DOE/SC-0140

Section:
Systems Biology for Environmental and Subsurface Microbiology

U.S. DEPARTMENT OF ENERGY
Office of Science

Joint Meeting 2011

Genomic Science Awardee Meeting IX

and

USDA-DOE Plant Feedstock Genomics for Bioenergy Awardee Meeting

[Revised: April 14, 2011]

Crystal City, Virginia
April 10-13, 2011

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

http://genomicscience.energy.gov

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

Systems Biology for Environmental and Subsurface Microbiology

156

Tracking Carbon Flows in a Model Microbial Community Using Genome-Enabled Methods and Stable Isotope Probing

Nicholas Justice[1]* (njustice@berkeley.edu), Chongle Pan[2]* (panc@ornl.gov), Chris Miller[1]* (csmiller@gmail.com), Curt Fischer,[1] Ben Bowen,[3] Itai Sharon,[1] Paul Wilmes,[1] Alexis Yelton,[1] Brett Baker,[1] Daniela Aliaga Goltsman,[1] Steven Singer,[3] Brian Thomas,[1] Robert Hettich,[2] Trent Northen,[3] and **Jill Banfield**[1]

[1]University of California, Berkeley; [2]Oak Ridge National Laboratory; and [3]Lawrence Berkeley National Laboratory

http://geomicrobiology.berkeley.edu/pages/amd.html

Project Goals: The primary goal is to develop integrated omics methods, including stable isotope probing, for tracking carbon flows in microbial communities.

Microbial communities both contribute to carbon fixation and drive the degradative portion of the carbon cycle. However, existing methods are inadequate to systematically examine carbon flows in complex natural communities at the molecular level. The primary goals of our project are to develop an integrated "omics" toolset enabled by community metagenomics and stable isotope probing and to apply this toolset to model microbial communities of tractable complexity. We aim to demonstrate that a comprehensive understanding of the carbon and energy flows in this model community can be achieved. The developed toolset will be broadly applicable to many microbial communities relevant to DOE missions.

The acid mine drainage (AMD) microbial community has been established as a model system for ecological and evolutionary studies and "omics" method development (Denef et al. 2010). Notably, a full carbon cycle is represented in this ecosystem. Organisms in the floating AMD biofilms fix CO_2 using chemical energy from iron oxidation. The biofilm undergoes successional changes, and eventually sinks to the sediment-water interface where it is anaerobically degraded. These communities have been extensively studied using a proteogenomic approach via 5 years of collaborative research funded by DOE. We have reconstructed near-complete genomes of essentially all the bacterial (8 lineages) and archaeal (12 lineages) natural populations consistently detected in the system, as well as one fungal genome (see below) and the genomes of numerous viruses/phage and plasmids. In addition, we have mapped changes in population structure and protein abundance over space, time, and developmental stages.

A specific new target for ongoing research is the degradative portion of the AMD carbon cycle. We evaluated *in situ* change in community structure of sunken biofilms using fluorescent *in situ* hybridization (FISH) and 454 pyrosequencing, and analyzed metabolic activities with proteomics. FISH measurments indicate that the sunken decaying community transitioned from dominance by bacteria to archaea. Intriguingly, the sunken community proteome showed more bacterial proteins than expected based on FISH. We attribute this to persistence of bacterial proteins from dead cells in the degrading biomass. Indeed, many bacterial proteins have suffered amine hydrolysis, likely due to exposure to the acidic solution after cell lysis. Twelve archaeal species identified in the sunken biofilms belong predominantly to the order *Thermoplasmatales* (A through I-plasma), but also include the deeply branched ARMAN Euryarchaeota. Genomic analyses of the archaea indicated that they are likely obligate heterotrophs, but that inorganic electron donors may also play an important role in their metabolism. Genotypic differences indicate varied roles for different archaea in community functioning (Justice *et al.,* in prep. Yelton *et al.,* in prep). To understand the roles of different community members, anaerobic communities were cultivated in the laboratory and analyzed using stable isotope probing, proteomics, and metabolomics.

Fungi often colonize late-developmental-stage biofilms at low relative abundance, and may play an important role in recycling carbon in the community. To test this hypothesis, we have targeted "streamer" biofilms dominated by fungal hyphae for genomics, proteomics, and transcriptomics. By developing methods for assembly of genomes from short-read Illumina community DNA sequencing data, we have reconstructed the near-complete genome (27 Mbp) of the dominant fungal community member, *Acidomyces richmondensis*. Predicting proteins from this genome allowed for identification of thousands of fungal proteins, covering many important metabolic pathways (Miller et al., in prep.). Comparative transcriptomic analysis is also being performed to better characterize the potential differences in fungal metabolism between streamer and floating biofilms.

