

Robert DeBoy
J. Craig Venter Institute
rdeboy@jcvi.org

Steve Decker
National Renewable Energy Laboratory
steve.decker@nrel.gov

Paramvir Dehal
Lawrence Berkeley National Laboratory
psdehal@lbl.gov

Adam Deutschbauer
Lawrence Berkeley National Laboratory
amdeutschbauer@lbl.gov

Patrik D'haeseleer
Lawrence Livermore National Laboratory
Joint BioEnergy Institute
patrikd@llnl.gov

Valentina Di Francesco
National Institute of Allergies and Infectious Diseases
National Institutes of Health
vdifrancesco@niaid.nih.gov

Brian Dill
Oak Ridge National Laboratory
dillbd@ornl.gov

Shi-You Ding
National Renewable Energy Laboratory
shi.you.ding@nrel.gov

Mitchel Doktycz
Oak Ridge National Laboratory
doktyczmj@ornl.gov

Bryon Donohoe
National Renewable Energy Laboratory
bryon.donohoe@nrel.gov

Timothy Donohue
Great Lakes Bioenergy Research Center
tdonohue@bact.wisc.edu

Kenneth Downing
Lawrence Berkeley National Laboratory
khdowning@lbl.gov

Daniel Drell
U.S. Department of Energy
daniel.drell@science.doe.gov

Inna Dubchak
Lawrence Berkeley National Laboratory
ildubchak@lbl.gov

Andrew Dykstra
Oak Ridge National Laboratory
dykstra@ion.chem.utk.edu

Thomas Earnest
Lawrence Berkeley National Laboratory
tnearnest@lbl.gov

Amanda (Nicole) Edwards
Oak Ridge National Laboratory
edwardsan@ornl.gov

Stephanie Eichorst
Los Alamos National Laboratory
seichorst@lanl.gov

Dwayne Elias
Oak Ridge National Laboratory
eliasda@ornl.gov

James Elkins
Oak Ridge National Laboratory
elkinsjg@ornl.gov

Barbara Evans
Oak Ridge National Laboratory
evansb@ornl.gov

Xiaojiang Feng
Princeton University
xfeng@princeton.edu

Matthew Fields
Montana State University
matthew.fields@erc.montana.edu

Curt Fischer
University of California, Berkeley
tentrillion@gmail.com

Madeline Fisher
University of Wisconsin, Madison
mmfisher@wisc.edu

Ronan Fleming
University of Iceland
ronan.mt.fleming@gmail.com

Virginia Fowler
Oak Ridge National Laboratory
fowlervs@ornl.gov

Brian Fox
Great Lakes Bioenergy Research Center
bgfox@biochem.wisc.edu

Jim Fredrickson
Pacific Northwest National Laboratory
jim.fredrickson@pnl.gov

Andrew Frock
North Carolina State University
adfrock@ncsu.edu

Katrin Gaertner
Great Lakes Bioenergy Research Center
kgaertn@msu.edu

Lila Ghamsari
Dana-Farber Cancer Institute
lila_ghamsari@dfci.harvard.edu

Maria Ghirardi
National Renewable Energy Laboratory
maria.ghirardi@nrel.gov

Richard Giannone
Oak Ridge National Laboratory
giannonerj@ornl.gov

Carol Giometti
Argonne National Laboratory
csgiometti@anl.gov

John Gladden
Joint BioEnergy Institute
jmgladden@lbl.gov

John Glass
J. Craig Venter Institute
jglass@jcvi.org

Adam Godzik
Burnham Institute for Medical Research
adam@burnham.org

Peter Good
National Human Genome Research Institute
goodp@mail.nih.gov

Andrey Gorin
Oak Ridge National Laboratory
agor@ornl.gov

Amita Gorur
Lawrence Berkeley National Laboratory
agorur@lbl.gov

Joseph Graber
U.S. Department of Energy
joseph.graber@science.doe.gov

