# Speaker Abstracts

## U.S. Department of Energy
## Genomic Science 2010

## Awardee Workshop VIII
## and
## Knowledgebase Workshop

**February 7–10, 2010**
**Hyatt Regency Crystal City**

U.S. DEPARTMENT OF
ENERGY | Office of Science

# Table of Contents

## Monday Plenaries



## Monday Breakout Sessions

### *A: Genomic Structure and Epigenetics*

**Location: Roosevelt/Lincoln**



### *B: Integrated Omics Approaches to Understand Environmental Processes*

**Location: Jefferson/Kennedy**



### *C: Imaging and Technology for Systems Biology*

**Location: General Session (Regency Ballroom E and F)**

## Tuesday Plenaries



## Tuesday Breakout Sessions

### *D: Systems Biology Knowledgebase Workshop for Genomic Sciences Applications*

**Location: General Session (Regency Ballroom E and F)**



### *E: Synthetic Biology*

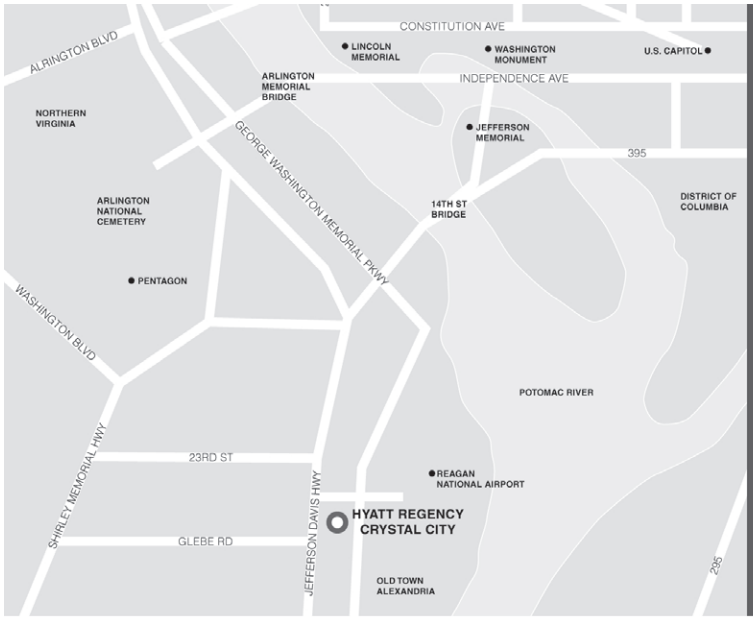
**Location: Roosevelt/Lincoln**

### *F: Structural Biology Applications in Genomic Sciences*

**Location: Jefferson/Kennedy**



## Wednesday Plenaries

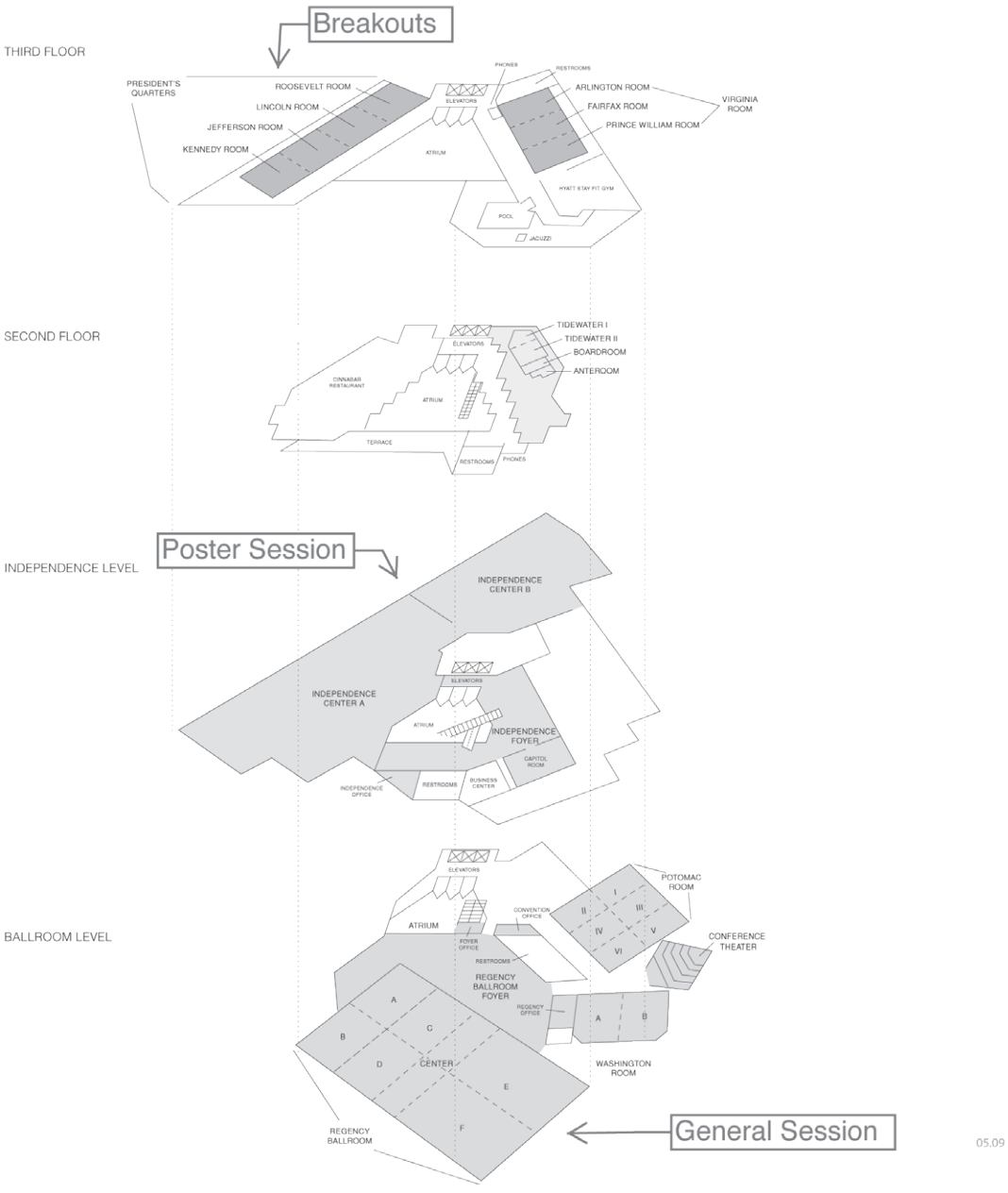



# Hyatt Regency Crystal City

## 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.



**Monday, February 8, 9:00 a.m.–9:30 a.m.**
**Plenary Session: DOE Bioenergy Research Centers**

## Improving Bioenergy Microbes by Multidimensional Evolutionary and Functional Genomics

**Robert Landick**,[1,2,3] David Keating,[1] Jeffrey A. Lewis,[1,4] Dana J. Wohlbach,[1,4] Michael Schwalbach,[1,5] Mary Tremaine,[1] Frances Tran,[1,2] Trey Sato,[1] Patricia Kiley,[1,5] and Audrey Gasch[1,4,6]

[1]Great Lakes Bioenergy Research Center; Departments of [2]Biochemistry, [3]Bacteriology, [4]Genetics, and [5]Biomolecular Chemistry; [6]Genome Center of Wisconsin; University of Wisconsin-Madison

A central goal of the Great Lakes Bioenergy Research Center is to combine information from multiple, cellular, and genome-scale "omic" analyses to define bottlenecks in microbial biofuel production, and to identify genes whose expression or altered regulation can overcome these bottlenecks. Toward that goal, analysis pipelines for genomics, transcriptomics, ChIPSeq, proteomics, metabolomics, and lipidomics have been established in collaboration with the DOE Joint Genome Institute, Pacific Northwest National Laboratory, and the University of Washington Biotechnology Center. The application of these methods to two central problems in microbial biofuels, ethanol tolerance in yeast and bacteria and xylose utilization in yeast, will be described. Through the use of novel comparative and functional genomic analyses of microbial species related by natural or directed evolution, we have identified core sets of genes involved in ethanol tolerance and xylose utilization. The core ethanol-tolerance genes include unexpected genes that suggest novel targets for strain engineering. Similarly, the core xylose utilization genes encode some functions not previously associated with xylose utilization.

**Monday, February 8, 9:30 a.m.–10:00 a.m.**
**Plenary Session: DOE Bioenergy Research Centers**

## High-Throughput Multiplexed GlycoChip Enzymatic Assays for Biofuels Development

**Trent Northen**, Wolfgang Reindl, Kai Deng, John Gladden, Steven Singer, Anup Singh, Blake Simmons, Paul Adams, and Jay Keasling

Technology Division, Joint BioEnergy Institute

Critical to the development of lignocellulosic biofuel processes is the efficient deconstruction of this highly recalcitrant material into fermentable sugars. At the Joint BioEnergy Institute (JBEI) this is being approached by optimizing both the feedstock and the deconstruction processes. In both cases, new technologies are required for identification of unknown activities and high-throughput screening of "targeted" reactions (e.g., glycosyl hydrolases). GlycoChips are constructed using perfluoronated tags to immobilized glycans on NIMS mass spectrometry chips. By changing the mass of the linker, multiple substrates can be assayed simultaneously. This has been used to characterize beta-glucosidase and beta-xylanase activities including ionic liquid tolerance, optimal pH, and temperature directly from isolates and microbial communities. We are currently automating this assay to screen JBEI enzyme libraries and environmental samples. Together these GlycoChip assays will help identify and optimize conversion of lignocellulose into fermentable sugars for biofuel production.

**Monday, February 8, 10:30 a.m.–11:00 a.m.**
**Plenary Session: DOE Bioenergy Research Centers**

# Consolidated Bioprocessing: From Science to Impact

**Lee Lynd**,[1,9] Mike Adams,[2,9] Steve Brown,[3,9] Brian Davison,[4,9] Jim Elkins,[3,9] Larry Feinberg,[4,9] Jim Flatt,[4,9] Adam Guss,[1,9] Bob Hettich,[3,9] Michael Himmel,[4,9] David Hogsett,[5,9] Javier Izqueirdo,[1,9] Bob Kelly,[6,9] Joel Kreps,[7,9] Jonathan Mielenz,[3,9] Vineet Rajgarhia,[4,9] Babu Raman,[3,9] Xiongjun Shao,[1,9] Daniel van der Lelie,[8,9] Kevin Wenger,[4,9] and Jan Westpheling[2,9]

[1]Dartmouth College, Hanover, New Hampshire; [2]University of Georgia, Athens, Georgia; [3]Oak Ridge National Laboratory; [4]National Renewable Energy Laboratory; [5]Mascoma Corporation, Lebanon, New Hampshire; [6]North Carolina State University; [7]Verenium Corporation; [8]Brookhaven National Laboratory; [9]BioEnergy Science Center, Oak Ridge National Laboratory

Microbial conversion of cellulosic biomass to ethanol or other products without added saccharolytic enzymes—consolidated bioprocessing (CBP)—is widely considered to be the ultimate low-cost configuration for cellulose hydrolysis and fermentation. This presentation will address CBP in terms of underlying fundamentals, technological progress, and applications of genomic science.