We have also developed a novel method for characterizing microbial community structure from Illumina metagenomic data. Our approach reconstructs full-length ribosomal small subunit genes from short sequencing reads (a previously intractable problem), and provides accurate estimates of relative taxon abundances (Miller et al., in review). Together with FISH, we have used the method to characterize the fungal community. Fungal streamer communities do not share the same community structure as floating biofilms: the typical primary producer is only found at low abundance. Yet neither do they share identical community structure with heterotrophic sunken biofilms; while *Thermoplasmatales*, ARMAN, and *Sulfobacillus* all are abundant, fungi which are absent in sunken biofilms still dominate the community.

Further proteomic and transcriptomics data should reveal if fungal-dominated biofilms represent a distinct degradative community in the carbon cycle.

Carbon flows in the community involve exchange and metabolism of small molecules that can be characterized using metabolomics. Identification of the chemical formula and structure for a detected-spectral feature in an untargeted metabolomics experiment is facilitated by the high resolution, sensitivity, and mass accuracy achieved by liquid chromatography-electrospray ionization quantitative time of flight mass spectrometry (LC-ESI-qTOF-MS). However, often the chemical identity is difficult to discern from mass-information alone. As an assist, heuristically filtered chemical formula generators and isotopic pattern fitting is utilized. Formulas and biological origin of the metabolites was then confirmed using stable isotope labeling. However, this approach does not provide structural information. A novel proton labeling technique was developed and applied to define the location of exchangeable (H/D) hydrogens and elucidate the most abundant metabolic features detected in a model mixed-species microbial community revealed the importance of a new class of lipids (Fischer et al., in review; Bowen et al. in preparation, and in press).

Stable isotopic labeling is also used to link metabolic activities to community structure. Nitrogen or carbon flows from ^{15}N- or ^{13}C-enriched substrates into the biomass of a microbial community can be traced by measuring the incorporation of these stable isotopes into biomarkers such as lipids, nucleic acids, and proteins. Existing stable isotope probing (SIP) methods can only measure a few selected lipids, DNA sequences or proteins. We developed a proteomic SIP method that can determine ^{13}C or ^{15}N atom% of thousands of identified proteins from multiple strains and species in the AMD community. The proteomic SIP method was validated using the AMD communities grown in laboratory at three known ^{15}N atom%: 0.4% ^{15}N (natural abundance), ~50% ^{15}N, or ~98% ^{15}N. 1408~2326 proteins were identified from each sample with false discovery rates of 0.4%~2%. The median ^{15}N atom% from each sample has less than 1% deviation from the experimentally imposed ^{15}N in the validation samples. The ^{15}N atom% was also measured with high precision. 95%~99% of identified proteins have a ^{15}N atom% estimated within the 5% range of the expected atom%. This indicates that our method can automatically identify the sequence of, and quantify the degree of heavy atom enrichment in, thousands of proteins from microbial community proteome samples. The method was used to monitor incorporation of ^{15}N into established and regrowing microbial biofilms. 1814~2407 non-redundant proteins were identified from four samples in a time course. Most proteins were either unlabeled (<5% ^{15}N atom %) or highly labeled (>80% ^{15}N atom %). The abundance ratio between two isotopologues of a protein was estimated using the ProRata algorithm. The different labeling patterns among microbial members of the community revealed distinct colonization behavior and metabolic activities among the microorganisms (Pan et al. in press).

Overall, in the first year of the project we have made considerable progress toward deploying metagenomic analyses coupled to isotope-enabled proteomic and metabolomic studies to track carbon and nitrogen flows in microbial communities.

157

Wiring Microbes to the Sun: Mechanisms for Energy Conservation During Microbial Electrosynthesis

Kelly P. Nevin* (knevin@microbio.umass.edu), Trevor L. Woodard, Sarah A. Hensley, Ashley E. Franks, Zarath M. Summers, Jianhong Ou, Oona L. Snoeyenbos-West, Muktak Aklujkar, Pier-Luc Tremblay, Ching Leang, and **Derek R. Lovley**

Department of Microbiology, University of Massachusetts, Amherst

http://www.geobacter.org

Project Goals: The primary goals of this project are to develop genome-based experimental approaches to elucidate the function of anaerobic microbial communities involved in groundwater bioremediation and novel bioenergy processes and to develop genome-scale models of these communities in order to better understand and optimize bioremediation and bioenergy applications.

