GTL Milestone 2

Develop Methods and Concepts Needed to Achieve a Systems-Level Understanding of Microbial Cell and Community Function, Regulation, and Dynamics

Section 1

OMICS: Systems Measurements of Plants, Microbes, and Communities

54 GTL

The Virtual Institute of Microbial Stress and Survival: An Overview of the Environmental Stress Pathway Project

Carl Abulencia,[4,14] Eric J. Alm,[4,12] Gary Anderson,[1,4] Edward Baidoo,[1,4,7] Peter Benke,[1,4,7] Sharon Borglin,[1,4] Eoin L. Brodie,[1,4] Romy Chakraborty,[1,4] Swapnil Chhabra,[1,4] Gabriela Chirica,[2,4] Dylan Chivian,[1,4] Michael J. Cipriano,[1,4] M.E. Clark,[4,9] Paramvir S. Dehal,[1,4] Elliot C. Drury,[4,8] Inna Dubchak,[1,4,5] Dwayne A. Elias,[4,8] Matthew W. Fields,[4,9] J. Gabster,[2,4] Sara P. Gaucher,[3,4] Jil Geller,[1,4] B. Giles,[4,8] Masood Hadi,[3,4] Terry C. Hazen,[1,4] Qiang He,[4,11] Zhili He,[4,10] Christopher L. Hemme,[4,10] E. Hendrickson,[2,4] Kristina L. Hillesland,[2,4] Hoi-Ying Holman,[1,4] Katherine H. Huang,[1,4] Y. Wayne Huang,[1,4] C. Hwang,[4,9] Janet Jacobsen,[1,4] Marcin P. Joachimiak,[1,4] Dominique C. Joyner,[1,4] Jay D. Keasling,[1,4,7] Keith Keller,[1,4] Martin Keller,[4,6] J. Leigh,[2,4] T. Lie,[2,4] Aindrila Mukhopadhyay,[1,4] Richard Phan,[1,4] Francesco Pingitore,[1,4,7] Morgan Price,[1,4] Alyssa M. Redding,[1,4,7] Joseph A. Ringbauer Jr.,[4,8] Rajat Sapra,[3,4] Christopher W. Schadt,[4,6] Amy Shutkin,[1,4] Anup K. Singh,[3,4] David A. Stahl,[2,4] Sergey M. Stolyar,[2,4] Yinjie Tang,[1,4] Joy D. Van Nostrand,[4,10] Chris B. Walker,[2,4] Judy D. Wall,[4,8] Eleanor Wozei,[1,4] Zamin K. Yang,[4,6] Huei-Che Yen,[4,8] Grant Zane,[4,8] Aifen Zhou,[4,10] Jizhong Zhou,[4,10] and **Adam P. Arkin**[1,4,7,13*] (aparkin@lbl.gov)

[1]Lawrence Berkeley National Laboratory, Berkeley, California; [2]University of Washington, Seattle, Washington; [3]Sandia National Laboratories, Livermore, California; [4]Virtual Institute for Microbial Stress and Survival, http://vimss.lbl.gov; [5]DOE Joint Genome Institute, Walnut Creek, California; [6]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee; [7]University of California, Berkeley, California; [8]Department of Biochemistry, University of Missouri, Columbia, Missouri; [9]Department of Microbiology, Miami University, Oxford, Ohio; [10]Institute for Environmental Genomics, University of Oklahoma, Norman, Oklahoma; [11]Civil and Environmental Engineering, Temple University, Philadelphia, Pennsylvania; [12]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts; [13]Howard Hughes Medical Institute, Chevy Chase, Maryland; and [14]Diversa, Inc., San Diego, California

Project Goals: Environmental Stress Pathway Project (ESPP) is developing computational models that describe and predict the behavior of gene regulatory networks in microbes in response to the environmental conditions found in DOE waste sites. The research takes place within the Virtual Institute for Microbial Stress and Survival (VIMSS). Based at Lawrence Berkeley National Laboratory (LBNL), VIMSS supports an integrated and multi-institutional program to understand the ability of bacteria and other microorganisms to respond to and survive external stresses. VIMSS was established in 2002 with funding from the U.S. Department of Energy Genomics:GTL Program for Rapid Deduction of Stress Response Pathways in Metal/ Radionuclide Reducing Bacteria. LBNL is operated by the University of California for the U.S. Department of Energy.

The Virtual Institute of Microbial Stress and Survival (VIMSS, http://vimss.lbl.gov) was established in 2002 through DOE Genomics:GTL funding of the Environmental Stress Pathway Project (ESPP). The mission of VIMSS is to create a coordinated core community of scientists and a supporting experimental and computational infrastructure for the large scale comparative systems biological study of microbes and the processes they perform in different environments.

As the founding project, the central goal of ESPP is to help build this VIMSS organization and apply the resources for the deduction of regulatory pathways and systems that impact the ability of bacteria to reduce metals and radionuclides in contaminated soil. To this end, over the past four and half years the ESPP team has built capabilities in environmental chemical and microbial monitoring and perturbation; environmental microbe isolation and characterization; controlled biomass production for anaerobic microbes and co-cultures of mixed microbes; phenotypic characterization of microbes; physiological imaging of microbes; development of genetic systems for environmental microbes; transcript, protein and metabolite profiling of single microorganisms and simple co-culture; and computational analysis of these diverse data types in comparative genomic context. With these tools at hand ESPP has focused on uncovering the pathways that the sulfate reducing bacterium *Desulfovibrio vulgaris* employs to respond to environmental cues, scavenge energy from the sulfate, reduce metals and interact with other members of its community. We have learned a great deal about the evolutionary biology of these pathways and how this has impacted the organization of its genome. We have been able to compare the physiology of *D. vulgaris* to other environmental microbes to discover how regulatory strategies may be tuned to deal with different environmental niches. This work has resulted in over 73 publications to date. To aid outside researchers to access this data and other results of the VIMSS team, ESPP has built the Experimental Information and Data Repository (EIDR, http://vimss.lbl.gov/EIDR/) and the MicrobesOnline comparative functional genomics Workbench (http://www.microbesonline.org/) which also incorporate community tools so other researchers can annotate and add data.

Since its inception with ESPP, VIMSS has grown to house or formally collaborate with seven large scale projects (three of which are other Genomics:GTL projects) and informally supports several others in genome annotation, data analysis, and use of the VIMSS experimental pipeline.

* Presenting author

55 GTL

VIMSS ESPP Functional Genomics Core: Cell Wide Analysis of Metal-Reducing Bacteria

Aindrila Mukhopadhyay,[1,6]* Edward Baidoo,[1,6] Peter Benke,[1,6] Swapnil Chhabra,[1,6] Gabriela Chirica,[2,6] Elliot Drury,[3,6] Matthew Fields[5,6] (fieldsmw@muohio.edu), Sara Gaucher,[2,6] Masood Hadi,[2,6] Qiang He,[4,6] Zhili He,[4,6] Chris Hemme,[4,6] **Jay Keasling**[1,6] (keasling@.berkeley.edu), Francesco Pingitore,[1,6] Alyssa Redding,[1,6] Rajat Sapra,[2,6] Anup Singh[2,6] (aksingh@sandia.gov), Yinjie Tang,[1,6] Judy Wall[3,6] (wallj@missouri.edu), Huei-Che Yen,[3,6] Grant Zane,[3,6] Aifen Zhou,[4,6] and Jizhong Zhou[4,6] (jzhou@rccc.ou.edu)

[1]Lawrence Berkeley National Laboratory, Berkeley, California; [2]Sandia National Laboratories, Livermore, California; [3]University of Missouri, Columbia, Missouri; [4]University of Oklahoma, Norman, Oklahoma; [5]Miami University, Oxford, Ohio; and [6]Virtual Institute for Microbial Stress and Survival (VIMSS; vimss.lbl.gov)

Project Goals: Environmental Stress Pathway Project (ESPP) is developing computational models that describe and predict the behavior of gene regulatory networks in microbes in response to the environmental conditions found in DOE waste sites. The research takes place within the Virtual Institute for Microbial Stress and Survival (VIMSS). Based at Lawrence Berkeley National Laboratory (LBNL), VIMSS supports an integrated and multi-institutional program to understand the ability of bacteria and other microorganisms to respond to and survive external stresses. VIMSS was established in 2002 with funding from the U.S. Department of Energy Genomics:GTL Program for Rapid Deduction of Stress Response Pathways in Metal/Radionuclide Reducing Bacteria. LBNL is operated by the University of California for the U.S. Department of Energy.

Heavy metal and radionuclide waste poses a serious problem as a source of environmental contamination. The discovery of dissimilatory sulfate and metal ion-reducing bacteria in such ecosystems presents the opportunity to develop effective biocontainment strategies. Such protocols have the potential to be cost-effective and involve limited physical perturbation towards the environment. *Desulfovibrio vulgaris* Hildenborough is an anaerobic mesophile that belongs to the sulfate-reducing class of bacteria found ubiquitously in nature. While historically studied because of its role in bio-corrosion of oil and gas pipelines, *D. vulgaris* has the ability to reduce a wide variety metals and radionuclides. A rigorous understanding of *D. vulgaris* physiology and its ability to survive in its environment will be critical to discern the biogeochemistry at metal contaminated sites, for bioremediation and natural attenuation for toxic metals. The availability of an annotated genomic sequence provides the foundation for such studies conducted in the functional genomics core of the VIMSS Environmental Stress Pathway Project (ESPP).

Utilizing our optimized pipeline for generating biomass for functional genomics studies, we conducted transcriptomics analysis to measure *D. vulgaris* response to mutations in important global regulators such as Fur, PerR and Zur. The response of these mutants to environmental stresses was also examined. Additionally we continued to collect data in *Shewanella oneidensis* and *Geobacter metallireducens* for comparative genomics. More importantly we extended transcriptomics analysis to examine alternate *D. vulgaris* physiological states such as in biofilms and growth in syntrophic co-culture with *Methanococcus maripaludis*. In order to address cell wide responses beyond the initial changes at the mRNA level, a quantitative proteomics workflow using the iTRAQ peptide tagging

strategy was optimized to provide proteins change measurements. This method provides data for 30% of the proteome and covers all known protein functional categories. Cell wide proteomics and transcriptomics measurements are now being used to develop models for several of these environmentally relevant stresses such as oxygen exposure, oxidative stress, as well as the altered physiology of biofilms and co-culture. Continued studies to map cell wide responses have also emphasized the importance of changes that require orthologous measurements – such as monitoring post translational modifications and protein-protein interactions. With this in mind a novel protocol to monitor redox state of the proteins has been developed as well exogenous protocols to study protein-protein interactions. Measurements at the metabolite and metabolic pathways level are necessary to complete our cell wide studies. Metabolite extraction and detection for several hundred metabolites can now be conducted for these non-model organisms using high resolution separation and mass spectroscopic methods. Flux analysis methods were established using ^{13}C-Lactate grown *D. vulgaris* and *S. oneidensis*. The use of a novel high resolution FTICR–MS method for flux analysis led to results that allowed us to re-evaluate the genome annotation for the central metabolic pathways in *D. vulgaris*. The integration of these methods enables a comprehensive understanding of *D. vulgaris* physiology.

56 GTL

Response of *Desulfovibrio vulgaris* Hildenborough to Acid pH

H.-C. Yen,[1,7]* T. C. Hazen,[2,7] Z. Yang,[3,7] J. Zhou,[4,7] K. H. Huang,[5,7] E. J. Alm,[5,7] **A. P. Arkin**[6,7] (aparkin@lbl.gov), and J. D. Wall[1,7]

[1]Department of Biochemistry, University of Missouri, Columbia, Missouri; [2]Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California; [3]Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee; [4]Department of Botany and Microbiology, University of Oklahoma, Norman, Oklahoma; [5]Massachusetts Institute of Technology, Cambridge, Massachusetts; [6]Department of Bioengineering, University of California, Berkeley, California; [7]Virtual Institute for Microbial Stress and Survival, http://vimss.lbl.gov/

Project Goals: The DOE oversees 350 cleanup projects involving soil contaminated with metals or radionuclides. The life-cycle cost of these projects is $220 billion over 70 years and could be $300 billion without breakthroughs. One breakthrough approach may be to exploit bacteria that can immobilize and detoxify metals in soil via reduction to less soluble and less toxic forms. This occurs naturally and be stimulated in situ. A thorough understanding of the biogeochemistry, especially stress responses in metal/radionuclide bacteria, will enable prediction of natural attenuation and new strategies for remediation that could save DOE billions in cleanup, risk assessment, and environmental stewardship. To achieve this understanding we will study three such organisms that occupy different niches at these sites by developing validated culture conditions similar to site conditions, and then using functional genomic and comparative genomic analysis to deduce the molecular basis of the stress responses that affect metal reduction efficiency. Models of the physiological processes will be made of accuracy suitable for computation development of field protocols for stimulating metal reduction. In addition, comparison of pathways among niches will give insight into how the bacteria adapt their pathways to different conditions.