Microbially mediated cellulose solubilization involves important and widely under-studied fundamental phenoma above and beyond the also important fundamentals associated with enzmatically mediated solubilization. Fundamental issues distinctive to CBP will be outlined, examples considered, and needed methodological development briefly addressed.

Development of CBP microbes can be pursued by two strategies: improving ethanol production by cellulose-utilizing microbes or improving cellulose-utilization in microbes that produce biofuels well. Technological progress, including proof of concept of CBP, will be presented for both strategies. In particular, improved ethanol production in thermophilic microbes and heterologous expression of a functional cellulase enzyme system in yeast will be considered.

Cellulose solubilization can be considered at successively increasing levels of aggregation—protein subcomponents, multiprotein systems and complexes, pure cultures of cellulolytic microbes, and mixed cellulolytic microbial consortia. Genomic science can be applied beneficially at each of these levels. Examples of such applications will be considered, including studies that have been completed, are under way, and potentially could be initiated in the future.

**Monday, February 8, 11:00 a.m.–12:00 noon**
**Keynote Presentation**

## The Systems Biology of Metabolism

**Bernhard Ø. Palsson**

Department of Bioengineering, University of California, San Diego

Metabolic systems biology has developed and matured over the past 10 years. Currently dozens of curated models exist for many target microorganisms generated by multiple laboratories. The concepts and workflows underlying the reconstruction process have been delineated, and recently a 96-step SOP for their generation has been developed. Automation of this process is also being developed. Similar efforts are being undertaken for genome-scale transcriptional networks, the transcription-translation machinery, and two-component signaling, all of which are driven by new and improved "omics" data types. Comprehensive network reconstructions for the core functions of prokaryotic organisms are thus emerging. A curated and validated network reconstruction effectively represents a knowledgebase. Network models represent query tools of this knowledgebase and form the basis for building models for a multitude of prospective uses. This talk will give an overview of the foundational concepts of this field, summarize its current state, and try to forecast the future challenges it faces.

**Monday, February 8, 2:00 p.m.–2: 25 p.m.**
**Breakout Session A: Genomic Structure and Epigenetics**

## Developmental Regulation and Individual Differences of Neuronal Epigenomes in the Prefrontal Cortex

**Dr. Zhiping Weng**
Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School

Little is known about the regulation of neuronal and other cell-type specific epigenomes from brain. Here, we map the genome-wide distribution of trimethylated histone H3K4 (H3K4me3), a histone mark associated with actual and potentiated transcription, in neuronal nuclei collected from prefrontal cortex (PFC) of nine individuals across a wide age range (0.5 to 69 years). Massively parallel sequencing identified 16,284 to 22,195 H3K4me3-enriched regions (peaks), the majority located proximal to (within 2 kb of) the transcription start site (TSS) of annotated genes. These included signatures specific to neurons as well as signatures specific to individual subjects. Our results reveal age-correlated genome reorganization in the postnatal PFC, including a general increase in TSS-associated H3K4me3 peaks concomitant with loss of peaks at many developmentally regulated genes and a three-fold decline in the proportion of individual-specific peaks. Mapping entire epigenomes from defined cell populations of the human brain can provide insights into novel mechanisms governing normal and diseased neurodevelopment.

**Monday, February 8, 2:25 p.m.–2:50 p.m.**
**Breakout Session A: Genomic Structure and Epigenetics**

## Genome-Wide Mapping and Integration of Chromatin Structure, Transcription Factors, and Transcriptome Output

**Barbara Wold**
Division of Biology and Beckman Institute, Pasadena, California

The recent development of high-throughput sequence-based assays (e.g., RNA-Seq and ChIP-Seq) has permitted genome-wide mapping of chromatin marks, sequence-specific transcription factor binding, RNA polymerase loading, and transcript output in large genomes of animals and plants. The challenges now are to integrate these diverse maps and relate them to different cell states. Major observations concern the relationship of specific regulatory complexes with the expression behavior and polymerase loading of neighboring genes. This talk will discuss strengths and weaknesses in these maps and progress toward integration and inference by machine-learning methods.

**Monday, February 8, 2:50 p.m.–3:15 p.m.**
**Breakout Session A: Genomic Structure and Epigenetics**

# Genboree Open Hosting System for Collaborative Genomic and Epigenomic Research

A. R. Jackson,[1] M.L. Gonzalez-Garay,[1] R. A. Harris,[1] C. Coarfa,[1] T. Charnecki,[1] S. Paithankar,[1] M. Smith,[2] S. G. Davis,[2] J. R. Easton-Marks,[1] S. Raghuraman,[1] and **A. Milosavljevic**[1]

[1]Bioinformatics Research Laboratory; Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas; [2]Bio::Neos Inc., Iowa City, Iowa

Comparative analysis of human epigenomes provides insights into epigenome variation associated with development, gene regulation, and human disease. Analysis of epigenomes of plants and other organisms has the potential to improve the search for efficient biofuel production methods and other industrial biotechnology applications. To address the computational challenges associated with comparative epigenome analyses in the course of the National Institutes of Health's Epigenomics Roadmap Initiative, we have developed the Genboree system. The Initiative aims to construct reference epigenomes, including methylomes, histone marks, small RNA species, and chromatin accessibility maps from more than 100 well-characterized human primary tissues, primary cells, and cell lines. The construction of the epigenomes and their comparative analyses require an informatic architecture that is scalable to accommodate rapidly increasing throughputs of massively parallel sequencing technologies and yet flexible to accommodate an increasing diversity of assays and associated analysis tools and pipelines. Using the software-as-a-service (Web 2.0) model, Genboree provides web-based hosting of data, analysis tools, and analysis results to support collaborative genomic and epigenomic research. In collaboration with the National Center for Biotechnology Information (NCBI) and based on standards emerging from several large-scale projects including TCGA, Encode, and the 1000 genomes project, data models for epigenomic data and metadata have been developed to facilitate multi-directional data flow between the Reference Epigenome Mapping Centers, the Epigenomics Data Analysis and Coordination Center (of which Genboree is the central component), the NCBI, and data users, thus creating an epigenomic data commons. To enable ready integration of analytical software over the web, Genboree functionality is made accessible programmatically via REST Application Programming Interfaces (API). REST-APIs open up the hosted services to extensions (open hosting), thus enabling Genboree to serve as a hub for integration of multiple software components to support data analysis pipelines, tools for integrative analysis, and tools for visualization of epigenomic data by geographically distributed collaborators. Genboree provides an online workbench and tools for visualization of genomic and epigenomic data for privately and publicly accessible databases and for comparative analyses of epigenomes. Analysis results, such as epigenomic states characteristic for specific cell types derived from epigenome comparisons, are organized into Genboree project pages—semi-structured editable and publishable web pages that point to structured data sets and visualizations.

**Monday, February 8, 3:15 p.m.–3:40 p.m.**
**Breakout Session A: Genomic Structure and Epigenetics**

## Whole-Genome DNA Methylation Profiles

**Matteo Pellegrini**
Molecular, Cell and Developmental Biology Department, University of California, Los Angeles

DNA methylation is an epigenetic mark that is important for the regulation of gene expression. In mammalian organisms DNA methyltransferases are essential enzymes, and it has been shown that the patterning of DNA methylation on the genome is a key regulatory component of cellular development. Bisulfite sequencing has been the standard approach for determining the methylation state of cytosines. Sodium bisulfite converts unmethylated cytosines to uracils and leaves methylated ones unchanged. Using next-generation sequencers, we have scaled this technique to the whole-genome level. We will present our results of whole-genome DNA methylation profiles across eight organisms and discuss the conservation and variability of DNA methylation from plants to animals. We also will discuss the variability of DNA methylation across multiple human stem cell lines and the mechanisms for demethylation and their implication in the development of germ cells.

**Monday, February 8, 2:00 p.m.–2:30 p.m.**
**Breakout Session B: Integrated Omics Approaches to Understand Environmental Processes**

# The (Micro)biology of Mercury

**Anne Summers**,[1] Susan Miller,[2] Mary Lipton,[3] Liyuan Liang,[4] Jeremy Smith,[4,5] Tamar Barkay,[6] Cindy Gilmour,[7] Stephen LaVoie,[1] Lyn Olliff,[1] Ben Polacco,[2] Ian Harwood,[2] Erika Zink,[3] Sam Purvine,[3] Alex Johs,[4] Jerry Parks,[4,5] and Hao-Bo Guo[4,5]

[1]University of Georgia; [2]University of California, San Francisco; [3]Pacific Northwest National Laboratory, [4]Oak Ridge National Laboratory; [5]University of Tennessee; [6]Rutgers University; [7]Smithsonian Institution

The toxic heavy metal mercury (Hg) occurs naturally as the ore cinnabar (HgS) in many places on Earth. Given its protean chemical and physical properties and despite its toxicity, Hg has been used in many ways by humans for millenia, leading to its dispersal well beyond its natural deposits. Modern spillage of Hg from large-scale industrial processes such as isotope recovery, battery manufacture, and the chlor-alkali process, among many others, has produced heavily contaminated sites worldwide from which Hg continues to disperse into soil, groundwater, and the atmosphere, resulting in exposure of humans and wildlife. Since organic and several inorganic forms of Hg are stable at standard temperatures and pressures, Hg experiences a global geocycle driven by biotic and abiotic processes. Several DOE supported groups collaborate to define key biotic steps in the Hg cycle. This talk will cover work on the molecular and cellular biology of bacterial Hg transformation, ranging from femtosecond molecular dynamics of Hg-transforming enzymes and metalloregulators to the emergence of these proteins during oxygenation of the atmosphere several billion years ago. In addition, proteomic dissection of the cellular and molecular bases of Hg toxicity in a model facultative bacterium has revealed key points of Hg vulnerability that are shared from bacteria to humans. These studies will enable more effective strategies for containment of this nearly ubiquitous toxic element, a subject of increasing public concern.