Microbial electrosynthesis is a novel bioenergy approach in which microorganisms consume electrons derived from electrodes to convert carbon dioxide to multi-carbon fuels or other desirable chemicals. When powered with solar technology microbial electrosynthesis has the same net reaction as plant-based photosynthesis, using solar energy to convert carbon dioxide and water to organic compounds and oxygen. However, microbial electrosynthesis has a number of potential benefits over biomass-based strategies for fuel production. These include: 1) solar technology is 100-fold more effective than photosynthesis in harvesting solar energy; 2) microbial electrosynthesis produces fuels directly whereas only a fraction of the energy harvested in biomass is converted to fuel; 3) microbial electrosynthesis eliminates the need for arable land for converting solar energy to fuel; 4) microbial electrosynthesis conserves water and avoids the pollution of water resources associated with biomass production. Thus, microbial electrosynthesis represents a new approach to photosynthesis that can convert solar energy to fuels more efficiently and with less environmental degradation than biomass-based strategies.

Previous studies in our laboratory demonstrated the concept of microbial electrosynthesis with the acetogenic microorganism *Sporomusa ovata*. *S. ovata* colonized graphite cathodes and used electrons derived from the cathodes to reduce carbon dioxide to acetate. Electron consumption and appearance of electrons in acetate were consistent with

the reduction of carbon dioxide to acetate with the cathode serving as the sole electron donor.

In order to better understand the mechanisms for microbial electrosynthesis, the possibility of catalyzing this process with other acetogenic microorganisms was evaluated, including several strains for which the complete genome sequence is available and/or strains that are genetically tractable. Additional *Sporomusa* strains were capable of acetate electrosynthesis, as were *Clostridium ljungdahlii*, *Clostridium aceticum*, and *Moorella thermoacetica*. The acetogen *Acetobacterium woodii* was unable to consume current.

Comparison of predicted mechanisms of energy conservation in the various acetogens suggested a mechanism for microorganisms to conserve energy for cell maintenance during electrosynthesis. Those acetogens capable of electrosynthesis all have proton-dependent ATPases. When carbon dioxide is reduced to acetate with electrons derived from extracellular electron transfer there is a net consumption of protons within the cytoplasm generating a proton-gradient across the inner membrane. Proton flux through proton-dependent ATPases to alleviate the proton imbalance can generate ATP. This mechanism of energy conservation would not be possible in *A. woodii*, which has a sodium-dependent ATPase. Thus, the model is consistent with experimental results.

Surprisingly, *Geobacter metallireducens* consumed current as well as the most active acetogenic microorganisms with carbon dioxide serving as the sole electron donor. *G. sulfurreducens*, which readily reduces a variety of other electron acceptors with an electrode as the sole electron donor, could not reduce carbon dioxide. Analysis of the *G. metallireducens* genome suggested two potential pathways by which *G. metallireducens* might reduce carbon dioxide to acetate. One of these is a reductive TCA cycle and the other is the recently discovered dicarboxylate/4-hydroxybutyrate cycle. The key enzyme of this pathway, 4-hydroxybutyryl-CoA dehydratase, is not found in *Geobacteraceae* other than *G. metallireducens*. The existence of two potential pathways for carbon dioxide reduction in *G. metallireducens* is promising for the development of strains engineered to produce a variety of organic compounds via electrosynthesis. The potential role of the two pathways in electrosynthesis is currently being evaluated with genetic approaches.

The genome of *G. metallireducens* contains genes for a wide diversity of redox-active proteins that could function as electrical contacts between cells and electrodes and for electron transfer into the cytoplasm. Effective strategies for gene deletion and expression of heterologous genes have recently been developed for *G. metallireducens*, making investigations of electron transfer mechanisms feasible. Examination of the genomes of *C. ljungdahlii* and *M. thermoacetica*, the two genetically tractable acetogens capable of electrosynthesis, has demonstrated that acetogens are likely to have much different strategies for electron transfer for electrosynthesis than *G. metallireducens*. We have recently significantly improved techniques for the genetic manipulation of *C. ljungdahlii* to facilitate investigation of electron transfer mechanisms in this gram-positive organism. Elucidation of these pathways is key to optimization of this potentially transformative bioenergy technology.