The anaerobic, sulfate-reducing bacterium *Desulfovibrio vulgaris* Hildenborough has been suggested to be useful for environmental bioremediation applications. Among the constraints likely to be encountered, that effect the growth of a bacterial community in contaminated sites, are low pH con-

* Presenting author

ditions. Growth of *D. vulgaris* on acid pH media was studied. It was evident that this bacterium was able to grow at pHs as low as 5 in batch cultures; however, the lag phase was prolonged and the final protein yield was proportionally lower with the decrease in pH. In medium with lactate as the carbon and reductant source and sulfate as the terminal electron acceptor, the final protein yields dropped to 50% or 29% of the control (pH 7) when the initial medium pH was pH 5.5 or pH 5.0, respectively. The average lag phase for initial pH 5.5 medium was 120 hours versus pH 7 which generally had two hours or less.

This bacterium incompletely oxidizes organic acids with the production of acetate. Thus at low pH, growth is limited by the accumulation of acetate in the medium that acts to shuttle protons across the membrane thereby dissipating the proton gradient and acidifying the cytoplasm. The production of sulfide gas through sulfate reduction consumes protons thereby countering the acid conditions. Finally deamination or decarboxylation of certain amino acids, e.g. lysine, arginine, tryptophan, and isoleucine, may gradually alkalize the acid medium. When the medium pH is high enough (~6.5) for lactate oxidation and sulfate reduction, cell growth follows.

Transcriptional profiling was performed for *D. vulgaris* exposed to pH 5.5 and 6.2. Typical acid shock responses for Gram-negative bacteria were not observed. Instead, from the transcripts with the highest differential expression, it was inferred that significant cellular damage had resulted because chaperone genes including heat shock genes and repair genes for both proteins and nucleic acids were greatly induced.

57 GTL

Global Gene Regulation in *Desulfovibrio vulgaris* Hildenborough

Aifen Zhou,[1,5]* Zhili He,[1,5] Chris Hemme,[1,5] Aindrila Mukhopadhyay,[2,5] Jay Keasling,[2,5] **Adam P. Arkin**[2,5] (aparkin@lbl.gov), Terry C. Hazen,[2,5] Judy D. Wall,[4,5] and Jizhong Zhou[1,5]

[1]Institute for Environmental Genomics, Department of Botany and Microbiology, University of Oklahoma, Norman, Oklahoma; [2]Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California; [3]Departments of Biochemistry and Molecular Microbiology & Immunology, University of Missouri, Columbia, Missouri; [4]Earth Science Division, Lawrence Berkeley National Laboratory, Berkeley, California; and [5]Virtual Institute for Microbial Stress and Survival, http://vimss.lbl.gov

Project Goals: Environmental Stress Pathway Project (ESPP) is developing computational models that describe and predict the behavior of gene regulatory networks in microbes in response to the environmental conditions found in DOE waste sites. The research takes place within the Virtual Institute for Microbial Stress and Survival (VIMSS). Based at Lawrence Berkeley National Laboratory (LBNL), VIMSS supports an integrated and multi-institutional program to understand the ability of bacteria and other microorganisms to respond to and survive external stresses. VIMSS was established in 2002 with funding from the U.S. Department of Energy Genomics:GTL Program for Rapid Deduction of Stress Response Pathways in Metal/ Radionuclide Reducing Bacteria. LBNL is operated by the University of California for the U.S. Department of Energy.

Desulfovibrio vulgaris Hildenborough (DvH) is an obligate anaerobe and has been used as a model organism for studying the energy metabolism of sulfate-reducing bacterium (SRB). However, experimental data about the transcriptional regulatory networks which are essential for understanding the cellular processes are very limited. One of the central goals of ESPP is to link laboratory measurements of stress responses and metabolism to activities of microbes in the field. Thus, towards this goal, we have performed laboratory study of gene expression and regulations in D. vulgaris in responses to oxidative stress and the importance of key global regulatory genes in stress responses. Several predicted global regulators are investigated via mutant characterization, transcriptomic assay and their *in vivo* gene regulations using ChIP-chip assay.

CRP/FNR. CRP/FNR regulators are DNA binding proteins function as positive transcription factors. There are four CRP/FNR homologues in the DvH genome (DVU2547, DVU0379, DVU3111, DVU2097). Evidence from other bacteria demonstrated that CRP/FNR regulators function in response to a broad spectrum of intracellular and exogenous signals such as oxidative and nitrosative stress, nitric oxide, carbon monoxide or temperature. Microarray data from DvH shows that their transcript levels are altered in response to nitrate, nitrite, heat shock, and oxygen stresses. To determine the function to the DvH CRP/FNR, knockout mutants for all four CRP/FNR proteins were generated. The mutants will be characterized using various electron donors and acceptors, different stressors, and transcriptomic analysis. To study the global gene regulation by CRP/FNR, recombinant proteins for all four CRP/FNR were obtained and polyclonal antibodies were generated. Immunoprecipitated DNA-protein complexes with specific CRP/FNR polyclonal antibodies will be hybridized to the DvH PCR-amplicon promoter array. And the CRP/FNR binding motif can be identified by computational and experimental approaches.

H_2O_2 stress response. Oxidative stress is one of the most common environmental stressors. Evidences show that DvH cells are aero-tolerant although they are strict anaerobe. But little is known about molecular mechanisms of oxidative stress responses. DvH is one of the few microorganisms that contain both defense systems which are typical for the aerobic (Sod and Kat) and the anaerobic (Rub, Rbr, Rbo etc.) microbes, but their roles remain elusive. In this study, DvH cells were stressed with two different concentrations of H_2O_2 (1 mM, 4 mM) and 5 time-points (30, 60, 120, 240 and 480 min) were used for the transcriptomic analysis. Microarray data demonstrated that higher concentration of H_2O_2 had broader effect on gene expression. The time-points with the greatest gene expression changes are 120 min (485 up and 527 down) and 240 min (750 up and 753 down) for 1 mM and 4 mM of H_2O_2 respectively. Rdl, Rbr2 were up-regulated, which suggest that these two proteins, rather than Rub-Rbo & Rbr suggested by Coulter's *in vitro* experiment data, may play major roles in H_2O_2 stress. Genes in the predicted PerR and FUR regulon were also up-regulated. Some interesting candidates such as DVU3269 (a hybrid histidine kinase (HK)), DVU3136 (a nitroreductase family protein) etc. were significantly up-regulated., The function of the putative candidates in H_2O_2 stress response are going to be confirmed by other approaches such as knockout mutant analysis.

FUR, PerR and ZUR. FUR, PerR and ZUR are Fur Paralogs in the DvH genome. Microarray data show that they are involved in iron acquisition, acid shock response and oxidative stress etc. The functional analysis of these three global transcription regulators is in progress.

* Presenting author

58 GTL

Desulfovibrio vulgaris Responses to Hexavalent Chromium at the Community, Population, and Cellular Levels

A. Klonowska,[1]* Z. He,[2,6] Q. He,[3,6] M.E. Clark,[1,6] S.B. Thieman,[1] T.C. Hazen,[4,6] E.L. Brodie,[4,6] R. Chakraborty,[4,6] E.J. Alm,[4,6] B. Giles,[5,6] H.-Y. Holman,[4,6] **A.P. Arkin**[4,6] (aparkin@lbl.gov), J.D. Wall,[5,6] J. Zhou,[2,6] and M.W. Fields[1,6]

[1]Department of Microbiology, Miami University, Oxford, Ohio; [2]Institute for Environmental Genomics, University of Oklahoma, Norman, Oklahoma; [3]Civil and Environmental Engineering, Temple University, Philadelphia, Pennsylvania; [4]Lawrence Berkeley National Laboratory, Berkeley, California; [5]Department of Biochemistry, University of Missouri, Columbia, Missouri; [6]Virtual Institute for Microbial Stress and Survival, http://vimss.lbl.gov/

Project Goals: Environmental Stress Pathway Project (ESPP) is developing computational models that describe and predict the behavior of gene regulatory networks in microbes in response to the environmental conditions found in DOE waste sites. The research takes place within the Virtual Institute for Microbial Stress and Survival (VIMSS). Based at Lawrence Berkeley National Laboratory (LBNL), VIMSS supports an integrated and multi-institutional program to understand the ability of bacteria and other microorganisms to respond to and survive external stresses. VIMSS was established in 2002 with funding from the U.S. Department of Energy Genomics:GTL Program for Rapid Deduction of Stress Response Pathways in Metal/ Radionuclide Reducing Bacteria. LBNL is operated by the University of California for the U.S. Department of Energy.

Desulfovibrio vulgaris is an anaerobic sulfate-reducing bacterium (SRB) able to reduce toxic heavy metals such as chromium and uranium, and *D. vulgaris* represents a useful SRB model for the bioremediation of heavy metal contamination. In order to correlate cellular responses to ecosystem changes with respect to heavy metal bioremediation, measurements are needed at different ecological scale. The VIMSS group has developed a multi-institutional collaboration to better understand the relationships between field site observations, cellular responses, and biochemical processes. The current work describes the observation of *Desulfovibrio* populations at a Cr-contaminated field site within the context of possible mechanisms for Cr bio-reduction. In a field scale trial of polylactate stimulated chromate bioreduction at Hanford, *Desulfovibrio* were rapidly enriched and remained at elevated densities for at least one year. Importantly this enrichment of *Desulfovibrio* corresponded to decreased chromate concentrations which remained below background concentrations during that time. Although much work has focused on Cr and U reduction via individual enzymes, less is known about the cellular response to heavy metal stress in *Desulfovibrio* species. Cells were cultivated in a defined medium with lactate and sulfate, and a sub-lethal concentration of Cr(VI) was added at mid-exponential phase growth. Based upon microarray data, the FMN-dependent nitroreductase might reduce Cr(VI) directly or reduce a Cr-complex. The FMN reductase could synthesize $FMNH_2$ and the NADP dehydrogenase might be used to regenerate $NADPH_2$. The *chr*AB genes on the mega-plasmid most likely play a key role in Cr(III) efflux based upon microarray data and growth data. Additional toxicological effects could be occurring once the Cr(III) is produced via protein denaturation in the cytoplasm, periplasm, and outer cell proper. Growth data with washed cells showed an increased sensitivity in the presence of Cr(VI). When exponential-phase cells were washed to remove hydrogen sulfide carry-over and inoculated into fresh medium with different levels of Cr(VI), lag time increased as the levels of Cr increased. Cells lagged approximately 5, 40, and 55 h in the pres-

* Presenting author

ence of 20, 50, and 100 μM Cr, respectively. When cells were transferred to 50 μM Cr, Cr(VI) levels declined within 2 h and lactate was consumed, but sulfate did not decline until growth was initiated approximately 40 h later. Lactate continued to be consumed at a slow rate during the lag but sulfate levels remained unchanged. The results indicated that lactate oxidation was decoupled from sulfate reduction in the presence of Cr(VI). These data coincide with the working model that electrons are re-routed away from sulfate reduction and used to reduce Cr(VI). The down-expression of sulfate permease coincided with these results. The results also suggested that cells responded to the presence of Cr even after reduction based upon growth responses, FTIR analyses on cellular macromolecules, and expressed genes detected with transcriptomics.

59 GTL

Energy Conserving Hydrogenases Drive Syntrophic Growth of *Desulfovibrio vulgaris* and *Methanococcus maripaludis*

C.B. Walker,[1,6]* Z.K. Yang,[2,6] Z. He,[3,6] S.S. Stolyar,[1,6] J. Jacobsen,[4,6] J.A. Ringbauer Jr.,[5,6] J.D. Wall,[5,6] J. Zhou,[3,6] **A. P. Arkin**[4,6] (aparkin@lbl.gov), and D.A. Stahl[1,6]

[1]University of Washington, Seattle, Washington; [2]Oak Ridge National Laboratory, Oak Ridge, Tennessee; [3]University of Oklahoma, Norman, Oklahoma; [4]Lawrence Berkeley National Laboratory, Berkeley, California; [5]University of Missouri, Columbia, Missouri; and [6]Virtual Institute for Microbial Stress and Survival, http://vimss.lbl.gov

Project Goals: Environmental Stress Pathway Project (ESPP) is developing computational models that describe and predict the behavior of gene regulatory networks in microbes in response to the environmental conditions found in DOE waste sites. The research takes place within the Virtual Institute for Microbial Stress and Survival (VIMSS). Based at Lawrence Berkeley National Laboratory (LBNL), VIMSS supports an integrated and multi-institutional program to understand the ability of bacteria and other microorganisms to respond to and survive external stresses. VIMSS was established in 2002 with funding from the U.S. Department of Energy Genomics:GTL Program for Rapid Deduction of Stress Response Pathways in Metal/Radionuclide Reducing Bacteria. LBNL is operated by the University of California for the U.S. Department of Energy.