Celestine Grady-Smith
University of California, Davis
csgrady@ucdavis.edu

Susan Gregurick
U.S. Department of Energy
susan.gregurick@science.doe.gov

Sol Gruner
Cornell University
smg26@cornell.edu

Lyn Gulley
Arizona State University
lyn.gulley@asu.edu

Robert Gunsalus
University of California, Los Angeles
robg@microbio.ucla.edu

Murray Hackett
University of Washington
mhackett@u.washington.edu

Carolyn Hall
Lawrence Livermore National Laboratory
hall94@llnl.gov

Bong-Gyoon Han
Lawrence Berkeley National Laboratory
bghan@lbl.gov

Loren Hauser
Oak Ridge National Laboratory
hauserlj@ornl.gov

Terry Hazen
Lawrence Berkeley National Laboratory
tchazen@lbl.gov

Terry Hazen
Lawrence Berkeley National Laboratory
tchazen@lbl.gov

Eric Hegg
Michigan State University
erichegg@msu.edu

Michael Held
Michigan State University
Great Lakes Bioenergy Research Center
heldm@msu.edu

Alyssa Henning
Cornell University
alh225@cornell.edu

Bernard Henrissat
Glycogenomics Architecture et Fonction des Macromolécules Biologiques
bernard.henrissat@afmb.univ-mrs.fr

Robert Hettich
Oak Ridge National Laboratory
hettichrl@ornl.gov

Nathan Hillson
Lawrence Berkeley National Laboratory
Joint BioEnergy Institute
njhillson@lbl.gov

Mike Himmel
National Renewable Energy Laboratory
mike.himmel@nrel.gov

Roland Hirsch
U.S. Department of Energy
roland.hirsch@science.doe.gov

Hoi-Ying Holman
Lawrence Berkeley National Laboratory
hyholman@lbl.gov

John Houghton
U.S. Department of Energy
john.houghton@science.doe.gov

Greg Hurst
Oak Ridge National Laboratory
hurstgb@ornl.gov

Yongqin Jiao
Lawrence Livermore National Laboratory
jiao1@llnl.gov

Jian Jin
Lawrence Berkeley National Laboratory
jjin@lbl.gov

Udaya Kalluri
Oak Ridge National Laboratory
kalluriudayc@ornl.gov

Marina Kalyuzhnaya
University of Washington, Seattle
mkalyuzh@u.washington.edu

Tatiana Karpinets
Oak Ridge National Laboratory
karpinetstv@ornl.gov

Abhijit Karve
Oak Ridge National Laboratory
karveaa@ornl.gov

Arthur Katz
U.S. Department of Energy
arthur.katz@science.doe.gov

Jay Keasling
Lawrence Berkeley National Laboratory
Joint BioEnergy Institute
keasling@lbl.gov

David Keating
Great Lakes Bioenergy Research Center
dkeating@glbrc.wisc.edu

Kenneth Keegstra
Michigan State University
keegstra@msu.edu

Kimberly Keller
University of Missouri
kellerkl@missouri.edu

Martin Keller
Oak Ridge National Laboratory
kellerm@ornl.gov

Robert Kelly
North Carolina State University
rmkelly@eos.ncsu.edu

Zvi Kelman
University of Maryland Biotechnology Institute
kelman@umbi.umd.edu

Madhu Khanna
University of Illinois, Urbana Champaign
khanna1@illinois.edu

George Khoury
Pennsylvania State University
gak5001@psu.edu

Jane Khudyakov
Joint BioEnergy Institute
jikhudyakov@lbl.gov

Joonhoon Kim
University of Wisconsin, Madison
jkim92@wisc.edu

Daniel Klein-Marcuschamer
Joint BioEnergy Institute
dklein@lbl.gov

William Klimke
National Center for Biotechnology Information
klimke@ncbi.nlm.nih.gov