158

Genomic Patterns of Amino Acid Usage in Fe(III)-Reducing, Sulfur-Reducing and Syntrophic Geobacteraceae

Muktak A. Aklujkar* (muktak@microbio.umass.edu) and **Derek R. Lovley**

Department of Microbiology, University of Massachusetts, Amherst

http://www.geobacter.org

Project Goals: The primary goals of this project are to develop genome-based experimental approaches to elucidate the function of anaerobic microbial communities involved in groundwater bioremediation and novel bioenergy processes and to develop genome-scale models of these communities in order better to understand and optimize bioremediation and bioenergy applications.

The family Geobacteraceae within the Deltaproteobacteria includes Fe(III)-respiring Geobacter species, sulfur-reducing Desulfuromonas species, and syntrophic Pelobacter species. Several genomes of Geobacteraceae have now been sequenced, leading to the discovery of intriguing differences not only of physiology and metabolism but even at the level of amino acid usage.

We have already demonstrated that the inhibitory effect of a short nucleotide sequence on histidyl-tRNA synthetase in the *Pelobacter carbinolicus* lineage could account for the loss of ancestral genes of respiratory metabolism, in which multiple closely spaced histidine codons may cause translation to stall. For that study, a histidine demand index, defined as the number of histidine codons divided by the harmonic mean distance between them, was computed for every gene in the genomes of *P. carbinolicus*, *Desulfuromonas acetoxidans*, *Geobacter bemidjiensis*, *Geobacter metallireducens*, and *Geobacter sulfurreducens*. The same analysis has now been performed on the genomes of *Pelobacter propionicus*, *Geobacter daltonii*, *Geobacter lovleyi*, *Geobacter* sp. M18, *Geobacter* sp. M21, and *Geobacter uraniireducens*, revealing that high-histidine-demand coding sequences are fewer in *P. propionicus* as well as *P. carbinolicus*, despite similar frequencies of histidine usage overall.

Similar analyses were performed for each of the other nineteen amino acids, in an attempt to identify patterns that distinguish the *Geobacter* clade that includes *P. propionicus* from the *Desulfuromonas* clade that includes *P. carbinolicus*, or that distinguish the two syntrophic *Pelobacter* species from their relatives that respire Fe(III) or sulfur. One of the most striking results is that genomic glutamine demand is lower in *G. sulfurreducens* and *G. metallireducens* compared

to subsurface *Geobacter* species, higher in *G. lovleyi*, and extraordinarily high in *D. acetoxidans*, suggesting that the common ancestor of *Geobacteraceae* had a high-glutamine-demand proteome that has shifted to ever lower glutamine demand in its Fe(III)-respiring descendants. Both *Pelobacter* species have much lower genomic glutamine demand than their closest relatives. Genomic glutamine demand can be correlated with the frequency of glutamine usage overall. Further investigations are ongoing to determine what high-glutamine-demand proteins have been acquired or lost, which orthologous proteins have variable glutamine content, and whether any protein sequence motifs are linked to mutations of glutamine. Metabolic pathways that involve glutamine are also being compared across the *Geobacteraceae*.

Lower genomic cysteine demand distinguishes *D. acetoxidans*, *P. carbinolicus*, *P. propionicus* and *G. lovleyi* from the majority of *Geobacter* species. Although several of the high-cysteine-demand genes of *Geobacter* species encode multiheme *c*-type cytochromes, this does not fully account for the difference, as the *D. acetoxidans* genome encodes 80 *c*-type cytochromes. Scrutiny of the remaining high-cysteine-demand genes may lead to the identification of redox-active proteins unique to *Geobacter* species and potentially important for respiration of Fe(III).

The median lengths of protein-coding sequences in the genomes of *Geobacter* species are almost all in the range from 283 to 287 codons, with *G. metallireducens* as the sole outlier at 294 codons. Interestingly, *G. bemidjiensis* is at the low end despite having more copies of rRNA and tRNA genes than *G. metallireducens*. *D. acetoxidans* has a median protein-coding sequence length of 289 codons. In contrast, the median lengths for *P. propionicus* (276 codons) and *P. carbinolicus* (275 codons) suggest that shorter proteins may be characteristic of the physiology of syntrophs, perhaps because respiration of extracellular electron acceptors requires more proteins to be made with signal peptides.

In conclusion, this study shows that the genomes of *Geobacteraceae* have unique amino acid clustering patterns that may reflect differences in their environment, physiology and metabolism as well as their phylogeny.