In the absence of an electron acceptor, many *Desulfovibrio* species grow on non-fermentable substrates via syntrophic association with hydrogen consuming methanogens. Building upon the ongoing Virtual Institute for Microbial Stress and Survival (VIMSS) investigation into the response of *Desulfovibrio vulgaris* Hildenborough to environmental stressors found in contaminated DOE sites, the Environmental Stress Pathway Project's (ESPP) Applied Environmental Core (AEC) developed and maintained a stable syntrophic consortium. *Desulfovibrio vulgaris* Hildenborough and *Methanococcus maripaludis* LL were continuously grown in a chemostat on minimal media amended with lactate but lacking electron acceptor. Replicated whole genome transcriptional analyses by the ESPP Functional Genomics Core (FGC) and the Computational Core (CC) identified 169 and 254 genes that were significantly up-regulated or down-regulated, respectively, relative to a sulfate-limited monoculture growing at the same generation time. The majority of up-regulated genes were associated with energy production/conservation, signal transduction mechanisms, and amino acid transport/metabolism. A number of the down-regulated genes were associated with signal transduc-

* Presenting author

tion mechanisms, inorganic ion transport/metabolism and amino acid transport and metabolism. Among those genes most highly up-regulated were a suite of hydrogenases including the putative carbon-monoxide induced hydrogenase (Coo, DVU2286 - 93). Coo is a multi-subunit membrane-bound complex with high similarity to an energy conserving protein in *Rhodospirillum rubrum*. In order to further elucidate the possible role energy conserving hydrogenases play in syntrophic growth, we examined transposon mutants generated by the FGC of both the Coo hydrogenase (*cooL*) and a structurally related homolog, Ech (energy conserving hydrogenase, *echA*, DVU0429-34). Both mutants grew to the same cell density on lactate-sulfate, although the *cooL* mutant grew significantly slower. When grown in coculture with *M. maripaludis* without any sulfate, the *cooL* mutant grew significantly slower and to approximately 25% yield, while the *echA* mutant showed a less pronounced difference in growth rate and yield (approximately 80%). Together these data suggest an important role for the Coo hydrogenase in energy conservation of *D. vulgaris* Hildenborough during syntrophic growth, possibly through proton translocation, although the exact physiological mechanism remains to be elucidated. Continued collaborative work by the VIMSS three ESPP core groups should provide a more complete mechanistic understanding of sulfate-reduction and syntrophic cooperation between microbes.

60 GTL

A Large Number of Hypothetical Proteins are Differentially Expressed during Stress in *Desulfovibrio vulgaris*

Elliot C. Drury,[1,5]* Alyssa M. Redding,[2,5] Aindrila Mukhopadyay,[2,5] Katherine H. Huang,[3,5] Terry C. Hazen,[2,5] **Adam P. Arkin**[2,4,5] (aparkin@lbl.gov), Judy D. Wall,[1,5] and Dwayne A. Elias[1,5]*

[1]Department of Biochemistry, University of Missouri, Columbia, Missouri; [2]Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California; [3]Massachusetts Institute of Technology, Cambridge, Massachusetts; [4]Department of Bioengineering, University of California, Berkeley, California; and [5]Virtual Institute for Microbial Stress and Survival, http://vimss.lbl.gov

Project Goals: Environmental Stress Pathway Project (ESPP) is developing computational models that describe and predict the behavior of gene regulatory networks in microbes in response to the environmental conditions found in DOE waste sites. The research takes place within the Virtual Institute for Microbial Stress and Survival (VIMSS). Based at Lawrence Berkeley National Laboratory (LBNL), VIMSS supports an integrated and multi-institutional program to understand the ability of bacteria and other microorganisms to respond to and survive external stresses. VIMSS was established in 2002 with funding from the U.S. Department of Energy Genomics:GTL Program for Rapid Deduction of Stress Response Pathways in Metal/Radionuclide Reducing Bacteria. LBNL is operated by the University of California for the U.S. Department of Energy.

Hypothetical and conserved hypothetical proteins consistently make up 30% or more of sequenced bacterial genomes, with few reports confirming their expression at either the rRNA or protein level. It is likely that many of these proteins serve significant functions ranging from regulation to presently unknown steps in carbon or electron flux pathways. Hence, the elucidation of their function(s) is highly relevant to the Virtual Institute Microbial Stress and Survival/Environmental Stress Pathway Project mission. We have compiled expression profiles for the expected 1167 hypothetical

and conserved hypothetical proteins in *D. vulgaris* from data obtained in the VIMSS/ESPP project over 10 environmental stresses, along with corresponding transcriptomic and MS-based iTRAC proteomic datasets from controlled cultures.

The genes were divided into two groups; those in polycistronic operons and those that are monocistronic. For operonic and monocistronic genes respectively, we observed 37 and 46 genes that are not expressed, 36 and 12 that show no stress-related response but are expressed at high rates, 0 and 173 that show no stress-related response but are expressed at low rates, 445 and 199 with significant response in two or more stresses, and 104 and 123 that only showed significant differential expression in one stress. While the number of transcription studies outweighs protein analysis, the abundance values indicating differential expression at the protein level were highly consistent with the microarray results when data were available.

Therefore, we are presently able to confirm, at both the mRNA and protein level, the expression of 253 hypothetical and conserved hypothetical proteins and show no evidence for 83 genes encoding a protein. Those that are expressed should be re-annotated to "expressed protein" while the remainder should be described as "non-coding gene". Of those that did show expression, elucidating function without a stress-related expression pattern is difficult, particularly for monocistronic genes in this category. The proteins that showed differential expression in response to one or more stresses are theoretically easier to deduce a putative function, especially for those in operons, and these assignments have been completed. Finally, the validity of such putative assignments can only be ascertained by interruption or deletion of the gene with further analysis. We are in the process of testing six mutants that have been isolated from a random transposon library. As the library continues to be sequenced, we will test the interrupted hypothetical and conserved hypothetical proteins either to assign a function or to confirm the putative assignments. By confidently confirming the function of these proteins and the effect of their removal in *D. vulgaris*, there will be a more through understanding of the mechanisms by which this bacterium survives stresses likely experienced at DOE contaminated sites.

61 GTL

Phenotypic Correlations in *Desulfovibrio*

K.L Hillesland,* C.B. Walker, and D.A. Stahl

University of Washington, Seattle, Washington and Virtual Institute for Microbial Stress and Survival, http://vimss.lbl.gov

Project Goals: Environmental Stress Pathway Project (ESPP) is developing computational models that describe and predict the behavior of gene regulatory networks in microbes in response to the environmental conditions found in DOE waste sites. The research takes place within the Virtual Institute for Microbial Stress and Survival (VIMSS). Based at Lawrence Berkeley National Laboratory (LBNL), VIMSS supports an integrated and multi-institutional program to understand the ability of bacteria and other microorganisms to respond to and survive external stresses. VIMSS was established in 2002 with funding from the U.S. Department of Energy Genomics:GTL Program for Rapid Deduction of Stress Response Pathways in Metal/Radionuclide Reducing Bacteria. LBNL is operated by the University of California for the U.S. Department of Energy.

* Presenting author

An ESPP goal is to distinguish genetic pathways that evolved as an adaptation to stressors from those that arose simply through inheritance from ancestral species. To do so, it is necessary to characterize genetic relationships that are conserved across broader taxonomic groups. Phenotypic correlations that are evident within a genus are likely to result from conserved pleiotropic relationships among traits (i.e. both traits share at least some genes in common). Such relationships may accelerate or limit evolutionary adaptation depending on whether they are positive or negative, respectively. Thus, identifying pleiotoropy may lead to new insights into the evolutionary trajectories of a taxon as well as its physiology. To expand our knowledge of the physiology and evolution of the genus *Desulfovibrio*, we measured several features of growth of 14 strains in the presence of different electron donors (lactate or pyruvate) and acceptors (sulfate, no electron acceptor, or in coculture with the hydrogenotrophic methanogen, *Methanococcus maripaludis*). We observed a strong positive correlation between growth rate on lactate and pyruvate when sulfate was the electron acceptor ($r = 0.79$, $p = 0.0007$), but not when *M. maripaludis* was the surrogate electron acceptor. However, there was a positive correlation between the biomass achieved on lactate versus pyruvate with *M. maripaludis* as the electron acceptor ($r = 0.8$, $p = 0.0033$). It is possible that the growth rate of *Desulfovibrio* on lactate and pyruvate was also correlated, but variability in growth kinetics of *M. maripaludis* obscured our ability to detect it. The relationships among growth phenotypes on different electron acceptors were not consistent among electron donors. Growth rate on pyruvate with *M. maripaludis* was positively correlated with growth rate on pyruvate and sulfate ($r = 0.8$, $p = 0.0028$) and lactate and sulfate ($r = 0.75$, $p = 0.0076$). These correlations were not evident when lactate was the electron donor. In fact, with lactate as the electron donor, growth rate with sulfate as the electron acceptor was negatively correlated with the level of biomass achieved from growth with *M. maripaludis* as the electron acceptor ($r = -0.77$, $p=0.006$). In addition to these correlations indicating potential pleiotropic relationships, individual strains showed unusual characteristics. For example, among this *Desulfovibrio* study set, only *D. vulgaris* Llanelly was unable to ferment pyruvate. Although this strain was incapable of syntrophic growth on lactate with *M. maripaludis*, syntrophic growth was possible with pyruvate as the electron donor. Together, these results suggest that some genes involved in mechanisms of energy acquisition are conserved among strains that have the capacity for syntrophic growth. The ecological consequences of these relationships between traits will be explored in experiments examining the relative fitness among *Desulfovibrio* growing under different conditions of electron donor and electron acceptor availability.

62 GTL

Nitrate Stress Response in *Desulfovibrio vulgaris* Hildenborough: Whole-Genome Transcriptomics and Proteomics Analyses

Qiang He,[1,8]* Zhili He,[2,8] Wenqiong Chen,[3] Zamin Yang,[4,8] Eric J. Alm,[5,8] Katherine H. Huang,[5,8] Huei-Che Yen,[6,8] Dominque C. Joyner,[7,8] Martin Keller,[3,4,8] **Adam P. Arkin**[5,8] (aparkin@lbl.gov), Terry C. Hazen,[7,8] Judy D. Wall,[6,8] and Jizhong Zhou[2,8]

[1]Department of Civil & Environmental Engineering, Temple University, Philadelphia, Pennsylvania; [2]Institute for Environmental Genomics, Department of Botany and Microbiology, University of Oklahoma, Norman, Oklahoma; [3]Diversa Corporation, San Diego, California; [4]Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee; [5]Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California; [6]Departments of Biochemistry and Molecular Microbiology & Immunology, University of Missouri, Columbia, Missouri; [7]Earth Science Division, Lawrence Berkeley National Laboratory, Berkeley, California; and [8]Virtual Institute for Microbial Stress and Survival, http://vimss.lbl.gov

Project Goals: Environmental Stress Pathway Project (ESPP) is developing computational models that describe and predict the behavior of gene regulatory networks in microbes in response to the environmental conditions found in DOE waste sites. The research takes place within the Virtual Institute for Microbial Stress and Survival (VIMSS). Based at Lawrence Berkeley National Laboratory (LBNL), VIMSS supports an integrated and multi-institutional program to understand the ability of bacteria and other microorganisms to respond to and survive external stresses. VIMSS was established in 2002 with funding from the U.S. Department of Energy Genomics:GTL Program for Rapid Deduction of Stress Response Pathways in Metal/Radionuclide Reducing Bacteria. LBNL is operated by the University of California for the U.S. Department of Energy.

Sulfate reducing bacteria (SRB) are of interest for bioremediation with their ability to reduce and immobilize heavy metals. Nitrate, a common co-contaminant in DOE sites, is suggested to inhibit SRB via nitrite. Previous results indicate that nitrite is indeed inhibitory to the growth of *Desulfovibrio vulgaris*. However, growth inhibition by nitrate alone was also observed. One of the central goals of ESPP is to link laboratory measurements of stress responses and metabolism to activities of microbes in the field. Thus to this end, we have performed laboratory study of expression in D. vulgaris in responses to nitrate stress. In this study, growth and expression responses to various concentrations of nitrate were investigated using the Omnilog phenotype arrays and whole-genome DNA microarrays. Changes in the proteome were examined with 3D-LC followed by MS-MS analysis.

Microarray analysis found 5, 50, 115, and 149 genes significantly up-regulated and 36, 113, 205, and 149 down-regulated at 30, 60, 120, and 240 min, respectively. Both transcriptomic and proteomic profiles shared little similarities with those of salt stress, indicating a specific inhibitory mechanism beyond osmotic stress. Many of the genes (~50% at certain time points) with altered expression level were of unknown functions; however, the increasing number of ribosomal protein genes down-regulated with time could provide a direct explanation to the growth inhibition effect of nitrate. Further, several lines of evidence suggested that the down-regulation of genes coding the ribosomal proteins could be the result of the changes in the energy flow upon nitrate exposure: 1) The down-regulation of genes for the ATPase subunits indicated reduced level of energy generation; 2) the up-regula-

* Presenting author

tion of phage shock protein genes (*pspA* and *pspC*) might indicate a reduced proton motive force; although damages to the cell envelope could also contribute to this outcome; 3) the gene for the hybrid cluster protein, a redox protein with roles in nitrogen metabolism, was highly up-regulated 120 and 240 min following nitrate stress at both transcriptomic and proteomic level, suggesting that nitrate was being actively reduced, shifting reducing equivalents away from normal energy production; 4) genes in the methionine biosynthesis pathway were among the most highly up-regulated genes throughout the experiment, potentially providing a convenient mechanism for the simultaneous disposal of excess sulfur (from sulfate reduction) and nitrogen (from nitrate reduction); 5) One gene cluster consistently among the most up-regulated genes consisted of genes encoding two TRAP dicarboxylate family transporters, a formate acetyltransferase, and a pyruvate formate-lyase activating enzyme, which might be regulated to provide an increased carbon flow to keep pace with demand from amino acids biosynthesis. These observations indicated that the growth inhibition effect of nitrate might be due to energy limitation.