**Mon., February 8, 2:35 p.m.–3:05 p.m.**
**Breakout Session B: Integrated Omics Approaches to Understanding Environmental Processes**

# Understanding and Predicting the Response of Soil Microbial Communities to Global Change

**Eoin Brodie**
Lawrence Berkeley National Laboratory

Soil microbial communities consisting of many thousands of diverse species are typically represented by a single "black box" in current models that forecast climate change impacts on ecosystems. Accurately predicting the response of these communities to climate change is crucial since soil microorganisms are dominant mediators of carbon and nutrient transformations that ultimately shape biological productivity and environmental chemistry worldwide. Despite this global importance, the factors that determine the distribution of microbial species in soil remain unclear. Prediction of bacterial and archaeal responses to climate change is crippled by the sobering reality that the majority of microbial diversity remains uncultivated, and the fundamental physiological and ecological roles of most soil microbial species are not defined. Understanding how different microbial species or taxa respond to their environment is key to linking microorganisms with spatial models of ecosystem processes that are under microbial control. A critical first step towards predicting the impacts of climate change on terrestrial ecosystems involves determining the fundamental regulators of soil microbial community composition and structure and, subsequently, evaluating climate change scenarios that alter these factors.

Using a biogeographic approach, we studied the distributions of over 2,000 bacterial and archaeal taxa across a natural gradient from California to Puerto Rico. Despite the enormous phylogenetic and physiological complexity, soil bacterial and archaeal community structure could be predicted from a limited number of environmental factors. Soil water content was the most important environmental variable that related to differences in microbial communities among habitats. Ancestral trait reconstruction suggests that the soil moisture preferences of contemporary soil organisms exhibit a strong phylogenetic signal and have been retained over long periods of evolutionary time. Therefore, the future responses of some microbial groups to changing rainfall patterns may be predictable based on phylogenetic affiliation alone.

To test these predictions, we analyzed data from field and greenhouse rainfall manipulations representing environments with distinct climate histories. Many microbial groups responded as predicted. However, metabolomic analyses of isolates suggest that climate history should also be considered in addition to phylogenetic affiliation when attempting to predict the response of soil microorganisms to altered precipitation. To better predict responses to novel climate conditions, a mechanistic understanding of microbial community adaptation to altered soil moisture is required. To achieve this, current omic tools must be adapted to deal with the inherent differences in heterogeneity, scale, and complexity of soil microbial communities.

**Mon., February 8, 3:30 p.m.–4:00 p.m.**
**Breakout Session B: Integrated Omics Approaches to Understanding Environmental Processes**

# Novel Experimental Approaches for Proteogenomic Characterizations of Microbial Soil Communities*

**Robert L. Hettich**[1] (hettichrl@ornl.gov), Karuna Chourey,[1] Nathan VerBerkmoes,[1] Brian Dill,[1] Manesh Shah,[1] Mike Wilkins,[2] Phil Long,[2] Ken Williams,[3] Janet Jansson,[3] Kim Handley,[4] and Jillian Banfield[4]

[1]Oak Ridge National Laboratory, Tennessee; [2]Pacific Northwest National Laboratory, Washington; [3]Lawrence Berkeley National Laboratory, California; and [4]University of California, Berkeley, California

The integration of high-performance mass spectrometry (MS), computational bioinformatics, and whole-community genomics provides a powerful systems biology approach to investigate the functional activities of natural microbial communities. To this end, we have conducted extensive research 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.

One goal for this proteogenomics technology is to expand this approach into other environmental arenas, such as soil and sediments. As expected, the success in this case is greatly influenced by the ability to effectively lyse the microbial cells in the more recalcitrant soil matrix and quantitatively extract the proteome sample without significant losses or interferences. Two approaches can be envisioned: one that relies on removing the intact microbial cells from the media (indirect method) prior to cellular lysis and proteome extraction and one that attempts to perform the lysis/proteome extraction *in situ* (direct method). While the indirect approach has had some limited success in previous work, the direct method has been stymied by unacceptable sample loss, likely due in major part to irreversible binding to the hydrophobic constituents of the soil matrix. After considerable effort in evaluating standard approaches for the direct *in situ* extraction, we have formulated a new method of detergent-based lysis (SDS) and proteome extraction, coupled with a TCA precipitation cleanup. This approach integrates the power of SDS for lysing cells and solubilizing proteins with the necessary cleanup of TCA for removal of the majority of SDS prior to MS measurement. We have developed and optimized this approach with microbial isolates (*P. putida, A. chlorophenolicus*) spiked into or grown in soil microcosms. For example, 5 grams of soil (DOE Hanford site) was spiked with *P. putida* and harvested after 5 hours of growth. The SDS-TCA method yielded identification of 925 non-redundant proteins, as compared with 1343 proteins from a control isolate sample (no soil). Based on this initial success, we have extended this approach to a series of sediment samples from the DOE Rifle Integrated Field Research Challenge site. By using an LTQ-Orbitrap-MS instrument, we were able to identify more than 1000 proteins from the sediment sample, an amount which compares favorably to proteomic measurements of planktonic biomass recovered from groundwater at the same site. In total, this proteogenomic methodology provides a powerful approach for the systems biology interrogation of natural microbial communities in soils, and it should be broadly applicable for systems of relevance for environmental remediation and bioenergy production.

*This research sponsored by the U.S. DOE-BER Environmental Remediation Science Program and Genomic Science program. Oak Ridge National Laboratory is managed by UT-Battelle, LLC, for the U.S. Department of Energy under Contract No. DE-AC05-00OR22725.

**Mon., February 8, 4:05 p.m.–4:35 p.m.**
**Breakout Session B: Integrated Omics Approaches to Understanding Environmental Processes**

# BUGS Modeling: Predicting the Geochemical Consequences of Environmental Perturbations on Microbial Processes

**Derek R. Lovley**
Department of Microbiology, University of Massachusetts

Advances in molecular ecology have significantly increased the understanding of which microorganisms, genes, and proteins are present in specific environments, but such descriptive studies have little predictive capability. The ability to predict how the composition and activity of microbial communities will change in response to environmental perturbations and the geochemical consequences of these changes is essential in order to understand the likely impacts of global climate change or more localized environmental perturbations, such as those induced during engineered bioremediation of contaminated groundwater.

An approach termed Bottom Up Genome-Scale (BUGS) Modeling was developed to meet this need. The bottom up approach inherent in BUGS Modeling differs in many aspects from more commonly considered top-down metagenomic and metaproteomic approaches to the study of microbial communities. It is predictive rather than descriptive and makes modeling possible because it generates a knowledgebase that can be applied to a diversity of environments. This contrasts with the site-specific databases that result from meta-omics studies.

The BUGS Modeling approach was initially applied to predicting the outcome of potential strategies for *in situ* bioremediation of uranium-contaminated groundwater at the DOE *in situ* subsurface uranium bioremediation site in Rifle, Colorado. This case study was chosen because during the initial phase of active uranium removal the microbial community is comprised of over 90% *Geobacter* species, which are considered to be responsible for the uranium bioremediation. This simplified the microbial components that needed to be included in the initial model.

In the first phase of the BUGS Modeling approach, the *Geobacter* species that predominate during active bioremediation were recovered in pure culture, their genomes sequenced, novel aspects of their physiology investigated experimentally, and genome-scale metabolic models constructed. This process identified key metabolic genes whose expression levels are diagnostic of crucial aspects of the physiological status of the subsurface *Geobacter* community. The *in situ* levels of transcripts for these genes in the subsurface *Geobacter* community were quantified to determine such factors as whether the microorganisms were limited for important nutrients including phosphorous, ammonium, acetate, or iron; other stress responses; and *in situ* growth rates. Subsequent studies have demonstrated that highly sensitive antibody-based analysis of key proteins can provide similar diagnoses to guide in modulating bioremediation strategies to best promote the desired microbial activity in the subsurface.

As will be described in detail in the companion talk, coupling the genome-scale *Geobacter* metabolic model that resulted from these studies with geochemical and hydrological models made it possible to predict the removal of uranium from the groundwater during a field experiment at the Rifle site. Subsequent studies have demonstrated that the BUGS modeling approach can be scaled to deal with more complex communities. Sequentially including additional genome-scale metabolic models for other organisms that can be found at the Rifle site is leading to increased understanding of cooperative and competitive interactions that impact both on natural attenuation and engineered bioremediation of the uranium contamination.

The results have demonstrated that BUGS Modeling can change bioremediation from an empirical practice to a knowledge-based strategy. It is expected that BUGS Modeling will enable first principles–based prediction of the geochemical consequences of other types of perturbations, such as climate change, in a diversity of environments.

**Mon., February 8, 4:40 p.m.–5:10 p.m.**
**Breakout Session B: Integrated Omics Approaches to Understanding Environmental Processes**

## Integrated Modeling of Environmental Processes using Dynamic Genome-Based Models

**Krishna Mahadevan**
Department of Chemical Engineering and Applied Chemistry, University of Toronto

Recent advances in genome-wide characterization of microorganisms have motivated the development of computational approaches for predicting microbial physiology. More recently, Bottom Up Genome-Scale (BUGS) Modeling has been developed to better represent the microbial physiology based on the identification and subsequent genome-wide characterization of the dominant members in different environments. The application of the BUGS approach for the analysis of competition in an Fe(III)-reducing microbial community at the DOE *in situ* subsurface uranium bioremediation site in Rifle, Colorado, will be presented earlier in a companion talk.

Acetate-amendment in uranium-contaminated subsurface environment can promote the simultaneous reduction of U(VI) and Fe(III) by *Geobacter* species. It was previously observed that the initial bloom of *Geobacter* and successful removal of uranium from the groundwater following acetate addition at the Rifle site are followed by an increase in sulfate-reducing microorganisms that are ineffective in uranium reduction. Hence, it is important to model the competition between sulfate reducers and *Geobacter* species. In order to better 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 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. In addition to representing microbial community competition, another important component is the ability of predicting the extent of uranium reduction by *Geobacter* species during acetate amendment. Recently, we have developed a novel kinetic model based on recent experimental evidence of temporary electron storage in *Geobacter* species. This model was able to explain the correlation of high U(VI) removal with high fractions of planktonic cells in subsurface environments.

Finally, in order to effectively predict the outcome of uranium bioremediation, spatiotemporal physical processes need to be integrated with microbial physiology and geochemistry. Therefore, a genome-scale, constraint-based model of the metabolism of *Geobacter sulfurreducens* was coupled with the reactive transport model HYDROGEOCHEM in an attempt to model *in situ* uranium bioremediation. The coupled genome-scale and reactive transport model predicted acetate concentrations and U(VI) reduction rates in a field trial of *in situ* uranium bioremediation that were comparable to the predictions of a calibrated conventional model, but without the need for empirical calibration. Efficient approaches for integrating the genome-scale models with reactive transport models are being developed to facilitate the integration of such models with the microbial community models.