159

Environmental Functional Genomics: Elucidating the Function of Highly Expressed Proteins During Subsurface Uranium Bioremediation

Jiae Yun, Toshiyuki Ueki* (tueki@microbio.umass.edu), and **Derek R. Lovley**

Department of Microbiology, University of Massachusetts, Amherst

http://www.geobacter.org

Project Goals: The primary goal of this project is to develop genome-based experimental approaches to elucidate the function of anaerobic microbial communities involved in groundwater bioremediation and novel bioenergy processes and to develop genome-scale models of these communities in order to better understand and optimize bioremediation and bioenergy applications.

One of the common findings in metagenomic, metatranscriptomic, and environmental proteomic studies is that a function can not be assigned to the vast majority of the diverse sequences recovered. This severely limits interpretation of the environmental data. Therefore, tools for experimentally elucidating the function of genes recovered in environmental studies need to be developed.

For example, extensive sequencing of subsurface mRNA collected during *in situ* bioremediation of uranium-contaminated groundwater at the DOE study site in Rifle, CO, as well as proteomic analysis of similar samples, has identified many genes that are actively expressed during the bioremediation process. Most of these genes can be assigned to *Geobacter* species. However, lack of information on the function of these genes has stymied progress in diagnosing the physiological status of the subsurface *Geobacter* species and their activities.

Analysis of proteins in the groundwater during active bioremediation revealed that one of the most abundant proteins stained positively for heme, suggesting that it was a *c*-type cytochrome. A peptide sequence in the protein, identified with mass spectrometry, was identical to that in a putative *c*-type cytochrome in the genome sequence of *Geobacter* strain M18, an isolate from the Rifle site. A homologous sequence was found in the genomes of multiple *Geobacter* species in Subsurface Clade I, a cluster of closely related *Geobacter* species that typically predominate in subsurface environments. Analysis of the subsurface proteome with an antibody designed to recognize the conserved portion of this protein in Subsurface Clade I *Geobacter* species confirmed its presence and also demonstrated that the abundance of the cytochrome was positively correlated with the success of uranium removal from the groundwater.

The lack of a developed genetic system for any of the Subsurface Clade I *Geobacter* species prevented evaluation of the cytochrome function in the available pure cultures and a homologous gene was not found in the genetically tractable *Geobacter sulfurreducens*. However, the function of the cytochrome could be evaluated by testing the ability of the gene to complement strains of *G. sulfurreducens* in which other cytochrome genes were deleted. OmcS is an abundant cytochrome in *G. sulfurreducens* growing on Fe(III) oxide. It is aligned along the conductive pili of *G. sulfurreducens* and is thought to facilitate electron transfer from the pili to Fe(III) oxides. Expression of the subsurface cytochrome gene in a *G. sulfurreducens* strain in which the gene for OmcS was missing restored the capacity for Fe(III) oxide reduction, suggesting that the function of the subsurface cytochrome gene is similar to that previously described for OmcS.

These studies demonstrate that, with the appropriate experimental design, it is possible to elucidate the function of abundant environmental proteins, even when the protein is part of a complex process, such as extracellular electron transfer. Additional functional genomic studies of other genes of unknown function that are highly expressed during bioremediation of uranium-contaminated groundwater are underway.

160

Functional Characterization of Aromatic-Solute Binding Proteins in Environmental Organisms

Ruth Pietri,[1]* Sarah Zerbs,[2] Danielle Corgliano,[2] **Frank Collart,**[2] and Lisa M. Miller[1]

[1]National Synchrotron Light Source, Brookhaven National Laboratory, Upton, N.Y.; and [2]Biosciences Division, Argonne National Laboratory, Lemont, Ill.

Project Goals: see below

This project addresses the *Genomic Science Program* mission by identifying the functional characteristics of proteins from key microorganisms and plants that are known to respond to complex environmental conditions. In particular, the program focuses on proteins that facilitate communication between the cell and the environment such as ABC transporters from microbial organisms living in soil. The goal is to characterize these proteins using genomic, biophysical and computational approaches and to develop model systems to address fundamental environmental concerns.