Similar to the observations made during salt stress, the glycine/betaine transporter gene was among genes highly up-regulated, suggesting that $NaNO_3$ also constituted osmotic stress which was relieved by the mechanism of osmoprotectant accumulation. Osmoprotectant accumulation as the major resistance mechanism was further validated by the partial relief of growth inhibition by glycine betaine. It is also noted that, similar to nitrite stress, the ferric iron transporter genes were up-regulated during nitrate stress, suggesting an increased demand for iron. Unlike nitrite stress, however, no other genes in the Fur regulon were co-regulated during nitrate stress, pointing to a yet-to-known regulatory signal.

In conclusion, excess $NaNO_3$ resulted in both osmotic stress and nitrate stress. *D. vulgaris* shifted nitrogen metabolism and energy production in response to nitrate stress. Resistance to osmotic stress was achieved primarily by the transport of osmoprotectant.

63 GTL

Redox Proteomics in *Desulfovibrio vulgaris* Hildenborough: Search for Proteins that Mediate Stress Response via Post-Translational Modification of the Cys Residues

Rajat Sapra,[1,2]* Sara Gaucher,[1,2] Gabriela Chirica,[2] Carrie Kozina,[2] George Buffleben,[2] Richard Phan,[1,3] Dominique Joyner,[1,3] Terry C. Hazen,[1,3] **Adam P. Arkin**[1,3] (aparkin@lbl.gov), and Anup K. Singh[1,2]

[1]Virtual Institute for Microbial Stress and Survival, http://vimss.lbl.gov; [2]Sandia National Laboratories, Livermore, California; and [3]Lawrence Berkeley National Laboratory, Berkeley, California

Project Goals: Environmental Stress Pathway Project (ESPP) is developing computational models that describe and predict the behavior of gene regulatory networks in microbes in response to the environmental conditions found in DOE waste sites. The research takes place within the Virtual Institute for Microbial Stress and Survival (VIMSS). Based at Lawrence Berkeley National Laboratory (LBNL), VIMSS supports an integrated and multi-institutional program to understand the ability of bacteria and other microorganisms to respond to and survive

external stresses. VIMSS was established in 2002 with funding from the U.S. Department of Energy Genomics:GTL Program for Rapid Deduction of Stress Response Pathways in Metal/ Radionuclide Reducing Bacteria. LBNL is operated by the University of California for the U.S. Department of Energy.

Desulfovibrio vulgaris Hildenborough (*D. vulgaris*) is a sulfate reducing bacterium that grows in the absence of oxygen in a reducing environment. From a physiological as well as ecological perspective, anaerobic bacteria have to redox stress in the form of reactive oxygen species (ROS) produced when the cells are exposed to molecular oxygen or chemicals like hydrogen peroxide. To counteract the deleterious effects of ROS, many anaerobic organisms have developed defense systems similar to those found in aerobes. In *D. vulgaris* enzymes that detoxify ROS include superoxide reductase (Sor) which can reduce superoxide to water and rubrerythrin (Rbr), which can reduce H_2O_2 to water without the regeneration of oxygen, a feature that is important for oxygen detoxification in anaerobe. Another mechanism which anaerobic bacteria have evolved to overcome redox stress is using post translational modifications of proteins, specifically the modifications of Cysteine (Cys) residues in proteins, to minimize the deleterious effects of the ROS and redox associated stress.

Cys is one of the most rarely used amino acids in the proteins of most organisms studied so far. Cys containing proteins are key in maintain the reducing environment of the cytosol and alleviating redox stress due to the presence of ROS. Furthermore, Cys residues in redox active proteins mediate redox reactions where transfer of electrons proceeds via thiol-disulfide exchange reactions. Importantly, all of these activities of Cys containing cytosolic enzymes usually depend on the preservation of the reduced state of the cysteine residue(s) involved. These redox active thiol groups of Cys residues can thus be post-translationally modified to form intra- and inter-molecular disulfide bonds or oxidized to sulfinic, sulfenic acid in response to redox stress.

We are investigating the mechanisms by which *D. vulgaris* proteins are post-translationally modified to counter redox stress induced by oxygen and Cr(VI) reduction using a combination of proteomics techniques that leverage the interdisciplinary cores of the VIMSS. D. vulgaris cells are grown in the environmental microbiology lab (Hazen Lab, Lawrence Berkeley National Labs). The protein samples generated are modified to block all free Cys residues with N-ethylmaleimide to reduce the complexity of the sample and to specifically target the modified Cys residues. The samples thus generated are surveyed using a combination of 2D-DIGE and Isotope Coded Affinity Tag (ICAT) proteomics (Singh Lab, Sandia National Labs) to identify post translational modifications and, finally, ITRAQ proteomics is used to identify the relative abundance of the proteins (Keasling Lab, Lawrence Berkeley National Labs). The stress response at the proteome level is compared to the transcriptomics data generated from the same samples by the bioinformatics core (Arkin Lab, Lawrence Berkeley National Labs) to develop a comprehensive picture of stress response and stress-induced post translational modifications of proteins in *D. vulgaris* . Using this approach, we have identified more than 50 proteins and have mapped the Cys residues in the corresponding proteins that have function associated with redox stress in *D. vulgaris*.

* Presenting author

64 GTL

A Survey of Protein Post-Translational Modifications Found in the Sulfate-Reducing Bacterium *Desulfovibrio vulgaris* Hildenborough

Sara P. Gaucher,[1,5]* Alyssa M. Redding,[2,5] Gabriela S. Chirica,[1] Rajat Sapra,[1,5] George M. Buffleben,[1] Carrie Kozina,[1] Aindrila Mukhopadhyay,[2,5] Dominique C. Joyner,[2,5] Jay D. Keasling,[2,5] Terry C. Hazen,[2,5] **Adam P. Arkin**[2,5] (aparkin@lbl.gov), David A. Stahl,[4,5] Judy D. Wall,[3,5] and Anup K. Singh[1,5]

[1]Sandia National Laboratories, Livermore, California; [2]Lawrence Berkeley National Laboratory, Berkeley, California; [3]University of Missouri, Columbia, Missouri; [4]University of Washington, Seattle, Washington; and [5]Virtual Institute for Microbial Stress and Survival, http://vimss.lbl.gov

Project Goals: Environmental Stress Pathway Project (ESPP) is developing computational models that describe and predict the behavior of gene regulatory networks in microbes in response to the environmental conditions found in DOE waste sites. The research takes place within the Virtual Institute for Microbial Stress and Survival (VIMSS). Based at Lawrence Berkeley National Laboratory (LBNL), VIMSS supports an integrated and multi-institutional program to understand the ability of bacteria and other microorganisms to respond to and survive external stresses. VIMSS was established in 2002 with funding from the U.S. Department of Energy Genomics:GTL Program for Rapid Deduction of Stress Response Pathways in Metal/ Radionuclide Reducing Bacteria. LBNL is operated by the University of California for the U.S. Department of Energy.

Sulfate reducing bacteria (SRB), found widely in nature, use sulfate as the terminal electron acceptor in their respiratory cycle, leading to the production of hydrogen sulfide. These bacteria have both ecological and economic importance. SRB play a role in various biogeochemical cycles including the sulfur and carbon cycles. They have a negative economic impact on the oil industry, where their metabolism causes corrosion and clogging of machinery, and fouling of oil wells. However, they have also been shown to reduce and/or immobilize toxic water-soluble metals such as copper (II), chromium (IV) and uranium (VI), and thus are candidates for bioremediation applications.

Desulfovibrio vulgaris Hildenborough (DvH) is a member of the most well studied genus of SRBs. A goal of the Environmental Stress Pathway Project (ESPP) in the Virtual Institute for Microbial Stress and Survival (VIMSS) is to understand the regulatory networks in DvH for applications to bioremediation. One aspect of this is the elucidation of protein post-translational modifications (PTMs) in DvH.

PTMs play various roles in the cell. Some modifications play a role in protein structure, such as lipid anchors or some disulfide bonds. Others are directly involved in regulation of protein function such as phosphorylation and glycosylation. Still others arise through cellular damage such as irreversible oxidation events. Whatever the role these PTMs play, they must be characterized at the protein level because they are not directly coded for in the genome. Furthermore, DvH may be particularly likely to use PTMs as a regulatory mechanism: Evidence for this includes the observation that the DvH genome encodes an abnormal number of histidine kinases. Our goal is to determine the types of protein modifications that arise in DvH and how these modifications affect the ability of DvH to survive or adapt to its environment.

This work leverages the unique resources of the Virtual Institute for Microbial Stress and Survival: Quality controlled biomass produced at LBL (Hazen lab) is used for all proteomic LC/MS/MS measurements at LBL (Keasling lab). Our initial survey of PTMs in DvH was obtained by mining these numerous proteomic LC/MS/MS data sets acquired over the course of ESPP for evidence of modified peptides. Data mining for PTMs is performed at Sandia National Labs. The searched-for modifications were determined based on literature precedence and a genome search for the existence of relevant transferases. To date we have found preliminary evidence for cysteine oxidation, lysine acetylation, and methylation of lysine and arginine. Data mining for additional PTMs is ongoing. Future work will focus on validation of these findings and determining which, if any, of these modifications play a regulatory role in DvH. Validation will require selective isolation of the proteins of interest for further characterization. Here, protein isolation is made possible through the work being performed at LBL and the University of Missouri to generate DvH mutants containing tagged versions of DvH proteins.

65 GTL

The Ech Hydrogenase is Important for Growth of *D vulgaris* with Hydrogen

S.M. Stolyar,[1,3]* J. Wall,[2,3] and D.A. Stahl[1,3]

[1]University of Washington, Seattle, Washington; [2]University of Missouri, Columbia, Missouri; and [3]Virtual Institute for Microbial Stress and Survival, http://vimss.lbl.gov

Project Goals: Environmental Stress Pathway Project (ESPP) is developing computational models that describe and predict the behavior of gene regulatory networks in microbes in response to the environmental conditions found in DOE waste sites. The research takes place within the Virtual Institute for Microbial Stress and Survival (VIMSS). Based at Lawrence Berkeley National Laboratory (LBNL), VIMSS supports an integrated and multi-institutional program to understand the ability of bacteria and other microorganisms to respond to and survive external stresses. VIMSS was established in 2002 with funding from the U.S. Department of Energy Genomics:GTL Program for Rapid Deduction of Stress Response Pathways in Metal/ Radionuclide Reducing Bacteria. LBNL is operated by the University of California for the U.S. Department of Energy.

One objective of the Virtual Institute for Microbial Stress and Survival (VIMSS) and the Environmental Stress Pathway Project (ESPP) is to determine the genetic and physiological basis for cooperative and competitive interactions among environmental microbial populations of relevance to the DOE. The ESPP Applied Environmental Core (AEC) and Functional Genomics Core (FGC) have identified a number of genes that may participate in cooperative interactions between sulfate reducers and methanogens under low sulfate conditions. Specifically, the gram-negative Deltaproteobacterium *D. vulgaris* is able to grow in the absence of an electron acceptor via syntrophic growth with hydrogenotrophic organisms. Despite decades of research, energy conservation in *D. vulgaris* is not well understood. The presence of multiple hydrogenases, including those located in the periplasm in all studied *Desulfovibrio* strains—and the observation that hydrogen is produced and then consumed during growth of *D. vulgaris* Miyazaki with lactate and sulfate (Tsuji&Yagi, 1980)—lead to the formulation of the hydrogen cycling hypothesis as a mechanism for energy

* Presenting author

conservation (Odom & Peck, 1981). The availability of a completed genome sequence of *D. vulgaris* Hildenborough has since revealed genes for at least six different hydrogenases: four periplasmic and two cytoplasmic. Although several have been partially characterized biochemically and genetically, their roles in *D. vulgaris* under different growth conditions is not well understood. We examined the growth and metabolite production of an *echA* (DVU0434) *D vulgaris* Hidenborough mutant under three different growth conditions: i) in medium amended with lactate and sulfate and ii) in medium amended with acetate, hydrogen and sulfate, and iii) in coculture the hydrogenotrophic methanogen *M. maripaludis*, lacking an electron acceptor. On lactate, the mutant demonstrated a comparable growth rate and yield to the wild type strain, but evolved more hydrogen as measured by its accumulation in the headspace during growth in batch culture. In a medium containing 5 mM acetate and an atmosphere of H_2/CO_2 (80:20), growth of the mutant was severely impaired relative to the wild type. A coculture consisting of the mutant strain and a hydrogenotrophic methanogen (*M. maripaludis*) demonstrated only slightly reduced growth rate and biomass relative to the wild type. Although this suggested some role in energy conservation, the more obvious phenotype was its greatly limited growth in monoculture with acetate, hydrogen and sulfate. Thus, the available data suggest that the primary role of the Ech Hydrogenase is oxidation of hydrogen during sulfate respiration, possibly also contributing to the production of reduced ferredoxin required for conversion of Acetyl CoA to pyruvate by pyruvate oxidoreductase, as was previously demonstrated for the homologous hydrogenases in *M. barkeri* and *M. maripaludis* (Meuer et al., 2002; Porat et al., 2006).