Niels Klitgord
Boston University Bioinformatics Program
niels@bu.edu

Bernhard Knierim
Lawrence Berkeley National Laboratory
bknierim@lbl.gov

Lori Knowles
University of Alberta
loripknowles@gmail.com

Wayne Kontur
Great Lakes Bioenergy Research Center
wskontur@wisc.edu

David Koppenaal
Pacific Northwest National Laboratory
david.koppenaal@pnl.gov

Julia Krushkal
University of Tennessee Health Science Center
jkrushka@uthsc.edu

Cheryl Kuske
Los Alamos National Laboratory
kuske@lanl.gov

Rick Laguna
Ohio State University
rlaguna22@gmail.com

Andrew Lancaster
University of Georgia, Athens
alancaster@bmb.uga.edu

Miriam Land
Oak Ridge National Laboratory
landml@ornl.gov

Robert Landick
University of Wisconsin, Madison
Great Lakes Bioenergy Research Center
landick@bact.wisc.edu

Paul Langan
Los Alamos National Laboratory
langan_paul@lanl.gov

Dan Lauffer
Great Lakes Bioenergy Research Center
dlauffer@wisc.edu

Ching Leang
University of Massachusetts, Amherst
leang@microbio.umass.edu

Daehee Lee
University of California, San Diego
dlee@bioeng.ucsd.edu

Taek Soon Lee
Joint BioEnergy Institute
tslee@lbl.gov

John Leigh
University of Washington
leighj@u.washington.edu

Jeffrey Lewis
Great Lakes Bioenergy Research Center
jalewis4@wisc.edu

Zhuo Li
University of Maryland Biotechnology Institute
lizhuo@umbi.umd.edu

Jan Liphardt
Lawrence Berkeley National Laboratory
liphardt@berkeley.edu

Chang-Jun Liu
Brookhaven National Laboratory
cliu@bnl.gov

Derek Lovley
University of Massachusetts
dlovley@microbio.umass.edu

Lee Lynd
Dartmouth College
lee.r.lynd@dartmouth.edu

Yukari Maezato
University of Nebraska
yukarimaezato@hotmail.com

Radhakrishnan Mahadevan
University of Toronto
krishna.mahadevan@utoronto.ca

Nikhil Malvankar
University of Massachusetts, Amherst
nikhil@physics.umass.edu

Yuzuki Manabe
Joint BioEnergy Institute
ymanabe@lbl.gov

Kalyaneswar Mandal
University of Chicago
kmandal@uchicago.edu

David Mann
University of Tennessee
dmann1@utk.edu

Costas Maranas
Pennsylvania State University
costas@psu.edu

Gary Marchant
Arizona State University
gary.marchant@asu.edu

Betty Mansfield
Oak Ridge National Laboratory
mansfieldbk@ornl.gov

Koushik Mazumder
University of Georgia, Athens
koushik@ccrc.uga.edu

Lee Ann McCue
Pacific Northwest National Laboratory
leeann.mccue@pnl.gov

James McKinlay
University of Washington
mckinla1@uw.edu

David Mead
Lucigen
Great Lakes Bioenergy Research Center
dmead@lucigen.com

John (Jack) Meeks
University of California, Davis
jcmeeks@ucdavis.edu

Matthew Melnicki
Pacific Northwest National Laboratory
matthew.melnicki@pnl.gov

Barbara Methe
J. Craig Venter Institute
bmethe@jcvi.org

Folker Meyer
Argonne National Laboratory
mishur@mcs.anl.gov

Jonathan Millen
University of Rochester
jonathan_millen@hotmail.com

Marissa Mills
Oak Ridge National Laboratory
millsmd@ornl.gov

Debra Mohnen
University of Georgia, Athens
dmohnen@ccrc.uga.edu

Jennifer Morrell-Falvey
Oak Ridge National Laboratory
morrelljl1@ornl.gov

Aindrila Mukhopadhyay
Lawrence Berkeley National Laboratory
amukhopadhyay@lbl.gov

Harish Nagarajan
University of California, San Diego
nh@ucsd.edu

Kelly Nevin
University of Massachusetts, Amherst
knevin@microbio.umass.edu

Basil Nikolau
Iowa State University
dimmas@iastate.edu

Kenneth Noll
University of Connecticut
kenneth.noll@uconn.edu

Jaspreet Notey
North Carolina State University
jsnotey@ncsu.edu

Yasuhiro Oda
University of Washington
yasuhiro@u.washington.edu

Daniel Olson
Dartmouth College, Mascoma
dan268@gmail.com

John Orban
University of Maryland
orban@umbi.umd.edu

Andrei Osterman
Burnham Institute
osterman@burnham.org

Elizabeth Ottesen
Massachusetts Institute of Technology
ottesen@mit.edu

Anna Palmisano
Office of Biological and Environmental Research
U.S. Department of Energy
anna.palmisano@science.doe.gov