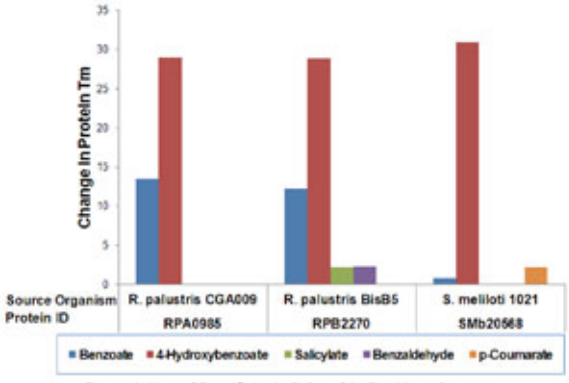

Figure 1. Ligand Specificity in Solute-binding Homologues

In this study we evaluated the ABC transporters from several soil organisms with an emphasis on proteins involved in the utilization of lignin-degradation products. The analysis set includes protein encoded in the genomes of *Rhodopseudomonas palustris, Bradyrhizobium japonicum*, and *Sinorhizobium meliloti*. These microbes share common features in transporter, sensor and metabolic characteristics but the individual strains and organisms were isolated from diverse ecological niches. They metabolize organic compounds derived from lignin-degradation products and have the potential for bioremediation of xenobiotic compounds. We hypothesize the ABC transporters profile for these aromatic compounds are predictive for the metabolic pathways and regulatory networks associated with utilization lignin degradation products. To test this hypothesis, we screened and characterized a set of transporter proteins from soil organisms using fluorescence thermal shift assay (FTS), isothermal calorimetry (ITC), circular dichroism spectroscopy (CD), infrared spectroscopy (FTIR), and small/wide angle x-ray scattering techniques (SAXS/WAXS).The FTS assay is used as a preliminary screen to profile the aromatic compound binding specificity of proteins in soil organisms known to utilize compounds derived from lignin degradation (Figure 1). The kinetic and thermodynamic properties of these proteins, obtained by ITC, showed that the proteins exhibit high affinity for the aromatic substrates with dissociation constants (K_d) in the submicromolar range, similar to other transporter binding proteins. Moreover, analysis of the flanking genomic regions reveals the co-localization of these transporter genes with metabolic genes associated with utilization of the transported compounds. The profile and number of transport proteins specific for aromatic compounds is consistent with ecological and laboratory studies, which demonstrate the capabilities of these organisms for the utilization of plant degradation products such as lignin-derived aromatic compounds.

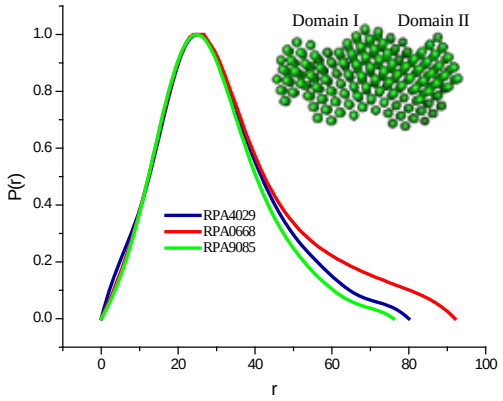

Figure 2. Electron distribution functions of transporter proteins from *R. palustris*. The structure predicted by SAXS/WAXS in shown in the upper right corner.

The specificity of three aromatic-solute binding transporters was further assessed by structural analysis using CD, FTIR, and SAXS/WAXS. CD and FTIR provide complementary information about protein secondary structure, and were therefore employed to determine the secondary motifs of the proteins in the presence and absence of aromatic ligands. SAXS/WAXS on the other hand, are capable of providing information about the size and shape of proteins and were also used to define the three dimensional fold of the transporters in the presence and absence of ligands. CD and FTIR analyses showed that the proteins share similar secondary structures with a mixture of α-helical and β-sheet motifs. The results also demonstrated that the proteins retain their native secondary structures in the presence of aromatic ligands. Similarly, SAXS/WAXS analysis

showed that the proteins have analogous topologies, which remain unaffected in the presence of aromatic ligands. The overall shape of the proteins, derived from their electron distribution functions (Figure 2), suggests that they have ellipsoid-like structure with two distinct domains. Indeed, theoretical structural modeling of the proteins predicted a three dimensional fold consisting of two domains connected by a hinge region, which is conserved in the solute-binding proteins of ABC transporters.

Overall, the FTS assay and ITC results showed that three ABC transporter proteins from *R. palustris* bind aromatic ligands with high affinity. CD, FTIR and SAXS/WAXS data suggested that these proteins have the conserved structure with two mixed α-helix and β-sheet domains connected by a hinge. As with other transporters, binding of the aromatic ligands might induce small movement of amino acid residues in the hinge region. Taken together, the study has shown the capability of indentifying solute-binding proteins that interact with lignin-degradation products using genomic and biophysical approaches, which can be extended to other organisms. These genome-scale studies will enable the identification of other proteins with similar ligand-binding profiles and characteristics and enable refinement of sequence-based methods for extension to other organisms and systems.