66 GTL

Monitoring of Microbial Reduction and Reoxidation Activities in the FRC Sites using a Comprehensive Functional Gene Array

Zhili He,[1,2,6]* Joy D. Van Nostrand,[1,6] Liyou Wu,[1,2] Terry J. Gentry,[2,3] Ye Deng,[1] Christopher W. Schadt,[2,6] Weimin Wu,[4] Jost Liebich,[2] Song C. Chong,[2] Baohua Gu,[2] Phil Jardine,[2] Craig Criddle,[4] David Watson,[2] Terry C. Hazen,[5,6] and Jizhong Zhou[1,2,6]

[1]Institute for Environmental Genomics and Department of Botany and Microbiology, University of Oklahoma, Norman, Oklahoma; [2]Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee; [3]Department of Soil and Crop Sciences, Texas A&M University, College Station, Texas; [4]Department of Civil and Environmental Engineering, Stanford University, Stanford, California; [5]Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California; [6]Virtual Institute for Microbial Stress and Survival, http://vimss.lbl.gov

Project Goals: Environmental Stress Pathway Project (ESPP) is developing computational models that describe and predict the behavior of gene regulatory networks in microbes in response to the environmental conditions found in DOE waste sites. The research takes place within the Virtual Institute for Microbial Stress and Survival (VIMSS). Based at Lawrence Berkeley National Laboratory (LBNL), VIMSS supports an integrated and multi-institutional program to understand the ability of bacteria and other microorganisms to respond to and survive external stresses. VIMSS was established in 2002 with funding from the U.S. Department of Energy Genomics:GTL Program for Rapid Deduction of Stress Response Pathways in Metal/ Radionuclide Reducing Bacteria. LBNL is operated by the University of California for the U.S. Department of Energy.

* Presenting author

A novel comprehensive functional gene microarray, termed GeoChip, has been developed. This array contains 24,243 oligonucleotide (50mer) probes and covers > 10,000 genes in >150 functional groups involved in nitrogen, carbon, sulfur and phosphorus cycling, metal reduction and resistance, and organic contaminant degradation. This array contains 24,243 oligonucleotide (50mer) probes and covers > 10,000 genes in >150 functional groups involved in nitrogen, carbon, sulfur and phosphorus cycling, metal reduction and resistance, and organic contaminant degradation. Due to the nature of functional gene sequences (highly homologous and incomplete), it is extremely challenging to select specific oligonucleotides for some functional genes using routine probe design strategies. To tackle those problems, we used the following strategies: (1) Retrieved sequences were aligned, and only the shared regions of the functional genes were used for probe design; (2) Experimentally established oligonucleotide design criteria and a novel software tool, CommOligo that was specifically developed to deal with highly similar sequences, were used for GeoChip; (3) To detect both divergent and closely related sequences, gene- and group-specific probes were designed; and (4) To increase the confidence of detection, multiple probes for each sequence or each group of sequences were designed.

The developed GeoChip is a powerful generic tool, and can be used: (1) to survey any environmental samples, such as soil, groundwater, sediments, oil fields, deep sea, animal guts, etc; (2) to study biogeochemical processes and functional activities of microbial communities important to human health, agriculture, energy, global climate change, ecosystem management, and environmental cleanup and restoration; (3) to explore direct linkages of microbial genes/populations to ecosystem processes and functions; and (4) to detect functional genes and/or organisms in a particular environment. Here, we present three related studies on the dynamics and stability of microbial genes and associated communities during bioremediation and reoxidation periods at the Oak Ridge Field Research Center (FRC) and Hanford site using the developed array.

First, Geochip was used to track the dynamics of metal-reducing bacteria and associated communities for an *in situ* bioremediation project at the FRC site in Oak Ridge. Samples were taken from different wells after ethanol injections (after day 166). During the uranium reduction period, both FeRB and SRB populations reached their highest levels at Day 212, followed by a gradual decrease over 500 days. Consequently, the uranium in groundwater and sediments was reduced, and the uranium concentrations in the groundwater were significantly correlated with the total abundance of *c*-type cytochrome genes from *Geobacter*-type FeRB and *Desulfovibrio*-type SRB. Mantel test also indicated that there was significant correlation between the differences of uranium concentrations and those of total c-cytochrome gene abundance or *dsrAB* gene abundance. These results suggested that *Geobacter*-type FeRB and SRB played significant roles in reducing uranium to a level below the drinking standard (<30 μg/L).

Second, the developed array was applied to study the processes of reoxidation (a period after microbial reduction) in microbial communities. Samples were taken before, during, and after a reoxidation period, during which air-saturated tap water (9-12 mg L^{-1} DO) was injected into the FBR for a period of 77 d. DO levels in well FW101-2 and FW102-2 increased to 2 and 0.4-0.5 mg L^{-1}, respectively, during the reoxidation period and changes in the relative abundance of functional groups were apparent in both wells. For example, at 40 d post reoxidation, well FW101-2 showed an increase in the relative abundance of genes involved in ammonification, nitrification, and denitrification and a decrease for those associated with cytochromes, methane generation, N fixation, and sulfate reduction. During the post-reoxidation (Day 77), the relative abundance of ammonification, denitrification, and ammonia oxidation genes had returned to pre-oxidation levels in FW101-2, while genes for methane oxidation and nitrate reduction increased and those for nitrite reduction decreased.

* Presenting author

In addition, the developed GeoChip was used to evaluate functional communities at a lactate-fed chromium reduction system at the Hanford site. Samples were taken from different depths within injection and extraction wells. Extraction well samples showed higher numbers of functional genes than the injection well at the same depth. Within the extraction well, abundance decreased with depth. However, the relative abundance of chromium resistance gene increased with depth in this same well.

The developed GeoChip is the most comprehensive functional gene array for environmental studies so far, but due to exponential increases in the numbers of genes and the number of sequences for each gene, we expect to continuously update the array to reflect the sequence information currently available in public databases and personal collections if possible. Thus, we are working on the third generation array, which covers almost three times more gene sequences, and has more features. For example, *gyrB* has been added for phylogenetic analysis. In addition, a software package (including databases) has been developed for sequence retrieval, probe and array design, probe verification, array construction, array data analysis, information storage, and automatic update.

67 GTL

Towards High-Throughput and High Sensitivity Approaches for Uncovering Total Environmental Gene Expression Patterns

Zamin Yang,[1,3] Christopher W. Schadt,[1,3]* Terry Hazen,[2,3] and Martin Keller[1,3]

[1]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee; [2]Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California; and [3]Virtual Institute for Microbial Stress and Survival, http://vimss.lbl.gov/

Project Goals: Environmental Stress Pathway Project (ESPP) is developing computational models that describe and predict the behavior of gene regulatory networks in microbes in response to the environmental conditions found in DOE waste sites. The research takes place within the Virtual Institute for Microbial Stress and Survival (VIMSS). Based at Lawrence Berkeley National Laboratory (LBNL), VIMSS supports an integrated and multi-institutional program to understand the ability of bacteria and other microorganisms to respond to and survive external stresses. VIMSS was established in 2002 with funding from the U.S. Department of Energy Genomics:GTL Program for Rapid Deduction of Stress Response Pathways in Metal/ Radionuclide Reducing Bacteria. LBNL is operated by the University of California for the U.S. Department of Energy.

In recent years, tremendous progress has been made in understanding microbial communities due to emergence of newly developed genomics-based technologies. Current technologies that have been applied to environmental samples for RNA transcriptional profiling include RT-PCR and functional gene microarrays using total RNAs. While these methods have provided considerable insights, they bear significant limitations that prevent their application in a high throughput manner to *de novo* communities. Both methods require background genomic information to allow for design of specific primers and/or microarray probes. Consequently, these methods can only reveal transcription activity of targeted conserved genes that are first surveyed by PCR and sequencing methods, or those that are obtained through comprehensive metagenomic shotgun sequencing. Thus each of these methods

* Presenting author

have high upfront costs in time, effort and materials. To circumvent this limitation, we are developing a method involving direct sequencing cDNA from the environment samples utilizing a high throughput sequence analysis system such as Bio454. These types of tools will be especially useful in understanding the basis for microbial survival under extreme environmental stressors which is a primary goal of the VIMSS:ESPP project.

Since about 80% of total RNA from microbial organisms consists of ribosomal rRNAs it is crucial to first remove rRNAs as completely as possible without degrading mRNA quality and quantity prior to HT sequence based screening. Since bacteria primarily produce mRNAs without a poly-A tail and thus cannot be enriched using oligo-dT methods, as first step to the application of HT sequencing we have compared three different methods to remove rRNAs and enrich mRNAs of *Desulfovibrio vulgaris Hildenborough* samples. The first method utilizes biotin modified oligos complementary to conserved regions in 16S & 23S rRNA and subtractive hybridization with streptavidin-coated magnetic beads. The second uses a commercially available exonulclease that specifically digests rRNAs bearing a 5'monophosphate group. The third method uses two rounds of reverse transcription, where rRNAs are first reverse transcribed with multiple universal primers for 16S & 23S RNAs, subsequently the RNA/DNA hybrids and cDNA are removed by sequential digestion with RNaseH and DNaseI, and the enriched mRNAs are then reverse transcribed using random primers.

We evaluated these three methods by comparing disappearance of the 16S and 23S bands via electrophoresis, and their effect on mRNA quality by analysis of transcription levels of control (total RNA) vs. enriched mRNA as measured by a whole genome microarray. While all three methods were able to significantly enrich mRNA from rRNA, the microarray analysis revealed differences in measured mRNA levels. In control vs. control (unenriched) hybridizations, less than 0.2 % of genome (5 of 3604 total genes) exhibited significant ($P < 0.05$) changes in the levels of their transcripts. Enriched mRNAs from the first two methods generated on average more genes with altered transcript levels compared to untreated total RNA, with 19 genes (0.5%) for the exonuclease method & 74 genes (2%) for subtractive hybridization exhibiting significantly different than controls. In fact, each of these methods appeared to increase the sensitivity of detection, as average signal intensity's corrected for background were 16% higher for the exonuclease method and 113% for the subtractive hybridization method. Microarray comparisons for the third method are currently under analysis. After completing this initial evaluation, we will use each of these methods to construct cDNA libraries for HT sequencing with the 454 to further optimize and validate this approach in single species, as well as make comparisons of HT sequence based methods with existing microarrays. Subsequent validation and application of the developed methods will be performed on mixed cultures (*Desulfovibrio & Methanococcus*) and incorporation of amplification steps. The developed tools will then be deployed to understand microbial survival in stressed environmental systems.

* Presenting author

68 GTL

Experimental and Computational Approaches to Enhance Proteomics Measurements of Natural Microbial Communities

Nathan VerBerkmoes,[1]* Mark Lefsrud,[1] Chongle Pan,[1] Brian Erickson,[1] Manesh Shah,[1] Chris Jeans,[2] Steven Singer,[2] Michael P. Thelen,[2] Vincent Denef,[3] I. Lo,[3] **Jillian Banfield**[3] (jill@eps.berkeley.edu), and Robert L. Hettich[1]

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

Project Goals: The interdisciplinary research in our GTL project is motivated by the desire to understand how microbial communities assemble, the interplay amongst community members, and the mechanisms of microbial adaptation and evolution. Chemoautotrophic acid mine drainage (AMD) microbial biofilms have proven particularly tractable for these analyses due to their comparatively low species richness. The goal of this subproject is to develop experimental and computational approaches for the comprehensive characterization of the proteome of the AMD system to investigate the nature of the gene expression and conservation amongst the various microbial members of this consortium. Proteomic information will be integrated with genomic and biochemical datasets to help elucidate the structure and activity of microbial communities in their natural environmental context.

Microorganisms comprise the majority of extant life forms and play key roles in a wide variety of health and environmental processes, yet little is known about the nature and driving forces of their diversification. Most research to date has focused on physiological and genomic characterization of a relatively small number of isolated microbial species maintained in monoculture. While providing crucial insights to connect genes and function, these studies are unable to capture some aspects of the organism's behavior in its natural environment. Perhaps the greatest knowledge gap is in the understanding of how microorganisms function within natural multi-species consortia. Recently, genomic characterization has been extended to natural microbial communities, opening the way for cultivation-independent analysis of microbial activity in environmental context.