**Monday, February 8, 2:00 p.m.–2:25 p.m.**
**Breakout Session C: Imaging and Technology for Systems Biology**

## NanoSIMS Analysis of Carbon and Nitrogen Partitioning in the *Anabaena-Rhizobium* Microbial Consortium

**Alfred Spormann**
Department of Civil and Environmental Engineering, Stanford University

The epibiotic bacterium *Rhizobium* sp. WH2K forms a highly specific association with the filamentous cyanobacterium, *Anabaena* sp. SSM-00, in a two-member coculture. Although the phylogeny of this coculture's two members is known, details of the metabolic exchange between the cell types remain poorly understood. High-resolution Nanometer Secondary Ion Mass Spectrometry (NanoSIMS) was applied in conjunction with stable isotopic labeling to examine the fixation and partitioning of carbon and nitrogen between the three cell types in this consortium. Our results clearly indicate that there was significant exchange of both carbon and nitrogen from the cyanobacterial hosts to the epibionts. Consistent with the strict spatial segregation of metabolisms, distinct isotopic signatures were observed for the three cell types within a single filament-epibiont association. However, the isotopic enrichment of the same cell type found in different filaments varied greatly. This high phenotypic diversity suggests that the coculture was made up of highly heterogeneous filaments that differed greatly in their primary metabolisms and nutrient assimilation. Finally, sequential imaging of thin cellular sections with transmission electron microscopy and NanoSIMS also provides preliminary evidence of subcellular structures with distinct isotopic enrichment. These data revealed the usefulness of NanoSIMS to obtain functional insights in metabolic interaction in complex communities. Further applications of NanoSIMS in functional microbial ecology will be discussed.

**Monday, February 8, 2:25 p.m.–2:50 p.m.**
**Breakout Session C: Imaging and Technology for Systems Biology**

# Yellowstone Hot-Spring Systems Geobiology: Quantification and Prediction of Mineral-Water-Microbe Feedback Interactions in the Subsurface

**Bruce W. Fouke**

Department of Geology, Department of Microbiology, and Institute for Genomic Biology, University of Illinois, Urbana-Champaign (fouke@illinois.edu)

A fundamental shift is under way in the geosciences in response to the recognition that microorganisms have played a defining role in the co-evolution of our planet and biosphere. This realization has been shaped by the application of DNA biotechnology to a wide variety of geological marine and terrestrial environments around the world. As a result, geoscientists are now probing one of the foremost theoretical and practical scientific questions of our time: *How have Microbial Life and Earth co-evolved through geological history and what will future interaction yield in the face of ongoing global environmental change and resource utilization?* The development of a "systems geobiology" approach, which integrates techniques from the geological, chemical, physical, and life sciences, is now required to develop and utilize subsurface carbon sequestration and enhanced oil recovery. Mammoth Hot Springs at the northern border of Yellowstone National Park has served has a natural lab for simultaneously measuring multiple chemical and biological species at multiple scales. Cross-disciplinary parameterization of hot-spring travertine ($CaCO_3$)–water–microbe systems at this site has been linked with a process-based sedimentary depositional facies model for paleoenvironmental and paleobiological reconstructions. Results have identified microbial influences on carbonate rock fabric, isotopic geochemistry and precipitation rate that are applicable to understanding feedback interactions between biotic growth and activity, mineral precipitation dynamics, and rapid environmental change in other modern and ancient geological settings.

The spring water at Mammoth Hot Springs in northern Yellowstone is derived from rain and snowmelt runoff in the Gallatin Mountains that flows down along faults into the rock subsurface. This groundwater is then heated by the Yellowstone supervolcano to ~100°C (212°F), chemically dissolves deeply buried ~350 million year old marine limestone, and flows back up to the surface to emerge from vents at a temperature of 73°C (163°F). The limestone rock (called *travertine*) that precipitates to form the classic meter-scale terraced steps of Mammoth Hot Springs are composed of a form of calcium carbonate ($CaCO_3$) mineral called aragonite. Parameterization across broad spatial and temporal scales of the Mammoth hot-spring drainage system has included (a) **Travertine**: analyses of geomorphology, crystal growth (size, shape, porosity, and structural organization at multiple scales), mineralogy, and elemental and isotopic geochemistry (Ca, Mg, Sr, Mn, Fe, $SO_4$, $\delta^{13}C$, $\delta^{18}O$, $\delta^{34}S$, and $^{87}Sr/^{86}Sr$); (b) **Spring Water**: triplicate measurements of temperature, pH, flow dyanamics (primary flow path, depth, and velocity) and elemental and isotopic geochemistry (DO, DIC, alkalinity, $S^{-2}$, $SO_4$, $NO_4$, Ca, Mg, Sr, Mn, Fe, Si, $\delta^{13}C$, $\delta^{18}O$, $\delta^{34}S$, $^{87}$and Sr/$^{86}Sr$); and (c) **Microbial Community**: metagenome-enabled analyses of community structure and phylogenetic diversity (field and microscope observations, 16S rRNA gene sequence clone libraries, and T-RFLP), metabolic activity (inference from 16S rRNA gene sequence composition, lipids and their $\delta^{13}C$, gene expression, and metagenomic analyses). Ongoing quantitative modeling and synthesis are being simultaneously focused on mechanistic, process or pattern hierarchical levels, including the integration of sedimentary dynamics and mineral precipitation with microbial metabolic energetics.

**Monday, February 8, 2:50 p.m.–3:15 p.m.**
**Breakout Session C: Imaging and Technology for Systems Biology**

## Super-Resolution Optical Methods for the Genomic Sciences: Possibilities and Problems

**Jan Liphardt**
Lawrence Berkeley National Laboratory

Over the last several years, there have been enormous breakthroughs in light microscopy. The diffraction limit, which restricts the resolution of conventional light microscopy to about 250 nanometers, can now be bypassed in several ways. Optical technologies such as PALM, STORM, and STED are providing unprecedented, exquisitely detailed views of microbes, the plant cell wall, and, in certain cases, even molecular machines. Like any technology, the new "super-resolution" optical methods also have various particularities and limitations, which at present can make it very hard to actually apply these methods to DOE genomic science problems. I will brief you on what is currently possible and enumerate unsolved problems in probe chemistry, imaging hardware, and data processing algorithms.

**Monday, February 8, 3:30 p.m.–3:55 p.m.**
**Breakout Session C: Imaging and Technology for Systems Biology**

# Chemical Imaging using Mass Spectroscopy

**Jonathan V. Sweedler**[1] and Paul W. Bohn[2]

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

Given the rapid advances in measurement sciences over the past decades, which measurement challenges remain that impact DOE programs? While multiple techniques can characterize a surface and provide either chemical or spatial information, they tend not to provide such information at the single cell resolution and multidimensional information often is lacking. Understanding the functioning of cells within microbial communities or within a mammalian organ requires approaches that provide multidimensional and contextual information. Here a range of mass spectrometry (MS)–based approaches are highlighted that provide unmatched chemical information on the cellular metabolome and proteome. While traditional MS-based approaches provide little spatial information, several approaches combine the chemical information–rich data from mass spectrometry with spatial information. The addition of imaging to mass spectrometry allows rare cells to be characterized, provides information on the functional context of the detected analytes, and allows us to probe the interface between cells and their environment. MS-based imaging can use several MS platforms, including secondary ion mass spectrometry and laser desorption ionization mass spectrometry. Using these approaches, we have characterized large numbers of cell-to-cell signaling molecules in several well-defined cellular networks. In order to enhance the chemical and spatial information, Raman microspectroscopy and MS imaging are combined to provide unmatched figures of merit for spatially resolved imaging of lignocellulosic materials. The application of such heterocorrelation spectroscopies to visualizing the structural and chemical changes during the processing of lignocellulosic materials demonstrates the advantages of multimodel chemical imaging approaches. The potential for other multidimensional information-rich characterization approaches will be discussed.

**Monday, February 8, 3:55 p.m.–4:20 p.m.**
**Breakout Session C: Imaging and Technology for Systems Biology**

## Applications of Full-Field Transmission X-Ray Microscopy to Nano- and Thick Bio-Materials

**P. Pianetta**

Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, Menlo Park, California

Exciting developments in full-field hard X-ray transmission microscopy have resulted in imaging performance of better than 30-nm resolution and uniform fields of view of 30 $\mu m^2$ that are enabling applications in nanoscience, biomaterials, and environmental science. A specially modified instrument from Xradia, operating in the 5- to 15-keV photon energy range using both absorption and phase contrast for 2D and 3D tomographic imaging, has been installed at beam line 6-2 at the Stanford Synchrotron Radiation Lightsource. Phase contrast imaging of trabeculae from mouse tibia, loaded *in vivo* to study the effects of weight bearing on bone structure, revealed a complex network of osteocytes and canaliculi, as well as yielding quantitative information of bone density variations at the nanoscale. In the area of environmental science, measurements of cordgrass roots exposed to mercury revealed insights into the process of mercury transformation by microorganisms living in the biofilm that surrounds the root. X-ray tomography shows the presence of micron-sized spherical balls surrounded by a thin layer of mercury. These results provide insight into the mechanism of mercury methylation in the environment. Studies in the fields of dentistry, alternative energy, and nanoparticle formation by yeast also are being pursued. Future applications that make use of X-ray absorption near-edge spectroscopy for chemical speciation at the nanometer scale also will be discussed.

**Monday, February 8, 4:20 p.m.–5:00 p.m.**
**Breakout Session C: Imaging and Technology for Systems Biology**

## Micro- and Nanotechnologies for Studying the Plant-Microbe Interface

**M. J. Doktycz**
Oak Ridge National Laboratory

Understanding the interfaces between organisms is key to realizing their functional roles in biological and environmental systems. The beneficial association of plants and microbes exemplifies a complex, multiorganism system that is shaped by its participating organisms and the environmental forces acting upon it. The interfaces between plants and microbes can benefit plant health and biomass production by affecting nutrient uptake, influencing plant hormone signaling, inducing catabolism of toxic compounds, or conferring resistance to pathogens. Plants and microbes function together to effect remediation processes, manage global carbon cycling, and offer alternative energy solutions. Practical extrapolation of the roles of organisms in complex natural settings will require understanding the correlations between genetic characteristics of the organisms involved and how information, energy, and materials are proportioned and exchanged between them. Numerous challenges are associated with understanding the interfaces between plants and microbes. Identifying and characterizing the diverse organisms involved and dissecting the molecular-based exchanges and genetically derived responses are just a few of the issues that must be addressed. Micro- and nanoscale technologies are well suited for aiding these investigations. Microfabrication techniques allow for the construction of mechanical and fluidic devices that are ideal for manipulating materials at the molecular and cellular level. Further, such structures can facilitate sampling and analysis of small quantities. By matching the functional scale of biology, structures fabricated at the nanoscale allow additional capabilities.