The submitted manuscript has been created by UChicago Argonne, LLC, Operator of Argonne National Laboratory ("Argonne"). Argonne, a U.S. Department of Energy Office of Science laboratory, is operated under Contract No. DE-AC02-06CH11357. Brookhaven National Laboratory is operated under Contract No. DE-AC02-98CH10886. This contribution originates from the "Environment Sensing and Response" Scientific Focus Area (SFA) program at Argonne National Laboratory. This research was supported by the U.S. Department of Energy (DOE), Office of Biological and Environmental Research (BER), as part of BER's Genomic Science Program (GSP).

161

Molecular Approaches for Elucidation of Sensory and Response Pathways in Cells

Sarah Zerbs,[1]* Michael Endres,[1] Elizabeth Landorf,[1] William Studier,[2] and **Frank R. Collart**[1]

[1]Biosciences Division, Argonne National Laboratory, Lemont, Ill.; and [2]Biology Division, Brookhaven National Lab, Upton, N.Y.

Project Goals: see below

Increased knowledge of protein function enhances our understanding of cellular functions and is ultimately required to model biological activities and systems. This program addresses the hypothesis that cellular behavior can be modeled through an understanding of the biological interface with the environment and the cellular responses that originate from the cell/environment interaction. The long term objective of the program is to define cellular sensory and regulatory pathways that respond to environmental nutrients thereby facilitating a system-level model that predicts the cellular response to environmental conditions or changes. One aspect of this program will develop tools to bridge the gap between genomes and systems biology. In this context we are developing methods for experimental validation of cellular sensory and regulatory networks by experimental validation of the function of two component sensors and transcription factors.

To improve prospects for identification of effectors for bacterial transcriptional regulators, we are evaluating *in vitro* ligand mapping strategies to identify environmental molecules or their metabolic derivatives that bind to transcription regulatory proteins and identify the DNA binding regions for these proteins. One approach maps effectors to TRs by assessment of increased stability attributable to the TR-ligand interaction. Effector binding confers increased thermal stability which can be used as a surrogate to infer binding specificity. This feasibility of this approach was validated with a set of *Escherichia coli* TRs with known effector molecule binding specificity and the method then extended to identify effector ligands for a set of TR from *Shewanella oneidensis*.

In a complimentary approach, we are applying biolayer interferometry to characterize the interaction of TR with DNA binging regions and effector molecules. The *forté*BIO Octet Red instrument is designed to enable a label free approach for the detection of interactions between biological molecules. The instrument uses an array of fiber-optic sensors that detect biomolecular interactions via biolayer interferometry. This instrument can be used for the identification of protein-protein and protein-ligand interactions using relatively small amounts of material in a microwell plate format. Identification of regulatory ligands coupled with knowledge of the DNA-binding regions of the transcription factors allows the association of metabolic pathways with the regulatory network. The functional assignments and ability to define specific sensory and regulatory pathways will increase the predictive capability of current models and support the development of predictive systems-level models. This increased knowledge of the molecular components and control features of cellular sensory and response pathways is essential for our understanding of natural biological processes related to carbon management, sustainability and bioenergy.

The submitted manuscript has been created by UChicago Argonne, LLC, Operator of Argonne National Laboratory ("Argonne"). Argonne, a U.S. Department of Energy Office of Science laboratory, is operated under Contract No. DE-AC02-06CH11357. Brookhaven National Laboratory is operated under Contract No. DE-AC02-98CH10886. This contribution originates from the "Environment Sensing and Response" Scientific Focus Area (SFA) program at Argonne National Laboratory. This research was supported by the U.S. Department of Energy, Office of Biological and Environmental Research (BER), as part of BER's Genomic Science Program.

162
Molecular Processes of Mycorrhizal Symbiosis

Peter E. Larsen,[1]* Geetika Trivedi,[2] Avinash Sreedasyam,[2] Leland Cseke,[2] and **Frank R. Collart**[1]

[1]Biosciences Division, Argonne National Laboratory, Lemont, Ill.; and [2]Department of Biological Sciences, University of Alabama, Huntsville

Project Goals: see below

This fungal-plant symbiosis is widespread and a process of major ecological importance. Establishment of this symbiosis involves a progressive series of complex developmental steps accompanied by radical changes in metabolism and plant/fungal interactions with the environment. The scientific goal of this theme is the delineation of the molecular events associated with alteration of the nutrient assimilation in the plant and fungal organisms.