We have used proteomic methods to analyze biofilm samples from acid mine drainage (AMD) biofilms previously characterized by cultivation-independent genomic methods (Ram, Science, 2005). This community consists of bacteria, archaea, and eukaryotes that have adapted to survive in an extreme environment (pH <1.0, high metal content, high temperature). LC/LC-MS/MS based "shotgun" proteomics with LTQ and LTQ-Orbitrap technologies identified over 2,000 proteins, yielding functional information about each of the five dominant species. The recent acquisition of additional genome data indicated the presence of two major strain variants of the dominant *Leptospirillum* II species. Detailed proteomic measurements provided the first evidence for large-scale genome recombination. The key methodological advance here is the finding that it is possible to deduce the sequences of gene variants, so long as genomic data from relatively closely related organisms are available.

Based on this initial work, we have established an experimental/computational proteomics pipeline at ORNL for the microbial community samples. To date, we have characterized six distinct biofilm samples from different regions in the Richland mine. In each sample, we were able to measure between 2,000 – 3,000 non-redundant proteins, including 1,000 – 1,200 unknown proteins (i.e. proteins that are predicted from the community genomic sequence whose existence has not previ-

ously been confirmed and which have no functional annotations). In many cases, these unknown proteins were identified in multiple samples, verifying the notion that they play key functional roles in the community operation, while other unknown proteins were found in individual biofilms suggestions potential spatial and/or temporal expression. We are in the process of examining the correlation between the fluorescent *in-situ* hybridization (FISH) data (which provides information about the microbial species abundance in each sample) and the proteome data (which provides protein information primarily from the most abundant organisms in each sample). In general, the proteomic information is providing important feedback for the genomic annotation and assembly. Whereas the genome information reveals genetic potential for the community members, the proteome data indicates the nature of the gene expression and conservation in the various species.

We have been able to acquire and integrate a high performance mass spectrometer into our proteomics pipeline this last year. In particular, this hybrid linear trapping quadrupole-Orbitrap (LTQ-Orbitrap) MS provides high resolution, accurate mass measurements on both parent and fragment ions in a high-throughput data-dependent fashion and on liquid chromatography time scales. The improved dynamic range of proteome measurement achievable with this instrumentation, when combined with the high resolution metrics stated above, provide enhanced capabilities for deeper and more accurate proteome measurements of these very complex samples. The LTQ-Orbitrap can routinely give mass accuracies on parent peptides with millimass accuracies (<5ppm mass error) while still permitting rapid MS/MS scans in the LTQ. With a 24-hour measurement, it is routine to measure 1500-2500 proteins from AMD samples. The ultra high mass accuracy combined with LTQ MS/MS spectra virtually eliminates false positives. A somewhat slower mode of operation is to collect both full-scan and MS/MS scans in the Orbitrap. This allows for very high mass accuracy on fragment ions, which facilitates *de-novo* sequencing. Even in this lower duty cycle mode of operation, 1000-1500 proteins can be identified in 24-hours and high abundance, high quality MS/MS spectra can be extracted and analyzed by *de-novo* sequencing algorithms.

We have also upgraded our existing Fourier transform ion cyclotron resonance mass spectrometer (FTICRMS) to incorporate a new electrospray ion source to enable advanced protein/peptide fragmentation techniques (infrared multiphoton dissociation (IRMDP) and electron capture dissociation (ECD)). These methods are complementary to the more conventional collisional fragmentation and are especially useful for larger peptides and proteins. In collaboration with our LLNL colleagues, we have been able to use this MS to investigate intact cytochrome proteins isolated from the extracellular biofilm fraction. Not only were we able to investigate the molecular mass heterogeneity of the various forms of cytochrome 579, but we also were able to use the IRMDP technique along with *de novo* sequencing approaches to discover and validate strain variations in the amino acid sequences. This provides important information that can be used to assist the "bottom-up" peptide identifications of strain variation in this natural community. We have begun work to use the new LTQ-Orbitrap technique for proteome quantification, achieved by isotopically labeling biofilm samples grown in laboratory bioreactors at UC-Berkeley. By comparing non-labeled and labeled biofilms, it should be possible to collect more definitive information about the relative abundance changes of proteins as a function of different biofilm growth states.

The explosion of experimental biofilm data this last year has prompted increasing bioinformatic needs. We have optimized our computational approach for data mining, interpretation, and dissemination. See http://compbio.ornl.gov/biofilm_amd/ and http://compbio.ornl.gov/biofilm_amd_recombination/ as examples. In addition, a new "peptide viewer" has been constructed to permit a more detailed examination of peptide sequence variation across different biofilm samples. A detailed informatic study was conducted to ascertain the level of strain variation that could be deciphered by MS-based proteomic measurements. The key concern motivating this study was the fact that even

* Presenting author

a single amino acid variation is sufficient to prevent *peptide* identifications; thus, it was necessary to evaluate what level of sequence variation at the amino acid level would preclude *protein* identifications. Further research has focused on false positives at different filtering levels for the AMD system and how this can be improved with high mass accuracy. The continuing goal is to limit both false positives and false negatives while obtaining as complete proteome coverage as possible.

On-going research continues to be directed at pushing the experimental and computational capabilities for deep and accurate proteome characterization in complex microbial communities. The limited complexity of the acid mine drainage system is the perfect system to begin to develop and evaluate these tools. In particular, we have learned a great deal about how to measure and decipher strain variation in microbial consortia. This work provides important information about how to deal with strain variation in complex systems of interest for both human health and environmental applications.

This research sponsored by the U.S. DOE-BER, Genomics:GTL Program. Oak Ridge National Laboratory is managed by UT-Battelle, LLC, for the U.S. Department of Energy under Contract No. DE-AC05-00OR22725.

69 GTL

Strain-Resolved Proteogenomics-Enabled Ecological Study of Natural Microbial Communities Associated with Acid Mine Drainage Formation

V.J. Denef[1]* (vdenef@berkeley.edu), N.C. VerBerkmoes,[2] P. Wilmes,[1] M. Shah,[2] D. Goltsman,[1] I. Lo,[1] G. DiBartolo,[1] L. Kalnejais,[1] B.J. Baker,[1] G.W. Tyson,[1] J.M. Eppley,[1] E.A. Allen,[1] R.L. Hettich,[2] M.P. Thelen,[3] and **J.F. Banfield**[1]

[1]University of California, Berkeley, California; [2]Chemical Sciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee; and [3]Biosciences Directorate, Lawrence Livermore National Laboratory, Livermore, California

Project Goals: The interdisciplinary research in our GTL project is motivated by the desire to understand how microbial communities assemble, the interplay amongst community members, and the mechanisms of microbial adaptation and evolution. Chemoautotrophic acid mine drainage (AMD) microbial biofilms have proven particularly tractable for these analyses due to their comparatively low species richness. Our goal is to use this system to develop cultivation-independent community genomic and proteomic methods and apply them to study the structure and activity of microbial communities in their natural environmental context.

Microbially promoted dissolution of metal sulfides leads to the formation of acid mine drainage (AMD), a major environmental problem associated with energy resources. It is also a process that underpins bioleaching-based metal recovery and coal desulphurization and mercury removal. Due to a limited spectrum of energy harvesting and metabolic opportunities, the chemoautotrophic microbial communities that populate acid mine drainage systems tend to have low species richness and thus are particularly amenable to high-resolution ecological analyses. Our approach is to use genomic data from two spatially and temporally separated natural microbial communities sampled at the Richmond mine (Iron Mountain, Redding, CA) to characterize the proteomes of multiple biofilms and to correlate the activities of organisms with community structure and environmental conditions.

To date, extensive community genomic data have been obtained from an air-solution interface biofilm (pH 0.83, 42 °C; Tyson et al., 2004) at the 5-way location and a subaerial biofilm (pH 1.1, 41 °C; Lo et al., resubmitted) at the UBA location. The *Leptospirillum* group II species from the 5-way CG and UBA datasets differ by 0.3% at the 16S rRNA gene level. The genomes are highly synteneous and share 83% and 76% orthologs (measured relative to the UBA 5-way CG gene inventories, respectively) with 95.24% average amino acid sequence identity (median = 96.69 %).

In order to evaluate the physiological and ecological significance of the *Leptospirillum* group II variants, we designed strain-specific fluorescent *in situ* hybridization (FISH) probes targeting the 23S rRNA gene to enable us to correlate the *Leptospirillum* group II type with environmental conditions. We detect no connection between genome type and pH, temperature, or ionic strength. However, the UBA type dominates in the early and middle stages of biofilm development whereas the 5-way CG type predominates in late successional stages when Archaea and fungi become important. Characterization of intact biofilm cross sections suggests that *Leptospirillum* group II most intensively colonize parts of the biofilm in direct contact with Fe-rich solutions whereas Archaea partition in the upper biofilm regions and show some association with fungal filaments. In contrast, *Leptospirillum* group III, the only nitrogen fixer yet identified in the system, is distributed throughout biofilms as single cells or microcolonies.

The reconstructed *Leptospirillum* group III genome encodes ~3,000 genes, with an average amino acid identity to *Leptospirillum* group II UBA type of ~58% (~1,800 orthologs). Compared to *Leptospirillum* group II, most biofilms contain a single near-clonal *Leptospirillum* group III type. However, genomic analyses reveal the same variants of the majority of *Leptospirillum* group III genes in all samples analyzed. Heterogeneity in *Leptospirillum* group II and III is typically in the form of differences in gene content, especially in putative prophage or plasmid-like regions. Analysis of genomic datasets revealed that recombination is an important process shaping archaeal populations in AMD biofilms. Quantitative genome-wide analyses indicate the frequency of recombination has a log-linear dependence on sequence divergence, with a significant discontinuity between rates within vs. between genomic clusters. The role of recombination in shaping the genetic potential of *Leptospirillum* group II was uncertain. Comparisons of the *Leptospirillum* group II UBA vs. 5-way CG genomes revealed regions of 10s to 100s kb in length and comprising a total of 421 genes sharing essentially identical nucleotide sequence. Excluding the subset attributed to integration of identical phage or IS elements, evidence suggests that the UBA and 5-way CG *Leptospirillum* species were shaped by recent homologous recombination between two organism types, followed by selective sweeps. Recombination is typically documented by sequence comparisons involving a few genes from organisms obtained in pure culture. In this study, we used shotgun proteomics to map, genome-wide, strain-specific expressed protein variants in a third community for which no genome sequence is available. Results reveal a *Leptospirillum* group II population dominated by a single organism with a genome of predominantly UBA type genes, but with chromosomal regions tens to hundreds of kilobases in length that derived from the 5-way CG genome type (confirmed by multi-locus sequence typing of isolates and uncultivated natural consortia). For both bacteria and archaea, within-gene recombination is an important source of new gene variants. Results suggest that formation of hybrid genome types is important in fine-scale environmental adaptation.

The ability to distinguish between proteins that differ by as few as a single amino acid enables strain-specific proteomics studies to resolve the behavior of closely related members of natural communities. Proteomic analyses of two mature biofilms that are colonized by both *Leptospirillum* types are underway. Replicate samples of biofilms representing different growth stages, as well as a biofilm showing evidence suggestive of phage predation, are in line for analysis in the next few months.

* Presenting author

Funding was provided by the DOE Genomics:GTL Program under grant number DE-FG02-05ER64134 (Office of Science), the NSF Biocomplexity Program, and the NASA Astrobiology Institute.

70 GTL

A Novel Iron Oxidase Isolated from an Extremophilic Microbial Community

Steven W. Singer,[1]* Christopher Jeans,[1] Jason Raymond,[1] Adam Zemla,[1,2] Nathan C. VerBerkmoes,[3] Robert L. Hettich,[3] Clara Chan,[4] **Jill Banfield**,[4] and Michael P. Thelen[1] (mthelen@llnl.gov)

[1]Chemistry, Materials and Life Sciences and [2]Computation Directorates, Lawrence Livermore National Laboratory, Livermore, California; [3]Chemical Sciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee; and [4]Department of Earth and Planetary Sciences, University of California, Berkeley, California

Project Goals: The interdisciplinary research in our GTL project is motivated by the desire to understand how microbial communities assemble, the interplay amongst community members, and the mechanisms of microbial adaptation and evolution. Chemoautotrophic acid mine drainage (AMD) microbial biofilms have proven particularly tractable for these analyses due to their comparatively low species richness. Our goal is to use this system to develop cultivation-independent community genomic and proteomic methods and apply them to study the structure and activity of microbial communities in their natural environmental context.

Proteins isolated directly from uncultivated microbial populations represent critical functional components of community structure and metabolism. Our investigation of a chemoautotrophic microbial community, typified by biofilm formation, iron oxidation and environmental acidification, has resulted in the identification of a large fraction of abundant proteins that do not correlate with any recognized sequences. As there are no systematic methods at hand to analyze "hypothetical proteins", we have developed an approach towards functional determination using proteogenomic, structural modeling and biochemical tools, with an initial focus on novel iron oxidizing proteins.