Anthony Palumbo
Oak Ridge National Laboratory
palumboav@ornl.gov

Morey Parang
Oak Ridge National Laboratory
parangm@ornl.gov

Joshua Park
Joint BioEnergy Institute
jipark@lbl.gov

Ljiljana Pasa-Tolic
Pacific Northwest National Laboratory
ljiljana.pasatolic@pnl.gov

Sivakumar Pattathil
University of Georgia, Athens
siva@ccrc.uga.edu

Dale Pelletier
Oak Ridge National Laboratory
pelletierda@ornl.gov

Maria Pena
University of Georgia, Athens
BioEnergy Science Center
mpena@ccrc.uga.edu

Gary Peter
University of Florida
gfpeter@ufl.edu

Loukas Petridis
Oak Ridge National Laboratory
petridisl@ornl.gov

Amanda Petrus
University of Connecticut
amanda.petrus@uconn.edu

Jennifer Pett-Ridge
Lawrence Livermore National Laboratory
pettridge2@llnl.gov

Chris Petzold
Lawrence Berkeley National Laboratory
cjpetzold@lbl.gov

Grigoriy Pinchuk
Pacific Northwest National Laboratory
grigoriy.pinchuk@pnl.gov

Sai Venkatesh Pingali
Oak Ridge National Laboratory
pingalis@ornl.gov

David Pletcher
Joint BioEnergy Institute
dcpletcher@lbl.gov

Malu Polanski
National Institutes of Health
polanskim@niaid.nih.gov

Iris Porat
Oak Ridge National Laboratory
porati@ornl.gov

Yu Qiu
University of California, San Diego
yuqiu@ucsd.edu

Sridhar Ranganathan
Pennsylvania State University
sur152@psu.edu

Prabhasa Ravikirthi
Pennsylvania State University
pxr173@psu.edu

Jennifer Reed
University of Wisconsin, Madison
reed@engr.wisc.edu

David Reiss
Institute for Systems Biology
dreiss@systemsbiology.org

Karin Remington
General Medical Sciences
National Institutes of Health
remingka@nigms.nih.gov

Scott Retterer
Oak Ridge National Laboratory
rettererst@ornl.gov

William Riehl
Boston University
briehl@bu.edu

Ashley Robinson
Pacific Northwest National Laboratory
ashley.robinson@pnl.gov

Robby Robinson
Pacific Northwest National Laboratory
errol.robinson@pnl.gov

Sebastien Rodrigue
Massachusetts Institute of Technology
s_rod@mit.edu

Margaret Romine
Pacific Northwest National Laboratory
margie.romine@pnl.gov

Cathy Ronning
U.S. Department of Energy
cathy.ronning@science.doe.gov

Douglas Rusch
J. Craig Venter Institute
drusch@jcvi.org

Brian Saar
Harvard University
saar@fas.harvard.edu

Rajib Saha
Pennsylvania State University
rus184@psu.edu

Kourosh Salehi-Ashtiani
Dana-Farber Cancer Institute
kourosh_salehi-sshtiani@dfci.harvard.edu

Simo Sarkanen
University of Minnesota
sarka001@umn.edu

Jan Schellenberger
University of California, San Diego
jschelle@ucsd.edu

Henrik Scheller
Lawrence Berkeley National Laboratory
Joint BioEnergy Institute
hscheller@lbl.gov

Marianne Schiffer
Argonne National Laboratory
mschiffer@anl.gov

Gerrit Schut
University of Georgia, Athens
gerti@uga.edu

Michael Schwalbach
Great Lakes Bioenergy Research Center
mschwalbach@glbrc.wisc.edu