This presentation will highlight three uses of micro- and nanotechnologies that are being developed for studying different aspects of the plant-microbe interface. In one project, a microfabricated device that enables the parallel, microscale culturing and characterization of individual members of a microbial community is being pursued. Using microfluidics, single cells derived from complex communities are being encapsulated into alginate gel microdroplets to allow for small-scale growth of thousands of isolated cells in parallel. Segregation of a cell into a gel bead can facilitate subsequent sorting and selection based on functional characteristics. The scale, throughput capabilities, and sensitivity of the proposed technology address key challenges facing the analysis of microbial consortia. In another application, microfluidic structures, combined with nanoscale features, are being developed for functionally assaying microbes. Chemotaxis assays are being developed in order to understand species variation in bacterial motility. Taxis plays an early and ongoing role in processes that affect gene expression, cell recruitment, surface recognition, invasion, and colonization. Finally, nanofabrication techniques are being used to create small volume reaction containers that mimic functional aspects of a cell. These "cell mimics" are being used to construct and test synthetic gene networks in order to understand their operation and to use them as interfaces to natural cells. Progress related to the development of these technologies will be presented.

**Tuesday, February 9, 8:30 a.m.–9:00 a.m.**
**Plenary Session: Facilities**

## DOE Joint Genome Institute Update

**Eddy Rubin**
DOE Joint Genome Institute, Lawrence Livermore National Laboratory, California

Over the past decade, the DOE Joint Genome Institute (JGI) has occupied a unique niche as a user facility focused on applying large-scale genomics to accelerate the progress of scientists working in DOE mission areas of biology. The DOE JGI's user facility approach is based on the concept that by focusing the most advanced sequencing and analysis resources on the best peer-reviewed proposals drawn from a diverse community of scientists, the DOE JGI will catalyze creative approaches to DOE mission challenges.

This has been coupled recently with the seismic technological changes presently under way in the field of genomics with the advent of next-generation sequencers that have produced a 50-fold increase in JGI's sequencing capacity over a 3-year period. New sequencing techonologies have specifically facilitated deep metagenome sequencing and variation detection by resequencing, while *de novo* sequencing by new technologies remains a significant challenge. Several examples will be presented of JGI projects exploiting next-generation sequencing capabilities to derive important biological insights.

**Tuesday, February 9, 9:00 a.m.–9:30 a.m.**
**Plenary Session: Facilities**

## Proteomic Insights into Microbes, Microbial Communities, and Plants: An EMSL Perspective

**Mary S. Lipton*** (mary.lipton@pnl.gov)

Biological Sciences Division and Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory

Significance: Characterization of biological systems using comprehensive global proteomic studies enhances scientific understanding through improved annotation of genomic sequences, elucidation of phenotypic relationships between environmentally important microorganisms, characterization of higher organisms such as plants, characterization of the metabolic activities within microbial communities, and identification of post-translationally modified proteins.

Proteomic applications support DOE missions and science by exploiting microbial and plant function for purposes of bioremediation, energy production, and carbon sequestration among other important areas. Inherent to exploiting microbial function or utilizing plants as biofuels is the detailed understanding of the physiology of the cell. This physiological state is determined by which proteins are expressed in the cell, where they are located, and their modification state.

EMSL is a national scientific user facility at Pacific Northwest National Laboratory that provides integrated experimental and computational resources for discovery and technological innovation in the environmental molecular sciences to support the needs of DOE and the nation. Proteomics serves as one of the foundational scientific capabilities within the EMSL and, as such, enables research critical

to the DOE mission. This presentation will illustrate how the proteomics pipeline in EMSL is utilized to address organism-specific scientific objectives developed in conjunction with biological experts for a number of different microbes and plants. We will highlight the ability to use proteomics data for genome annotation of microbes and fungi, characterization of microbial communities, advances in the characterization of protein modification state, and the identification of new proteins important to photosynthesis, and the determination of protein localization in stem, root, and leaf tissues of plant cells.

Additional information and supplementary material can be found at the PNNL proteomics website at http://ober-proteomics.pnl.gov/ and the EMSL website http://www.emsl.pnl.gov/emslweb/.

*Acknowledgements: 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 Office of Biological and Environmental Research. Pacific Northwest National Laboratory is operated for the U.S. Department of Energy by Battelle Memorial Institute through Contract No. DE-AC05-76RLO 1830.

**Tuesday, February 9, 10:00 a.m.–10:40 a.m.**
**Plenary Session: Systems Biology Developments to Enable a Genomic Sciences Knowledgebase**

## Systems Integration of Omics Data for Biological Network Discovery

**Cathy H. Wu**

Center for Bioinformatics and Computational Biology, University of Delaware

Systems integration is becoming the driving force for 21st Century biology. Researchers are systematically tackling gene functions and complex regulatory processes by studying organisms at different levels of organization, from genomes, transcriptomes, and proteomes to metabolomes and interactomes. Fully realizing the value of such high-throughput data requires advanced bioinformatics for integration, mining, comparative analysis, and functional interpretation. The Protein Information Resource (PIR) is developing a bioinformatics research infrastructure that links data mining with text mining and network visualization and analysis in the systems biology context for biological network discovery. Built on major bioinformatics community resources to promote interoperability and scalability, the system framework integrates data types critical for genome-scale network reconstruction. The underlying data warehouse integrates over 100 databases relevant to systems biology, extending from the UniProt Knowledgebase to cover biochemical, enzyme, and regulatory data required for genome-scale modeling. The system is interoperable with common standards and controlled vocabularies, as well as ontologies in the Open Biomedical Ontologies (OBO) Foundry framework to capture functional properties of proteins, modifications, and complexes in metabolic and signaling pathways. Linking with the system biology data and ontology, the text-mining component integrates publicly available text-mining resources and tools to extract literature-based experimental data and scientific knowledge. The network analysis and visualization component connects to a network prediction method based on profiles of interacting domains of known 3D protein structures from the Protein Structure Knowledgebase, as well as a visualization tool for interactive visual analysis of biological networks. The integrative approach will reveal hidden interrelationships among the various components of biological systems, allowing researchers to ask complex biological questions and gain better understanding of biological processes, thereby facilitating biological network discovery.

**Tuesday, February 9, 10:40 a.m.–11:20 a.m.**
**Plenary Session: Systems Biology Developments to Enable a Genomic Sciences Knowledgebase**

## A Knowledgebase for Rapid Inference and Re-Engineering of Biological Circuits

**Nitin Baliga**

Institute for Systems Biology, Seattle, Washington

There is much hype about leveraging 3.5 billion years of evolution to solve pressing problems concerning environment and energy. The basic strategy is to design novel (unnatural) biochemical capabilities for environmental cleanup and energy production by recombining and rationally re-engineering biological circuits from diverse organisms. In principle, it is possible to decipher and re-engineer the complex information processing circuits that dynamically reconfigure the physiology of a particular organism in response to environmental change. In practice, this would require extensive mining of systems measurements for conditional relationships among patterns of changes in gene expression (mRNA, protein, and ncRNA), interactions (P-P and P-D), modifications (protein and DNA), and metabolism (metabolite levels, enzyme activities). For synthetic biology applications, it is essential to capture relevant conditional relationships both at a systems level and at a sufficiently high resolution to mechanistcially describe and predict how environmental change influences the execution of these cellular algorithms at multiple scales. Clearly, the experimental and computational tools necessary for doing this type of multiscale network inference and modeling are diverse, disjointed, and constantly changing for a purely monolithic knowledgebase solution to be practical. A knowledgebase that is built upon an architecture of loosely coupled resources, on the other hand, is both highly adaptable to change and essential for large collaborative and interdisciplinary systems biology efforts for rapid inference and re-engineering of biological circuits.

**Tuesday, February 9, 11:20 a.m.–12:00 noon**
**Plenary Session: Systems Biology Developments to Enable a Genomic Sciences Knowledgebase**

## How Easy Is It To Interpret Function from CAZy Family Location?

**Harry Gilbert**

Complex Carbohydrate Research Center, University of Georgia

This lecture will briefly review the criteria for including enzymes in specific CAZy families, and what this tells us about the general features of these biocatalysts. Examples will be provided in which the CAZy location of proteins provides insight into specificity, and in which a more refined analysis is required to predict such activity. The bulk of the lecture will focus on how CAZy might be used to develop more sophisticated tools for predicting protein function. Finally, the lecture will suggest additional strategies that may be deployed to mine novel activities obtained from genome and metagenome programs.

**Tuesday, February 9, 2:00 p.m.–2:20 p.m.**
**Breakout Session D: Systems Biology Knowledgebase Workshop for Genomic Sciences Applications**

# Microbial Systems Biology Knowledgebase: Scientific Objectives and Current Prospects

**Robert W. Cottingham**

Computational Biology and Bioinformatics, Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee

The DOE Office of Biological and Environmental Sciences has begun the Systems Biology Knowledgebase program by funding as a first effort the development of the requirements and specifications called the **Knowledgebase R&D** project. Following on from the vision described in the original workshop report, *Systems Biology Knowledgebase for a New Era in Biology*, the Knowledgebase R&D project activities include workshops (such as this one), pilot projects, and a final report establishing an implementation plan necessary for the subsequent development of the Knowledgebase.

In order for the Knowledgebase development effort to be successful, it is very important that there be clear, achievable scientific objectives. For example, with the development of RNAseq technology, integration of genomic sequence from phylogenetically related organisms along with high-resolution expression data from multiple biological states provides the basis for improved determination of gene regulatory networks. Another objective could be to integrate phenotypic response with specific genotypes or pathways such that regulatory or genetic changes could be predictably associated with microbial behavior and response. We will discuss objectives based on current research activities and consider candidates and priorities to recommend. Ideally, these objectives should come from the research community and leaders, not from the bioinformatics or development teams.

The Knowledgebase will be an open, community-driven, computational infrastructure, designed to provide data and analysis in a more accessible manner than current systems. Everything will be available via a browser. The system will be developed in the open somewhat analogous to Wikipedia and encourage community contribution. Community development means that systems do not have to be dependent on only the initiative and interest of one researcher or a small team. Where different analysis methods are available, the system will be designed to facilitate comparison of methods. These are some of the expected features and benefits of the Knowledgebase. It is expected that such characteristics will be much more supportive of the scientific endeavor than current computational methods.