Two very abundant novel proteins isolated from acidophilic biofilms collected in Iron Mountain's Richmond Mine (described by JF Banfield in other presentations) were identified as *Leptospirillum* group II gene products and characterized as cytochromes with unique and unusual properties. A sulfuric acid wash of biofilm samples liberates a major 16 kDa cytochrome with an unusual α-band absorption at 579 nm, Cyt_{579} [Ram et al., *Science* 2005]. Using Cyt_{579}–specific antibodies, electron microscopic imaging of biofilm thin sections localizes the cytochrome to the outer cell surface of *Leptospirillum* bacteria. Purification from different microbial community samples unexpectedly indicated several different variations in protein mass and isoelectric points of Cyt_{579}. Further genomic sequencing of field samples indicated that several paralogs and variants of Cyt_{579} are present in distinct *Leptospirillum* strains, and MS proteomics confirmed that at least five of these variants occur in different amounts depending on the location and growth stage of the biofilm examined. Although the variations are due to one or more amino acid substitutions, multiple truncations at both ends of Cyt_{579} proteins were also determined, indicating processing that is perhaps important for cytochrome placement and function. Solution measurements of the purified protein point to a single domain, monomeric, α-helical structure. This is important in its interactions with other cytochromes involved in the oxidation of Fe(II) and electron transfer.

* Presenting author

In addition to the small soluble Cyt_{579}, membrane fractions of biofilm samples are dominated by a 60 kDa, heme-bearing protein with an α-band absorption at 572 nm. This cytochrome, Cyt_{572}, is located in the outer membrane of *Leptospirillum* group II, and again there are several variants and possible paralogs present in the community genome data. Cyt_{572} genes fall within a recombination hotspot between two strains of *Leptospirillum* group II, perhaps indicating the selection of variants of a protein essential to survival by its close association with changing geochemical conditions. Solution measurements of this purified cytochrome indicate a two-domain, oligomeric, and predominantly β-stranded structure.

Although both cytochromes 579 and 572 are so far unique to the *Leptospirilla* bacteria, a new "domain fusion" approach to structural modeling was used to test their similarity with known protein structures. Surprisingly, both cytochromes appear to have structural homology with bacterial nitrite reductases such as NirB. These are outer membrane proteins with homodimeric, two-domain structures. A small N-terminal heme binding domain is a good template for the entire Cyt_{579} polypeptide and also a relatively good fit to the N-terminal domain of Cyt_{572}. The large C-terminal domain of these nitrite reductases are 8-bladed β-propeller structures observed in many other proteins including those involved in signal transduction; this domain provides a structural scaffold for modeling the C-terminal domain of Cyt_{572}. These predictions are supported by solution measurements of purified cytochromes 579 and 572, including circular dichroism, protease digestion, size exclusion chromatography, and cytochrome oxidation by nitrite.

A general mechanistic model involving iron oxidation has resulted from our studies. Fe(II) oxidation by isolated Cyt_{572} occurs readily at pH 0.95 – 3, whereas Cyt_{579} is less reactive at low pH. Under specified conditions, Cyt_{572} transferred electrons to Cyt_{579}, perhaps representing an initial step in energy flow from the environment to the biofilm. Interestingly, a recently sequenced acidophilic bacterium contains a distinct operon of cytochromes that are homologous to cytochromes 579, 572, and a cytochrome oxidase. This indicates a coordinated regulation of these novel cytochrome genes and links these proteins to the generation of toxic mine drainage by acidophilic biofilm communities.

71 GTL

Functional Analysis of Protein Phosphorylation in *Shewanella oneidensis* MR-1

C. Giometti[1]* (csgiometti@anl.gov), G. Babnigg,[1] A. Beliaev,[2] G. Pinchuk,[2] M. Romine,[2] and J. Fredrickson[2]

[1]Argonne National Laboratory, Argonne, Illinois and [2]Pacific Northwest National Laboratory, Richland, Washington

Project Goals: This project is a component of the *Shewanella* Federation and as such contributes to the overall goal of applying the tools of genomics, leveraging the availability of genome sequence for 18 additional strains of *Shewanella*, to better understand the ecophysiology and speciation of respiratory-versatile members of this important genus. To understand these systems the SF is using genome-based approaches to investigate *Shewanella* as a system of integrated networks; first describing key cellular subsystems—those involved in signal transduction, regulation, and metabolism—then building towards understanding the function of whole cells and, eventually, cells within populations. As a general approach, the SF is collectively employing complimentary "top-down" bioinformatics-based genome functional predictions, high-throughput expres-

* Presenting author

sion analyses, and functional genomics approaches to uncover key genes as well as metabolic and regulatory networks. The "bottom-up" component employs more traditional approaches including genetics, physiology and biochemistry to test or verify predictions. This information will ultimately be linked to analyses of signal transduction and transcriptional regulatory systems and used to develop a linked model that will contribute to understanding the ecophysiology of *Shewanella* in redox stratified environments.

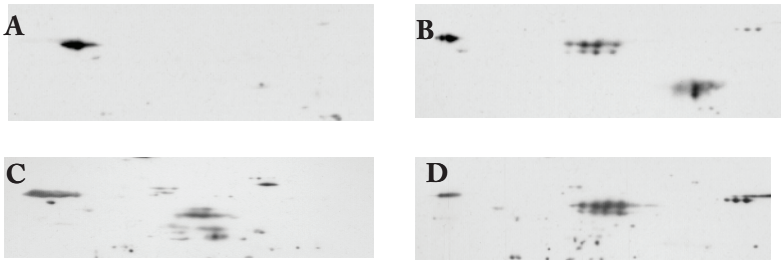


Figure 1. Western blot analysis of *S. oneidensis* proteins revealed proteins phosphorylated on serine or threonine residues. A, phosphothreonine proteins from MR-1 grown aerobically; B, phosphothreonine proteins from MR-1 grown anaerobically with fumarate; C, phosphoserine proteins from MR-1 grown aerobically; D, phosphoserine proteins from MR-1 grown anaerobically with fumarate.

Protein phosphorylation plays an important role in the regulation of cell physiology in both prokaryotes and eukaryotes. Using a suite of proteomics tools, including affinity chromatography, gel electrophoresis, Western blotting, and tryptic peptide mass analysis, we have identified a set of phosphoserine and phosphothreonine proteins expressed by *Shewanella oneidensis* MR-1 cells. Differential expression of a subset of these phosphoproteins, including pyruvate formate lyase and fumarate reductase (Figure 1) has been observed in cells grown with different electron acceptors. These results suggest that the phosphorylation status of these proteins is involved in regulating carbon and energy metabolism in *S. oneidensis* MR-1. To determine the functional significance of the observed differential phosphorylation in response to electron acceptor availability, protein phosphorylation in MR-1 strains deficient in key respiratory proteins, the tetraheme cytochrome c, CymA, and fumarate reductase, FccA, is being investigated. These mutants, generated as part of the *Shewanella* Federation efforts to characterize MR-1 carbon and energy metabolism, are being grown in parallel with MR-1 wild type with different electron acceptors. The phosphoserine and phosphothreonine protein profiles of these cultures are then generated by using Western blot analysis, and quantitative comparative analysis is used to detect significant differences in the phosphoprotein profiles. Characterization of the phosphorylation sites is being done using mass spectrometry. In addition, studies are being done to further characterize the differential phosphorylation of pyruvate formate lyase observed in response to oxygen availability and the phosphorylation of fumarate reductase, known to be a periplasmic protein in *S. oneidensis.*

This work is funded by the U.S. DOE Office of Biological and Environmental Research under Contract No. DE-AC02-06CH11357.

72 GTL

Enriching Metabolic Function Predictions for *Shewanella oneidensis* MR-1 with Growth and Expression Studies

Margrethe H. Serres[1]* (mserres@mbl.edu), Margaret F. Romine,[2] Mike E. Driscoll,[3] Tim S. Gardner,[3] Natalia Maltsev,[4] Miriam Land,[5] Andrei Osterman,[6] Mary Lipton,[2] and LeeAnn McCue[2]

[1]Marine Biological Laboratory, Woods Hole, Massachusetts; [2]Pacific Northwest National Laboratory, Richland, Washington; [3]Massachusetts Institute of Technology, Boston, Massachusetts; [4]Argonne National Laboratory, Argonne, Illinois; [5]Oak Ridge National Laboratory, Oak Ridge, Tennessee; and [6]Burnham Institute, La Jolla, California

Project Goals: This project is a component of the *Shewanella* Federation and as such contributes to the overall goal of applying the tools of genomics, leveraging the availability of genome sequence for 18 additional strains of *Shewanella*, to better understand the ecophysiology and speciation of respiratory-versatile members of this important genus. To understand these systems the SF is using genome-based approaches to investigate *Shewanella* as a system of integrated networks; first describing key cellular subsystems—those involved in signal transduction, regulation, and metabolism—then building towards understanding the function of whole cells and, eventually, cells within populations. As a general approach, the SF is collectively employing complimentary "top-down" -bioinformatics-based genome functional predictions, high-throughput expression analyses, and functional genomics approaches to uncover key genes as well as metabolic and regulatory networks. The "bottom-up" component employs more traditional approaches including genetics, physiology and biochemistry to test or verify predictions. This information will ultimately be linked to analyses of signal transduction and transcriptional regulatory systems and used to develop a linked model that will contribute to understanding the ecophysiology of *Shewanella* in redox stratified environments.

The genome sequence *Shewanella oneidensis* MR-1 (MR-1) was released in 2002 (1) and reannotated one year later (2). MR-1 is the model organism being used by the *Shewanella* Federation team to develop a better understanding of the ecophysiology and speciation of respiratory-versatile members of the *Shewanella* genus. These genome annotations have served as an essential resource for designing experiments to investigate MR-1 function and for interpreting experimental results generated from high through-put analyses such as microarray and global proteomics. In order to improve the accuracy of these initial annotations we have applied various new bioinformatics tools including Puma2, Gnare, and SEED as well as information from 15 new *Shewanella* genome sequences to improve functional predictions based on sequence information. In addition, we have used proteome data generated by the *Shewanella* Federation to improve ORF calls and to validate predictions of protein processing. The new ORF predictions were used to design an MR-1 specific Affymatrix chip and to create a modified protein sequence file for proteome analysis. They have also been used to improved predictions of operon structure and cellular localization of proteins.

Predicted functions have been used to generate a picture of the intermediate metabolism of MR-1 (3). Analogous to the pathways described in the BioCyc database as well as the literature, we have made approximately 150 pathway predictions, and we are constructing a *Shewanella* pathway database (SheonDB) for MR-1 using the Pathway Tools software (4). In addition public metabolic databases i.e. the KEGG map are being consulted.

We are now seeking to further improve the accuracy of our current predictions by interrogating additional experimental resources, such as microarray analyses of expression patterns and growth

* Presenting author

phenotypes. A comparative analysis of cells grown aerobically with one of 5 different carbon sources (casamino acids, N-acetyl-D-glucosamine, inosine, pyruvate and lactate) supported many previously reported pathway assignments (3) and suggested alternative assignments for others. Such a case was seen for the degradation of the branched chain amino acids of leucine, valine, or isoleucine. The first step, a deamination reaction, was proposed to be catalyzed by either the branched-chain-amino-acid aminotransferase IlvE (SO0340) or by the leucine dehydrogenase Ldh (SO2638). The microarray data suggests that SO2638 catalyzes this reaction during growth on casamino acids generating NADH and ammonia as byproducts. Also, according to the KEGG map the third step in the conversion of these amino acids to the respective methyl butanoyl-CoA intermediate (EC 2.3.1.-) was missing an assignment for MR-1. The prediction that SO2341 (EC 2.3.1.168) carried out this step (3) was supported by the microarray data. By contrast our earlier predictions for subsequent degradative steps in valine and isoleucine are not supported by the microarray data. We originally predicted that the sequential conversion of isobutyryl-CoA (valine degradation) to propionyl-CoA and acetyl-CoA was mediated via EC 1.3.99.12 (SO0021), EC 4.2.1.17 (SO0021, SO1681), EC 3.2.1.4 (SO0020), EC 1.1.1.31 (SO1682), and EC 1.2.1.27 (SO1678). Combined results from genome neighborhood analysis, literature analysis, and microarray data suggest that instead the process is mediated by EC 1.3.99.12 (SO1679), EC 4.2.1.17 (SO1681), EC 3.1.2.4 (1680), EC 1.1.1.31 (SO1682), and EC 1.2.1.27 (SO1678).

Our results suggest that integration of bioinformatics analysis with experimental data interrogation can provide improved annotations and hence can more effectively drive subsequent experimental design and data interpretation. Additional examples of how we have applied microarray data to predict new functions will be presented in the poster.

References

1. Heidelberg, J.F. et al. (2002) Genome sequence of the dissimilatory metal ion-reducing bacterium *Shewanella oneidensis*. *Nat Biotechnol*. 20:1118-23.
2. Darselia N. et al. (2003) Reannotation of *Shewanella oneidensis* genome.OMICS. 7:171-5.
3. Serres M.H., Riley, M. (2006) Genomic analysis of carbon source metabolism of *Shewanella oneidensis* MR-1: Predictions versus experiments. J Bacteriol. 188(13):4601-9.
4. Karp, P., Paley, S., Romero, P. (2002) The Pathway Tools Software. *Bioinformatics* 18:S225-32.