Robert Scott
University of Georgia, Athens
rscott@uga.edu

John Sedbrook
Great Lakes Bioenergy Research Center
jcsedbr@ilstu.edu

Daniel Segre
Boston University
dsegre@bu.edu

Michael Seibert
National Renewable Energy Laboratory
mike.seibert@nrel.gov

Pamela Seidenman
Joint BioEnergy Institute
psseidenman@lbl.gov

Margrethe Serres
Marine Biological Laboratory
mserres@mbl.edu

Adrian Sharma
Massachusetts Institute of Technology
aksharma@mit.edu

Rita Sharma
University of California, Davis
rasharma@ucdavis.edu

Xiling Shen
Cornell University
xilings@ece.cornell.edu

Louis Sherman
Purdue University
lsherman@purdue.edu

Blake Simmons
Joint BioEnergy Institute
basimmo@sandia.gov

Steven Singer
Lawrence Berkeley National Laboratory
swsinger@lbl.gov

Nicholas Skizim
Rutgers University
nskizim@princeton.edu

Steven Slater
Great Lakes Bioenergy Research Center
scslater@glbrc.wisc.edu

Richard Smith
Pacific Northwest National Laboratory
rds@pnl.gov

Melanie Spero
University of Wisconsin
spero@wisc.edu

Prem Srivastava
U.S. Department of Energy
prem.srivastava@science.doe.gov

Marvin Stodolsky
U.S. Department of Energy
marvin.stodolsky@science.doe.gov

Sergey Stolyar
Pacific Northwest National Laboratory
sergey.stolyar@pnl.gov

Garret Suen
Great Lakes Bioenergy Research Center
gsuen@wisc.edu

Zarath Summers
University of Massachusetts, Amherst
zsummers@microbio.umass.edu

Jun Sun
Genomatica, Inc.
jsun@genomatica.com

Christine Sun
University of California, Berkeley
christine_sun@berkeley.edu

Lan Sun
Joint BioEnergy Institute
lansun@lbl.gov

Anil Suresh
Oak Ridge National Laboratory
sureshak@ornl.gov

Patrick Suthers
Pennsylvania State University
suthers@engr.psu.edu

Anongpat Suttangkakul
Joint BioEnergy Institute
asuttangkakul@lbl.gov

Granger Sutton
J. Craig Venter Institute
gsutton@jcvi.org

Jonathan Sweedler
University of Illinois
jsweedle@illinois.edu

F. Robert Tabita
Ohio State University
tabita.1@osu.edu

Larry Taylor
National Renewable Energy Laboratory
larry.taylor@nrel.gov

Michael Teresinski
U.S. Department of Energy
michael.teresinski@science.doe.gov

Michael Thelen
Lawrence Livermore National Laboratory
Joint BioEnergy Institute
mthelen@llnl.gov

Ines Thiele
University of Iceland
ines.thiele@gmail.com

Michael Thorgersen
University of Georgia, Athens
mthorger@uga.edu

Andrew Tolonen
Harvard Medical School
tolonen@genetics.med.harvard.edu

Kimberly Tucker
University of South Florida
kpause@marine.usf.edu

Edward Uberbacher
Oak Ridge National Laboratory
uberbacherec@ornl.gov

Volker Urban
Oak Ridge National Laboratory
urbanvs@ornl.gov

Breeanna Urbanowicz
University of Georgia, Athens
BioEnergy Science Center
breeanna@uga.edu

Patanjali Varanasi
Joint BioEnergy Institute
pvaranasi@lbl.gov

Miguel Vega-Sanchez
University of California, Davis
Joint BioEnergy Institute
mevega-sanchez@lbl.gov

Nathan VerBerkmoes
Oak Ridge National Laboratory
verberkmoesn@ornl.gov

Peter Walian
Lawrence Berkeley National Laboratory
pjwalian@lbl.gov

Judy Wall
University of Missouri, Columbia
wallj@missouri.edu

Zhiwu Wang
drwangzhiwu@gmail.com

Sharlene Weatherwax
U.S. Department of Energy
sharlene.weatherwax@science.doe.gov

Peter Weber
Lawrence Livermore National Laboratory
weber21@llnl.gov

Zhiping Weng
University of Massachusetts Medical School
zhiping.weng@umassmed.edu