**Tuesday, February 9, 3:00 p.m.–3:30 p.m.**
**Breakout Session D: Systems Biology Knowledgebase Workshop for Genomic Science Applications**

## Near-Term Prospects for Functional Microbial Genomics: Moving Beyond the Monoculture Paradigm

**Robert M. Kelly**
Department of Chemical and Biomolecular Engineering, North Carolina State University, Raleigh

Over the past 15 years, microbial genomics has gone from "square one" to a point that would have been difficult to predict at the outset of the genomics era. New genome sequence information appears in public databases daily (hourly?), and the challenge is how to store, analyze, and utilize such data in a meaningful way to foster scientific and technological advances. A framework (i.e., systems biology) has arisen in response to this challenge, but strategies are needed to take advantage of this new context for examining microbial biology.

Most of what is now known about microbial biology was learned from the study of pure laboratory cultures. The "monoculture" paradigm has been extraordinarily productive and will continue to be at the heart of microbiology. However, monocultures are not representative of how microbial systems exist in nature (and, perhaps, not even in the laboratory). To this end, metagenomics has provided a means for examining microbial complexity, but complementary functional information is still needed to understand the "metaphenotype."

Over the next several years, efforts are needed to link the complexity reflected in metagenomes to what is already known from monoculture studies. This learning curve will necessarily start with relatively simple systems, for even co-cultures can exhibit phenotypes not easily predicted from pure culture information. The extension of functional microbial genomics beyond monocultures will be discussed with an eye towards the integration of experimental design, experimental methods, and data analysis strategies. Examples will be provided to illustrate the challenges that arise when moving beyond monocultures.

**Tuesday, February 9, 4:00 p.m.–4:30 p.m.**
**Breakout Session D: Systems Biology Knowledgebase Workshop for Genomic Sciences Applications**

# From Pathways to Populations and Back Again: Long-Term Prospects for the Microbial Systems Biology Knowledgebase

**Adam P. Arkin**

Head, Synthetic Biology, Physical Biosciences Division, Lawrence Berkeley National Laboratory; and Director, Bioinformatics, Joint BioEnergy Institute, Department of Bioengineering, University of California, Berkeley

The grand challenge to predict phenotype from genotype is particularly challenging in the microbial world. At its heart this challenge seeks to understand the principles of biological architecture and function sufficient for predicting behavior and, of course, for changing it. A systems biology knowledgebase should grow into an indispensible tool for molecular, environmental, evolutionary medical and epidemiological microbiologists, and biotechnologists to understand and engineer their systems. However, there are challenges in accomplishing this that are found in few other systems.

Microbes rarely work alone, but operate in complex communities that form spatial and temporal webs of mutual support, parasitism and predation. Perhaps unique to microbes and their communities is the astonishingly rapid mechanisms for evolution and the deeply intertwined ecology of mobile genetic elements that aid in the preservation, diversification, and dissemination of function and may be central drivers themselves of the architecture of microbial networks.

The knowledgebase, in the long term, will be faced with capturing and interrelating data about all these processes at scales from molecular to meters. Sequencing technologies give us information on identities of microbial players in these communities and can hone in on some aspects of gene expression. Structural techniques can give us key information on molecular identity and sometimes function. New imaging technologies can give us information on the arrangements and interactions among molecules, cells, and their environment. But the complexity of the data increases greatly as one moves away from the sequence of single genomes and crystal structures of single proteins. It also becomes far more conditional on unmeasured conditions and interactions and less precise and accurate metrologically, all of which present challenges for organizing and navigating this information.

A pass through how such information could be assembled, navigated, and used in such a knowledgebase will be given. At each level, the challenges and acuteness of need for the community will be discussed.

**Tuesday, February 9, 2:00 p.m.–2:30 p.m.**
**Breakout Session E: Synthetic Biology**

## Synthetic Genomics: Progress on Construction of a Synthetic Bacterial Cell

**John Glass**
Synthetic Biology Group, J. Craig Venter Institute

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 300 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.

**Tuesday, February 9, 2:30 p.m.–3:00 p.m.**
**Breakout Session E: Synthetic Biology**

## Integration of BioCAD Tools, Parts Registries, and Automated Assembly

**Nathan J. Hillson**
Fuels Synthesis Division, Joint BioEnergy Institute, Lawrence Berkeley National Laboratory

The specification, design, construction, and introduction of exogenous metabolic pathways, subcellular structures, control circuitry, and other various devices into host organisms of choice are common synthetic biology tasks. The recent emergence of biological parts registries, biological computer-aided design (BioCAD) tools, and automated DNA-assembly methods promises to greatly facilitate the execution of these tasks and to increase the scope of what is readily achievable experimentally. Current progress towards the development and integration of these three technologies at the Joint BioEnergy Institute will be discussed.

**Tuesday, February 9, 3:00 p.m.–3:30 p.m.**
**Breakout Session E: Synthetic Biology**

## Genome Engineering, Multi-Virus Resistance, and Accelerated Evolution for Industrial Chemicals

**Geroge M. Church**

Genetics Department, Harvard Medical School

We have reduced the costs of reading and writing DNA by five logs in the past 5 years via multiplexing and micron-scale chip chemistry. This enables genome-wide design, standardization, interoperable parts, and multi-genome Fab scale. But the next phase of molecular engineering can embrace a feature uncommon in other engineering disciplines, which is evolution—trillions of designed constructs competing in accelerated (1) Fab, (2) Sensing, (3) Selection loops. (1) A key new Fab tool is Multiplex Automated Genome Engineering [(MAGE), Wang et al., *Nature* July 2009]. (2) The Sensor Project collects, designs, and evolves molecular binding motifs coupled allosterically or enzymatically to high differential growth markers (ideally, pos/neg like TolC in *Escherichia coli* and URA3 in yeast). (3) Selection tools include (a) MAGE, (b) Polonator sequencer with DMD photorelease chemistry, and (c) *in vitro* ribosome display. Computer Aided Design (CAD) can direct MAGE to achieve specific genome designs (e.g., new translation code for multi-virus resistance in production strains) or to explore 1E5 to 1E10 combinatorial genome designs per day to optimize hydrocarbons (alkanes, Lycopene, fatty acid esters) or chemical feedstocks (Tyr and Trp).

**Tuesday, February 9, 3:50 p.m.–4:10 p.m.**
**Breakout Session E: Synthetic Biology**

## Engineering the *Bacillus subtilis* Metal Ion Homeostasis System To Serve as a Cadmium-Responsive Biosensor

**Alyssa Henning**

CUGEM, iGEN, Cornell Team, Cornell University

Cadmium is a toxic heavy metal that has no known biological function. Ingestion of water contaminated with cadmium can induce bone fractures and severe renal damage. Major sources of cadmium contamination include fertilizers, sewage sludge, manure, and atmospheric deposition. Cadmium-contaminated sewage is often used for irrigation purposes in many parts of the world, especially in developing nations. Current analytical methods such as atomic absorption spectroscopy, though highly sensitive, are too expensive for field deployment and are unable to measure the amount of bioavailable cadmium. The goal of our project is to engineer a cheap cadmium biosensor using *Bacillus subtilis* as the chassis. This *in vivo* cadmium biosensor is composed of two modules: the positive cadmium-responsive promoter, cadA, attached to the cyan fluorescent reporter protein (CFP), and the negative cadmium-responsive promoter, mntH, attached to the yellow fluorescent reporter protein (YFP).

We expect the two modules to respond differently to cadmium. As intracellular cadmium concentrations rise, we anticipate an increase in fluorescence at the peak emission wavelength for CFP, 476 nm, and a decrease in fluorescence at the peak emission wavelength for YFP, 527 nm. Both modules will work synergistically to enhance the signal-to-noise ratio and cancel out stochastic fluctuations. Because these modules do not respond exclusively to the cadmium ion, *in vivo* fluorescence measurements need to be calibrated according to baseline values to correct for false positive readings.

**Tuesday, February 9, 4:10 p.m.–4:30 p.m.**
**Breakout Session E: Synthetic Biology**

## BioBricks Without Borders: Investigating a Multi-Host BioBrick Vector and Secretion of Cellular Products

**Elisabeth Linton** and **Charles D. Miller**
Department of Biological and Irrigation Engineering, Utah State University

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 *Escherichia 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 because of 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.

**Tuesday, February 9, 4:30 p.m.–5:00 p.m.**
**Breakout Session E: Synthetic Biology**

# Synthetic Genomics: Oxymorons and Communication Bloopers—Why Researchers' Enthusiasm for "Tinkering with Life" Might not be Contagious

**Lori Knowles**
Health Law Institute, University of Alberta

When the benefits of synthetic biology are self-evident to those involved the research, why might society not rally behind biological engineering?

This seemingly simple question about what might concern people about synthetic genomics is a complicated story of communication missteps and the baggage of biotechnology—from GM crops to Dolly. Concerns about what research is being conducted, who is conducting the research and why, or what benefit it aims to achieve are central questions that researchers must be prepared to address. Regardless of whether synthetic genomics is just another biotechnological tool or is actually something novel, its ultimate acceptance by society will be predicated on a number of factors. The utility of first applications, who stands to benefit, whether innovation is privately owned or pooled or shared, and the scope of first failures will directly impact society's view of the risks and benefits synthetic genomics brings to bear. Of paramount importance is how researchers and scientists communicate what they are doing and why they are doing it.

Synthetic biology offers the possibility to construct novel organisms, and DNA synthesis provides a faster, more efficient research tool than isolating sequences from nature. DNA synthesis makes biological engineering available to a wide range of engineers and biologists. This democratization of science is exciting and has been embraced by the DIY biology community. While events such as the recent "Outlaw Biology" symposium are meant to explore the creativity these new tools permit, the underlying message that "pirates, ninjas, outlaws" are all welcome plays directly into public concerns about researcher responsibility.

To many people "synthetic biology" is as oxymoronic as "playing God." Part of the concern with the concept of "playing God" lies in the idea of *playing* at being a creator. How seriously do biological engineers take their work? Do they recognize the significance and power of constructing life or are they just "tinkering"? Does this ability to construct and design life devalue life itself or open up life's mysteries to us all? Researchers must engage in discussions with members of the public and be prepared to explain the motives of their work and why it's important. Talking clearly and carefully about research is now part of a researcher's responsibilities. Equally, what research would *not* be undertaken due to ethical or societal constraints is also part of a researcher-public dialogue. Lessons from past biotechnology debates are instructive.