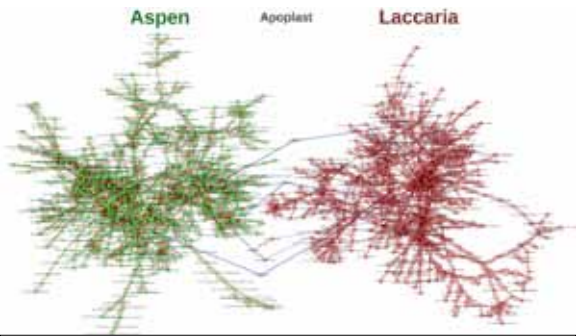

Figure 1. This model of mycorrhizal metabolome is predicted from transcriptomic NGS analysis. Edges are specific enzyme activities detected as statistically significantly expressed in the mycorrhizal transcriptome. Nodes are metabolites inferred by detected enzyme activities. Green edges are metabolic activities expressed in aspen root, red edges are metabolic activities expressed in Laccaria, blue edges are the action of metabolite transport predicted from detected, expressed mycorrhizal transporter genes, and blue dashed edges are predicted metabolite diffusion across cellular membranes. Light green nodes are metabolites predicted to be present in aspen roots, pink nodes are metabolites predicted to be present in Laccaria, and grey nodes are metabolites predicted to be present in mycorrhizal apoplast.

The molecular elements in these signaling and metabolic networks are identified by merging Next Generation Sequencing (NGS) transcriptomic analysis of plant roots and associated soil microorganism communities with established databases of metabolic networks, previously archived transcriptomic data, and experimentally-verified protein interactions. The transcriptomic data was used to identify statistically significantly expressed gene models using a bootstrap-style approach, and these expressed genes were mapped to specific metabolic pathways. This mycorrhizal metabolome (Figure 1) is comprised of the expressed metabolic enzymes in the mycorrhizal transcriptome and the transporters required for the exchange of metabolic compounds between these two organisms. Analysis of the models suggests they are consistent with experimental environmental data and provide insight into the molecular exchange processes for organisms in this complex ecosystem. For example, in mycorrhizal symbiosis, aspen exchanges with *L. bicolor* photosynthetic sugars for nutrients. This expectation is validated by the statistically enriched expression of sugar porters by *L. bicolor*. The enrichment of KEGG amino acid metabolism pathways with unique expressed enzyme activities and the enrichment for expressed amino acid transporters for both aspen and *L. bicolor* indicate that, for mycorrhizae formed in WPM, *L. bicolor's* debt to aspen for carbon is paid with organic nitrogen. *L. bicolor* expresses the metabolic capacity to synthesize nitrogenous compounds such as glycine, glutamate and allantoin, via pathways not expressed in aspen roots. In the growth conditions used here, the predicted exchange compounds are the fructose and glucose as well as organic nitrogen compounds, specifically glycine, glutamate and perhaps allantoin. The predictions suggest *L. bicolor* is an active metabolic partner in addition to passively extending the absorptive surface of aspen roots. This role encompasses uptake of ammonium from the medium and synthesis of more complex compounds provided to the plant. Additional experiments in different nutrient environments are expected to uncover additional mechanisms of mycorrhizal metabolic interactions.

These predictions conform closely to prior experimental observations in other plant species and alternate forms of symbiotic relationships between plant roots and soil microorganisms. By merging transcriptomic data with genomic annotation and previously published metabolic pathway information, knowledge of what fraction of an organism's metabolic capacity is being expressed during symbiotic interaction provides insights into the mycorrhizal metabolome that a single source of data could not. The model-based predictions lead directly to specific, testable biological hypotheses and target particular expressed proteins and transporters for molecular characterization. Additional experiments in different nutrient environments will uncover additional mechanisms of mycorrhizal metabolic interactions. Our available and interactive model of the mycorrhizal metabolome will serve as an important resource for other investigators.

The submitted manuscript has been created by UChicago Argonne, LLC, Operator of Argonne National Laboratory ("Argonne"). Argonne, a U.S. Department of Energy Office of Science laboratory, is operated under Contract No. DE-AC02-06CH11357. This contribution originates from the "Environment Sensing and Response" Scientific Focus Area (SFA) program at Argonne National Laboratory. This research was supported by the U.S. Department of Energy (DOE), Office of Biological and Environmental Research (BER), as part of BER's Genomic Science Program (GSP).