Gary Wesenberg
University of Wisconsin, Madison
gewesenb@wisc.edu

Janet Westpheling
University of Georgia, Athens
janwest@uga.edu

Elizabeth White
U.S. Department of Energy
elizabeth.white@science.doe.gov

Derrick White
University of Nebraska
derrick.white@yahoo.com

William Whitman
University of Georgia, Athens
whitman@uga.edu

Steven Wiley
Pacific Northwest National Laboratory
steven.wiley@pnl.gov

Curtis Wilkerson
Michigan State University
wilker13@msu.edu

Halina Ewa Witkowska
University of California, San Francisco
witkowsk@cgl.ucsf.edu

Dagmar Woebken
Stanford University
dwoebken@stanford.edu

Amy Wolfe
Oak Ridge National Laboratory
wolfeak@ornl.gov

J. H. David Wu
University of Rochester
davidwu@che.rochester.edu

Ying Xu
University of Georgia, Athens
xyn@bmb.uga.edu

Qi Xu
National Renewable Energy Laboratory
qi.xu@nrel.gov

Tae Hoon Yang
Genomatica, Inc.
tyang@genomatica.com

Lee Yang
Lawarence Berkeley National
Laboratory
llyang@lbl.gov

Yanbin Yin
University of Georgia, Athens
yinyb@csbl.bmb.uga.edu

Sung Ho Yoon
Institute for Systems Biology
syoon@systemsbiology.org

Karsten Zengler
University of California, San Diego
kzengler@ucsd.edu

Y.-H. Percival Zhang
Virginia Tech
biofuels@vt.edu

Alirezq Zomorrodi
Pennsylvania State Univeristy
alireza@psu.edu

Appendix 2: Websites

Genomic Science Program Websites

- Genomic Science program http://genomicscience.energy.gov
- Office of Biological and Environmental Research Genomic Science http://www.sc.doe.gov/ober/BSSD/genomics.html
- This book http://genomicscience.energy.gov/pubs/2010abstracts/
- Plant Feedstock Genomics Research http://genomicscience.energy.gov/research/DOEUSDA/
- Publications
  - *New Frontiers in Characterizing Biological Systems* (October 2009) http://genomicscience.energy.gov/characterization/
  - *Systems Biology Knowledgebase for a New Era in Biology* (March 2009) http://genomicscience.energy.gov/compbio/
  - *Sustainability of Biofuels: Future Research Opportunities* (March 2009) http://genomicscience.energy.gov/biofuels/sustainability/
  - *Genomics:GTL 2008 Strategic Plan* (February 2009) http://genomicscience.energy.gov/strategicplan/
  - *Carbon Cycling and Biosequestration: Integrating Biology and Climate Through Systems Science* (December 2008) http://genomicscience.energy.gov/carboncycle/
  - *Breaking the Biological Barriers to Cellulosic Ethanol: A Joint Research Agenda* (June 2006) http://genomicscience.energy.gov/biofuels/b2bworkshop.shtml

Bioenergy Research Center Websites

- Overview http://genomicscience.energy.gov/centers/
- BioEnergy Science Center http://www.bioenergycenter.org
- Great Lakes Bioenergy Research Center http://www.greatlakesbioenergy.org
- Joint BioEnergy Institute http://www.jbei.org

Project and Related Websites

- DOE Joint Genome Institute http://jgi.doe.gov
- Environmental Molecular Sciences Laboratory http://www.emsl.pnl.gov/emslweb/
- MicrobesOnline http://microbesonline.org
- Virtual Institute for Microbial Stress and Survival http://vimss.lbl.gov
- MAGGIE Data Viewer (BETA) http://maggie.systemsbiology.net
- CAZy toolkit and environment, part of BESC KnowledgeBase (BETA) http://bobcat.ornl.gov/besc/

Author Index

Indexed by Page Number

C

D

E

F

G

H

I

J

K

L

M

N

O

P

Q

R

S

T

U

V

W

X

Y

Z

Institutional Index

Indexed by Page Number

Institutions determined by affiliations provided by authors.