**Tuesday, February 9, 2:00 p.m.–2:25 p.m.**
**Breakout Session F: Structural Biology Applications in Genomic Sciences**

# Structural Studies of Heme Containing Proteins from *Geobacter Sulfurreducens* Using the Structural Biology Center Beam Lines at the Advanced Photon Source

**Marianne Schiffer**[1] (mschiffer@anl.gov), P. Raj Pokkuluri,[1] Yuri Y. Londer,[1] Derek R. Lovley,[2] and Andrzej Joachimiak[1]

[1]Biosciences Division, Argonne National Laboratory, Argonne, Illinois; [2]Department of Microbiology, University of Massachusetts, Amherst

The bacterium *Geobacter sulfurreducens* has been shown to exhibit electron transfer properties that are important for the development of bioremediation tools as well as biofuel cells. Specifically, multiheme cytochromes have been implicated in the electron transport chain(s) used by *G. sulfurreducens* for the reduction of soluble and insoluble Fe(III), and for transfer of electrons to electrodes. Interestingly, the *G. sulfurreducens* genome encodes 111 *c*-type cytochromes, some of which are responsible for the electron transfer functions in the versatile respiratory pathways characteristic of this organism. We have been using X-ray crystallography enabled by the Structural Biology Center (SBC) at the Advanced Photon Source (APS) to determine the structure of several of these *c*-type cytochromes in an effort to provide mechanistic insight into *Geobacter's* electron transport capabilities and to monitor its redox environment.

The periplasmic cytochromes from the $c_7$ family, for example, are of interest since they are required for the reduction of Fe(III) and U(VI). Five10-kDa, three-heme cytochromes $c_7$ are encoded in the *G. sulfurreducens* genome with sequence identity varying between 45% and 77% and having the same overall protein fold. Our detailed analysis of the structures provided a possible rationalization for the existence of five similar proteins that can have different functions required by the organism. (*BBA-Bioenergetics* 2010).

In a separate study, polymers of $c_7$-type domains were analyzed, 2 consisting of 4 domains and 1 of 9 domains that contain a total of 12 and 27 hemes per protein, respectively. Structure determination of one of the four domains of proteins revealed a new type of multiheme cytochrome that forms a "nanowire." Also, the structure showed that this family has unusual heme coordination; in each domain the third heme has His-Met coordination, whereas in $c_7$ all three hemes have bis-His coordination. (*Protein Sci.* 2004)

Ten of the G. sulfureducens proteins annotated as parts of two-component signal transduction and/or chemotaxis pathways contain c-type hemes. We determined the structures of two of the sensors from chemotaxis proteins that might sense the redox potential in the periplasm. In the crystal, both sensor domains form ìswappedî dimers and reveal a novel way of forming PAS-like domains using two chains. We suggest that the swapped dimer formation could be a mechanism for signal transduction from the periplasm to the cytoplasm. (*JMB* 2008)

These studies were carried out at the SBC by multiple-wavelength anomalous dispersion (MAD) method using the heme Fe atoms. SBC is a national scientific user facility at Argonne National Laboratory that operates two beamlines—one insertion device (ID) and one bending magnet (BM). Located at Sector 19 of the APS, the beamlines can deliver X rays onto a micrometer-size crystal sample with very low angular divergence. Diffraction data is recorded on large, fast, and efficient CCD area detectors and is processed on high-performance, integrated computing systems with beamline control and data analysis software designed specifically for the SBC.

**Tuesday, February 9, 2:25 p.m.–2:50 p.m.**
**Breakout Session F: Structural Biology Applications in Genomic Sciences**

# Using Neutron Crystallography To Reveal the Mechanism of Enzymes for Renewable Energy and the Environment

**Paul Langan** (Langan_paul@lanl.gov),[1] Jenny P. Glusker,[2] B. Leif Hanson,[3] L. Carrell,[2] Amy K. Katz,[2] S. Zoe Fisher,[1] Hannah Johnson,[1] Andrey Kovalevsky,[1] Marat Mustyakimov,[1] Benno, P. Schoenborn,[1] and Mary Jo Waltman.[1]

[1]Bioscience Division, Los Alamos National Laboratory, Los Alamos, New Mexico; [2]Fox Chase Cancer Center, Philadelphia, Pennsylvania; and [3]Department of Chemistry, Toledo University, Toledo, Ohio.

We are investigating several enzymes that are important to Genome Science program missions in renewable energy and the environment, with a view to understanding their detailed catalytic mechanisms. This new information is then being exploited to manipulate their performance and use. The talk will focus on a major problem for the DOE Bioenergy mission: the presence of xylose in biomass hydrolazates. Xylose is a pentose sugar that cannot be fermented by *Shewanella cerevisiae*.

Extensive efforts are being made to engineer *S. cerevisiae* with an efficient xylose metabolic pathway to produce xylulose, the fermentable keto isomer of xylose. One pathway is isomerization of xylose to xylulose by expression of *xylA*, the gene found in some anaerobic fungi and bacteria that encodes Xylose Isomeraze (XI). Heterologous expression of *xylA* has been problematic, leading to searches for orthologues that may express XI with improved performance. However, although XI from a different organism can have low sequence identity (<30%), X-ray crystallography has shown that their active sites and, by extension, their catalytic mechanisms are well conserved. We are elucidating the mechanism in detail in order to provide insights that will allow us make changes to the active site that improve catalytic performance and binding specificity and that are generalizable to a broad range of XIs with different thermal and pH stabilities.

Most of the biochemical reactions of importance in this system involve transferring hydrogen, an atom which is difficult to locate using crystallography with X rays, but which is easy to locate with neutrons. Using the neutron Protein Crystallography Station (PCS), a BER-funded experimental user capability at Los Alamos National Laboratory, we are revealing the movement of hydrogen in several key enzymatic processes, including the XI catalysis process.

XI catalysis is a multi-step process that can be understood at a molecular level only through a detailed characterization of each step. We have trapped the enzyme in complex with its metal cofactors and substrate at different stages of the chemical reaction by using different metal species. Unexpectedly, the locations and possible movements of hydrogens determined in this study do not support many aspects of the several proposals that have been advanced for the mechanism of XI. Rather they lead to other possible mechanisms, which we are using to guide protein engineering efforts.

In addition to using the PCS for neutron beam-time, we have made broad use of its user support for deuteration, the expression of proteins, and the synthesis of substrates with stable isotopes, as well as for support for data reduction and structure analysis.

**Tuesday, February 9, 2:50 p.m.–3:15 p.m.**
**Breakout Session F: Structural Biology Applications in Genomic Sciences**

# Developed Proteomics Scale Solution X-Ray Scattering (SAXS) Tools Applied to Metabolic Networks of Interest from the ENIGMA Program Project

**Greg Hura**
Lawrence Berkeley National Laboratory

Protein functional relationships involve interaction mosaics that self-assemble from independent macromolecular pieces tuned by modifications and metabolites. The molecular machines involved in genetic maintenance are a particularly illustrative case because many of the players and their functions are known. Members of these pathways encounter a large host of known metabolites and partners, which change conformations enabling catalysis and the transmission of information for subsequent steps and processes. We are now targeting the enzymes involved in the modified glycolytic pathway of the extremophile *Pyrococcus furiosus*, which grows optimally near 100°C. This organism employs a membrane-bound hydrogenase as part of its energy conservation respiratory mechanism.

Understanding the structural biology of protein machines not just in one condition but in varied contexts is key to understanding how they accomplish their tasks. With most common techniques, access to this comprehensive information is beyond reach given their bottlenecks. We are using a small-angle X-ray scattering (SAXS) facility (SIBYLS), which enables facile access to the shape and assembly of macromolecules in solution. We have applied SIBYLS to a host of enzymes involved in genetic maintenance under a variety of contexts with and without substrate or partners. We are providing a comprehensive view of the mechanisms and functions of these machines. Based on technologies developed in our studies of the DNA ligase PCNA complex from the extremophile *Sulfulobus sulfotaricus*, we are characterizing complexes formed as part of the pathway by which *P. furiosus* oxidizes sugars, produces hydrogen gas, and conserves energy—in particular, the steps that couple electron transfer to and from the redox protein ferredoxin in the oxidation of pyruvate to acetyl CoA and $CO_2$ by the enzyme pyruvate ferredoxin oxidoreductase (POR). Within the cell the reduced ferredoxin that is generated is subsequently oxidized by a membrane-bound hydrogenase that both pumps ions and evolves $H_2$ gas. The shapes and assemblies of these enzymes with substrates are undergoing characterization to ultimately provide a comprehensive picture of this pathway.

**Tuesday, February 9, 3:30 p.m.–3:55 p.m.**
**Breakout Session F: Structural Biology Applications in Genomic Sciences**

## Novel Insights into the Structural Mechanism of Lipid Accumulation Within Algae Using Soft X-Ray Tomography

**Blake Simmons**,[1] Carolyn Larabell,[2] M. A. Le Gros,[2] Huawen Wu,[1] Joanne Volponi,[1] and Seema Singh[1]

[1]Biomass Science and Conversion Technology Department, Sandia National Laboratories, Livermore, California; and [2] National Center for X-Ray Microscopy, Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California

Biofuels generated from algae hold great promise as a potentially significant source of renewable, sustainable, and low-carbon transportation energy. One class of algal biofuel that has received significant attention is biodiesel obtained from the transesterification of triacylglycerols (TAGs) isolated from algae. Although certain types of algae are known to accumulate a significant amount of these TAGs under specific conditions, there is still a lack of fundamental understanding as to the mechanism and process by which these lipids are sequestered within algae. We will present our results obtained using soft X-ray tomography to visualize whole algal cells at different stages of lipid accumulation. Using this technique, we have developed three-dimensional whole-cell images that highlight the structure and placement of these lipid pools within the algae. The knowledge gained by this technique provides new insight into the mechanism and extent of lipid accumulation within algae and is the foundation for subsequent work by us in the development of optimized algal strains for biouel production.

**Tuesday, February 9, 3:55 p.m.–4:20 p.m.**
**Breakout Session F: Structural Biology Applications in Genomic Sciences**

## Real-Time Monitoring of Chemical Environment in Cells During Stress-Adaptive Response

**Hoi-Ying N. Holman**,[1,2*] Eleanor Wozei,[1,4] Luis R. Comolli,[1] David. A. Ball,[1] Sharon Borglin,[1,2] Matthew W. Fields,[2,3] Terry C. Hazen,[1,2] and Kenneth H. Downing[1]

[1]Lawrence Berkeley National Laboratory, Berkeley, California; [2]Virtual Institute for Microbial Stress and Survival (http://vimss.lbl.gov); [3]Montana State University, Bozeman, Montana; [4]Current address: Makerere University, Kampala, Uganda

The transient chemical properties of the intracellular environment can elucidate the paths through which a biological system adapts to changes in its environment. For example, which mechanisms enable some obligate anaerobic bacteria to actually survive a sudden exposure to oxygen? Here we use high-resolution synchrotron radiation-based Fourier Transform Infrared (SR-FTIR) spectromicroscopy to continuously follow cellular chemistry within living obligate anaerobes, such as *Desulfovibrio vulgaris* Hildenborough, by monitoring hydrogen bonding in their cellular water. We observed a sequence of well-orchestrated molecular events that correspond to changes in cellular processes in those cells that survive, but we only saw accumulation of radicals in those that do not. We thereby can interpret adaptive response in terms of transient intracellular chemistry and link it to oxygen stress and survival. This ability to monitor chemical changes at the molecular level can yield important insights into a wide range of adaptive responses.

**Tuesday, February 9, 4:20 p.m.–4:45 p.m.**
**Breakout Session F: Structural Biology Applications in Genomic Sciences**

# Changes in Switchgrass Structure During Acid Pretreatment Examined with Small-Angle Neutron Scattering and Wide-Angle X-Ray Diffraction

Sai Venkatesh Pingali,[1] Volker S. Urban,[1] William T. Heller, [1] Hugh M. O'Neill, [1] Dean A. Myles, [1] Joseph McGaughey, [1] Dean A. Myles, [1] Loukas Petridis,[2] Jeremy C. Smith,[2] Marcus Foston,[3] Arthur Ragauskas,[3] **Barbara R. Evans**,[1] and Brian Davison[1]

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

In order to attain cost-effective production of biofuels from lignocellulosic biomass, it is necessary to define the minimal structural changes to plant cell-wall structure that pretreatment must make for subsequent conversion processes to reach maximum efficiency. The structural consequences of dilute acid pretreatment of switchgrass were examined across a wide range of length scales using a combined approach of small-angle neutron scattering (SANS) and wide-angle X-ray diffraction (WAXD). SANS is a powerful tool for studying bulk materials that probes length scales ranging from 10 to 10,000 Å, making it ideally suited to understanding the gross morphological changes in switchgrass resulting from pretreatment. In contrast, WAXD probes distances less than 10 Å, making it possible to also determine the effect of pretreatment on the crystallinity of the elementary cellulose fibrils. The BIO-SANS instrument at the High Flux Isotope Reactor at Oak Ridge National Laboratory and the WAXD station at the Advanced Photon Source (APS) at Argonne National Laboratory were used for these studies. The information from the SANS and WAXD experiments combined with chemical structural information obtained from NMR will enable building of computer models that accurately describe deconstruction of lignocellulosic biomass.

**Wednesday, February 10, 9:00 a.m.–9:30 a.m.**
**Plenary Session**

# Insights into Organelle Biogenesis and Near Molecule Protein Distribution Using Photoactivatable Fluorescent Protein Technology

**Jennifer Lippincott-Schwartz**

Cell Biology and Metabolism Program, National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland

Genetically encoded photoactivatable fluorescent proteins (PA-FP) are invisible at the imaging wavelength until switched on by activation at a different wavelength. After activation there is no fluorescence contributed by newly synthesized, nonactivated PA-FPs. This allows many new applications beyond what is possible with conventional fluorescent proteins. For example, the PA-FPs attached to specific proteins can be used to spatially pulse-label subpopulations of molecules in cells for determining protein traffic patterns and turnover rates. More recently, PA-FPs have been used as probes in the novel high-resolution optical imaging technique called photoactivation localization microscopy (PALM). In PALM, PA-FPs are stochastically activated, localized, and bleached one at a time. A super-resolution image is then constructed by summing all the single-molecule positions. This permits structures labeled by an ensemble of PA-FPs too dense to be imaged simultaneously to be resolved with nanometric precision. Various new applications of PALM will be discussed, including simultaneous, dual-color PALM using PA-Cherry and PA-GFP probes, and PALM combined with single-particle tracking to examine lateral diffusion of membrane proteins and directed transport of cytoskeletal elements in live cells. Finally, an interferometric PALM strategy will be described that utilizes single-photon, simultaneous multiphase interferometry to provide sub-20 nm, 3-D protein localization with genetic specificity. Examples such as these illustrate the value of PA-FP–based, super-resolution imaging for providing quantitative insights into protein organization and dynamics at the nanoscale.

**Wednesday, February 10, 10:00 a.m.–10:40 a.m.**
**Plenary Session: Systems Biology for $H_2$ Production**

# Metabolomic Data Reveals Pathways and Regulatory Principles

**Joshua D. Rabinowitz**
Lewis-Sigler Institute for Integrative Genomics, Princeton University

Microbial metabolism holds promise for clean and cost-effective conversion of sunlight or plant material into fuels such as ethanol, butanol, or hydrogen. The ability to rationally improve such processes relies on understanding metabolic pathways and their regulation. While genome sequencing and homology-based annotation provides a first-pass blueprint of metabolic pathways, many metabolic enzymes remain unannotated, especially in promising bioenergy species distant from canonical model organisms. Even in species as well studied as *Escherichia coli*, regulation of metabolism remains poorly understood.

Metabolomics—the measurement of hundreds of metabolites in parallel using mass spectrometry and/or nuclear magnetic resonance—provides a new window on cellular metabolic activity. With use of stable isotope tracers, one can measure not only relative metabolite concentration changes, but also absolute metabolite concentrations and metabolite flows or "fluxes." This information can then be used to drive the development of quantitative models of metabolism and its regulation.

*Clostridium acetobutylicum* is an obligatory anaerobic bacterium of bioenergy interest for its rapid production of hydrogen and butanol. As a prelude to attempting to improve its hydrogen production efficiency, we used isotopic tracers to quantitate its central metabolic fluxes. We found that glucose is converted to pyruvate exclusively by glycolysis, with no Entner-Doudoroff or oxidative pentose phosphate pathway activity. Most interestingly, while genome sequencing failed to reveal homologues of many enzymes of the citric acid cycle, we found a complete, albeit bifurcated, TCA cycle: Oxaloacetate flows to succinate both through citrate/ketoglutarate and via malate/fumarate. These observations enabled us to build a genome-scale model of *C. acetobutylicum* metabolism, including quantitatively accurate fluxes during both exponential growth and solventogenesis. The regulatory mechanisms accounting for these fluxes remain to be determined.

To explore flux regulation, we turned to the model organism *E. coli*. Absolute quantitation of its metabolome revealed that most substrates are present at levels exceeding the $K_m$ of the enzyme consuming them. Other metabolites are also abundant, however, resulting in competition for enzyme active sites between substrate, product, and competitive inhibitors. This "intrinsic" regulation by active site competition plays a key role in flux control and can produce metabolic homeostasis without relying on allostery or other overt regulatory mechanisms. By accounting for both overt and intrinsic regulation, we developed and validated a predictive dynamic model of nitrogen metabolism. Efforts to similarly model carbon metabolism and its links to nitrogen metabolism are ongoing. Such quantitative modeling efforts aim to obtain a sufficiently rigorous understanding of metabolic regulation to enable rational flux control. Development of similar models for bioenergy organisms will provide a valuable tool for optimizing biofuel production.

**Wednesday, February 10, 10:40 a.m.–11:20 a.m.**
**Plenary Session: Systems Biology for $H_2$ Production**

# Protons as Electron Acceptors: What Can This Teach Us About Bioenergy Generation?

**Caroline S. Harwood**

Department of Microbiology, The University of Washington

Hydrogen gas has good potential to be used in conjunction with hydrogen fuel cells as a transportation fuel for automobiles and other vehicles. There are several microbial routes for hydrogen production that depend on the environmentally benign use of biomass-derived feedstocks, solar energy, and other readily available resources. All microbes generate hydrogen gas by using protons as electron acceptors. Cells remove electrons from compounds that they acquire from their environment and then use hydrogenase or nitrogenase enzymes to combine the electrons with protons to make hydrogen gas. We and others have found that hydrogen yields are greatest when microbes are denied access to alternative electron acceptors. Also, the precise route that electrons take from donor to acceptor can influence the hydrogen yield. Finally, closely related strains of bacteria grown under identical conditions can differ markedly in the amount of hydrogen they produce. Approaches that we are developing to associate genomic variation with phenotype to infer causal relationships may reveal new proteins that facilitate the use of protons as electron acceptors.

**Wednesday, February 10, 11:20 a.m.–12:00 noon**
**Plenary Session: Systems Biology for $H_2$ Production**

# Pathway of Fermentative Hydrogen Production by Sulfate-Reducing Bacteria

**Judy D. Wall**,[1,4] Kimberly L Keller,[1,4] Barbara J. Giles,[1] Iris Porat,[2] Timothy J. Tschaplinski,[2] Adam Deutschbauer,[3,4] Adam Arkin,[3,4] and Steve D. Brown[2,4]

[1]University of Missouri, Columbia, Missouri; [2]Oak Ridge National Laboratory; [3]University of California, Berkeley; and [4] Virtual Institute for Microbial Stress and Survival (http://vimss.lbl.gov)

Molecular hydrogen is a candidate biofuel that has a large energy yield and does not pollute. Microbes generate hydrogen through three primary routes: photosynthetically through biophotolysis of water, by nitrogenase activity in photosynthetic organisms, or through fermentation pathways. The latter two processes require electron sources from organic matter or biomass. Among these three production mechanisms, fermentation has received the least enthusiastic interest as a process for biofuel production because the microbial enzymes responsible are hydrogenases that are essentially freely reversible and therefore self-damping. However, fermentative pathways do have an advantage relative to other microbial systems in that large amounts of substrate are oxidized with production of reduced end products during the generation of small amounts of cell material.

Sulfate-reducing anaerobes of the genus *Desulfovibrio* are the focus of our efforts to determine the wiring diagram for electron flow to hydrogen and to examine the complexity of the wiring. With the organic acids—lactate, pyruvate, or fumarate—these bacteria produce a series of reduced end products, among them being small amounts of hydrogen. Analysis of a mutant lacking the type-1 tetraheme cytochrome $c_3$ in *D. desulfuricans* G20 has revealed that this mutant does not reduce sulfate with pyruvate as electron donor. From this observation, we infer that electrons from pyruvate must traverse the cytoplasmic membrane twice, once to reach the cytochrome and a second time to return to the adenylylsulfate reductase responsible for sulfate reduction. Apparently, unlike pyruvate, electrons from lactate bypass that cytochrome and reach sulfate, although a greater share of electrons are diverted to hydrogen. Curiously, the mutant lacking the cytochrome $c_3$ also grows more robustly on pyruvate without sulfate than does the parental strain. Analyses of accumulated end products indicate potential pathways operating during these growth modes and show that respiration and fermentation are occurring simultaneously rather than being mutually exclusive.

Microarray and proteomic analyses of the expressed genes and proteins differentially expressed in the parental and mutant strains growing with sulfate or with pyruvate alone are in remarkable agreement regarding the enzyme systems responsible for some of these changes in growth. In addition, assays of mutants with transposon insertions in the genes identified are being used to confirm the pathways. Models consistent with the observations of electron flow will be presented.