73 GTL

Proteomics Technologies Advance the Understanding of Microbial Systems Allowing for In-Depth Characterization of Microbes Important for Bioenergy Production, Bioremediation and Carbon Sequestration and Cycling

Mary S. Lipton* (mary.lipton@pnl.gov), Joshua Turse, Stephen Callister, Kim K. Hixson, Xuixia Du, Angela Norbeck, Samuel Purvine, Feng Yang, Margie F. Romine, Carrie D. Nicora, Joshua Adkins, Richard D. Smith, and Jim K. Fredrickson

Biological Sciences Division, Pacific Northwest National Laboratory, Richland, Washington

Project Goals: The understanding of how cells function at the systems level will greatly benefit from the development of new approaches capable of making global measurements of protein expression (proteome). The aim of this project is to apply new capabilities that are being devel-

oped for quantitative and high throughput proteomic measurements, based primarily upon high resolution separations combined with the unique high field Fourier transform ion cyclotron resonance mass spectrometry technology developed at PNNL. The project focuses on biological applications involving studies of the proteomes of several microorganisms including *Shewanella oneidensis* MR1, *Deinococcus radiodurans* R1, *Rhodobacter sphaeroides*, *Peligobacter ubique*, *Caulobacter crecentus* and *Geobacter* species that are of interest to the DOE Genomics:GTL program. Each organism plays a significant role in bioremediation, carbon sequestration energy production or furthering the understanding of biological systems. These efforts are proceeding in collaboration with biologists with expertise in each of these organisms, and in a close collaboration with each of the biologists, the proteomic data is translated into biological implications of changes in cellular stress or state. The focus of each of the organisms is listed below as well as the collaborator that serves as the biological lead for these subprojects.

Collaborators (Laboratory leads): Michael Daly (USUHS), Sam Kaplan (UT-Houston Medical School), Tim Donohue (University of Wisconsin-Madison), Lucy Shapiro (Stanford University), Stephen Giovannoni (Oregon State University), Derek Lovley (University of Massachusetts), Andre Osterman (Burnham Institute), Pavel Pevzner (UCSD)

Exploiting microbial function for purposes of bioremediation, energy production, carbon sequestration and other missions important to the DOE requires in-depth systems level knowledge of the molecular components of the cell that confer function. Inherent to developing a systems level understanding is comprehensive characterization of cellular proteins and how their abundance, location and modification state respond to changing conditions. Recent advances in proteomics technologies at PNNL have allowed the determination of shifts in relative abundance, localization, interactions, and post-translational modifications of cellular proteins. We will present illustrations of how the high through put technologies at PNNL have been utilized to elucidate these post-transcriptional events in microbial cells.

Global proteomics is now capable of comprehensively identifying cellular proteins. This information has proved useful in the annotation of microbial genomes. For example, global proteomics has validated hypothetical proteins, n-and c- terminal sequences, various post-translational processing and signal peptide cleavage events. In *Shewanella oneidensis*, a blind search methodology was applied to identify peptides arising from post-translational enzymatic processing or chemical modification. Standard peptide identification methods were then used to refine protein-coding gene starts and predictions of signal peptidase cleavage greatly improving predictions of which proteins are secreted into the cell envelope. In addition to confirming many prior discoveries, this novel application of proteomics revealed 1) predicted to arise by natural mutation/variation in the cell population were in fact not the result of sequencing mistake 2) several new start codons including one instance of a rare ATA start codon, and 3) 390 instances of chemical modification of proteins.

While comprehensive identification of cellular proteins has important applications, , robust measurements of abundance are needed. The technique that shows the most promise when used in combination with the AMT tag approach is the utilization of absolute peak intensity from high-resolution FTICR instruments. Compared to stable isotope labeling and spectrum counting, this method produces quantitative data with a higher confidence from a single run, thus allowing for a statistical handling of multiple technical and biological replicates. Such analyses provide the global determination of quantitative protein response to the culture conditions or environmental condition. Changes in protein abundance levels were investigated in *Geobacter sulfurreducens* grown on fumarate or Fe(III) citrate. The abundance of proteins was determined and the results compared to identify proteins associated with these distinct modes of anaerobic respiration. Among the proteins that changed,

 * Presenting author

91 *c*-type cytochromes were identified. Relative abundance of some *c*-type cytochromes varied markedly with different growth conditions.

While the characterization of steady state cultures focus on the static survey of the cell, time course studies allow for understanding the dynamics of a system. High-throughput proteomics technologies make it possible to examine the dynamics and regulatory mechanism of biological pathways via measuring the protein expressions at multiple time points. For example, in *Rhodobacter sphaeroides*, the response of the bacteriochlorophyll production pathway increase over time as cells transition from an aerobic to a photosynthetic state suggesting the importance of this pathway in the synthesis of the photosynthetic reaction center.

While the characterization of protein abundance allows an understanding of how cells respond to environmental conditions, protein location within cells can change as a function of time and conditions. The global determination of these sub-cellular protein localizations is an ability that is unique to proteomics methodologies. We present results for *Rhodobacter sphaeroides* that characterize the proteome of aerobic and photosynthetic cell cultures by utilizing: 1) proteins extracted from whole cell lysate, soluble, insoluble, and global fractions, and 2) proteins extracted from sub-cellular fractions that include cytoplasm, cytoplasmic membrane, periplasm, outer membrane, and chromatophore. Additionally, the application of informatics techniques to those proteins that are assigned to multiple locales can aid in the determination of potential protein interaction partners since the co-localization events could arise from the formation of interactions among proteins.

Post-translational modifications are important components of protein function. Characterization of these modifications is another application that is unique to proteomics technologies. Types of modifications can vary from the addition of large moieties like hemes to addition of small ligands, such as phosphorylation. Heme modification plays an important role in electron transfer and enzyme catalysis, while phosphorylation at histidine and aspartate residues of the response regulator is essential to regulate the signal transduction pathway in the two-component system. Oxidation of proteins is often a response to a stress on the cell. The use of specialized separation or enrichment schemes in combination with high-resolution mass spectrometry allows the characterization and quantitative measurement of these post-translational modifications under different biological conditions. When used in conjunction with measurements of global protein abundance and subcellular localization, greater depth of understanding about cellular response, both to and upon the environment, will emerge.

Emergent work in our lab includes application of proteomics to microbial communities. New advances in both separations and instrumentation resolution have allowed characterization of the protein expression patterns of the microbes within these communities, thus furthering the understanding of how these microbes interact with their environment. For example, we have applied the AMT tag approach to *Pelagibacter ubique (a.k.a. "SAR11")*, perhaps the most abundant microbe in sea-water communities. The data are revealing the adaptive strategies that enable these alphaproteobacteria to recycle carbon efficiently throughout the oceans.

We are also building on this work to include other microbes such as *Caluobacter crescentus* focusing on the determination of important pathways for bioenergy production of both cultured and yet undiscovered or uncultured organisms open the potential for the increase in production of biofuels and the mitigation of the use of these fuels in the environment.

74 GTL

Functional Genomic Analysis of Current Production in High Power Density Microbial Fuel Cells

Kelly P. Nevin* (knevin@microbio.umass.edu), Sean F. Covalla, Jessica P. Johnson, Trevor L. Woodard, Raymond DiDonato Jr., Kim K. Hixson, Mary Lipton, and **Derek R. Lovley**

Department of Microbiology, University of Massachusetts, Amherst, Massachusetts

Project Goals: The overall purpose of this project is to develop experimental and computational tools to predictively model the behavior of complex microbial communities involved in microbial processes of interest to the Department of Energy. The five year goal is to deliver in silico models that can predict the behavior of two microbial communities of direct relevance to Department of Energy interests: 1) the microbial community responsible for in situ bioremediation of uranium in contaminated subsurface environments; and 2) the microbial community capable of harvesting electricity from waste organic matter and renewable biomass. The research in this abstract summarizes research under Subproject III. The purpose of Subproject III is to define the function of genes of unknown function involved in environmentally significant aspects of the physiology of *Geobacter* species.

Recent advances in the engineering of microbial fuel cells has greatly increased their power output, suggesting that expanded applications, such as powering mobile electronics and large-scale conversion of wastes to electricity, may be feasible goals. Furthermore, although it has widely been considered that mixed microbial communities are required for maximal power output of microbial fuel cells, more recent studies have demonstrated that pure cultures of *Geobacter sulfurreducens* can produce power densities equivalent to those observed with mixed communities. Surprisingly, *G. sulfurreducens* forms thick biofilms on the anodes of such systems and there is a direct correlation between the amount of biomass and current, suggesting that cells at substantial distance from the anode are as effective in contributing to current production as cells in close association with the anode surface. This contrasts with the concept, derived from studies of low power density microbial fuel cells, that cells must be in intimate contact with the anode in order to significantly contribute to power production.

In order to understand the mechanisms for long-range electron transfer to anodes in high power density microbial fuel cells, functional genomic studies of *G. sulfurreducens* growing in flow-through fuel cells are being conducted. For example, a series of deletion mutants which have deficiencies in Fe(III) reduction were evaluated for their ability to produce high levels of current. One gene of particular interest was *pilA*, which encodes the structural pilin protein. Previous studies demonstrated that deleting *pilA* prevents pilin production and Fe(III) oxide reduction. The pilin appear to be electrically conductive and these 'microbial nanowires' are proposed to be the final conduit for electron transfer between the cell and the oxides. Previous studies demonstrated that the *pilA*-deficient mutant produced power as well as wild-type cells in low power density microbial fuel cells. However, in high power density systems current was less than 10% of that observed for wild-type and there has been no adaptation for increased power production in long-term incubations. Expressing *pilA in trans* restored current production. Quantitative PCR analysis of *pilA* transcript levels demonstrated that expression of *pilA* increased as current levels, and hence the thickness of the biofilm, increased. These results demonstrate that pili are required for high density current production and suggest that the pilin contribute to long-range electron transfer through the anode biofilm. Pili might also contribute to the structure of the anode biofilm. However, a whole genome comparison of gene

* Presenting author

transcript levels with DNA microarrays demonstrated that *pilA* and the pseudopilin gene just downstream of *pilA* were the most highly upregulated genes in current-producing biofilms compared to equally thick biofilms grown on the same graphite surface but with fumarate as the electron acceptor. This indicates an increased need for pilin for electricity production over any potential structural role in the biofilm.

OxpG, is a putative pseudopilin in *G. sulfurreducens* which is part of a type II secretion system necessary for export of proteins essential for Fe(III) oxide reduction to the outer membrane. An *oxpG*-deficient mutant was severally limited in its ability to produce electricity, but produced pili. This suggests that there are outer-membrane proteins other than pili that are essential for current production. Mutants deficient in proteins that are known to be secreted by the type II system are now being evaluated for their capacity for electricity production.

Several outer-membrane *c*-type cytochromes also appear to play a role in high density current production. Previous studies have demonstrated that OmcS is important for low-power density current production. Although transcript levels for *omcS* were elevated in low power density fuel cells, this gene was down regulated in high power density fuel cells. In contrast, expression of genes for other outer-membrane *c*-type cytochromes, such as OmcB and OmcE increased at higher power levels. Although deleting the genes for one or two *c*-type cytochromes typically resulted an initial decrease in power production, these mutants eventually adapted to produce power as well as wild type. This suggests that there is some plasticity in the intermediary pathways of electron transfer to anodes, much more than has been observed in studies on Fe(III) oxide reduction. Proteomic and microarray studies of adapted strains are underway in order to better understand how blockages in electron transfer through important outer-membrane *c*-type cytochromes is overcome. Proteomics studies indicate that outer-membrane *c*-type cytochrome omcB, pilin, and numerous hypothetical proteins are present in substantially higher quantities in current harvesting biofilms than in biofilms grown using a soluble electron acceptor.

The periplasmic *c*-type cytochrome, PpcA, is one of the most abundant cytochromes in *G. sulfurreducens* and has previously been shown to be a key intermediary in electron transfer in Fe(III) reduction. Expression of *ppcA* was much 84 fold higher in cells producing current than cells reducing soluble Fe(III). Deleting *ppcA* resulted in a substantial lag in current production and even after long adaption periods the mutant only produced 70% as much power as wild-type cells. Global proteomics analysis to evaluate the mechanisms for adaptation to the loss of PpcA are underway.

The factors leading to optimal power production are complex because they are dependent upon proper biofilm formation, fuel consumption, and multifaceted pathways of extracellular electron transfer. This requires appropriate regulation of many genes. Therefore, the effect of disrupting global regulatory systems is under investigation. For example, deleting the gene for the sigma factor, RpoE, resulted in a significant lag in power production, lower maximum power production than wild type, and a biofilm that was less adherent to the anode surface. Evaluation of a number of other regulatory mutants is in progress.

75 GTL

Genome-Scale Analysis of Adaptive Evolution of *Geobacter* for Improved Metal Reduction and Electricity Production

Zarath Summers,[1]* Kelly Nevin,[1] Chris Herring,[2] Richard Glaven,[1] Shelley Haveman,[1] James Elkins,[2] Bernhard Palsson,[2] and **Derek Lovley**[1] (dlovley@microbio.umass.edu)

[1]Department of Microbiology, University of Massachusetts, Amherst, Massachusetts and [2]Department of Bioengineering, University of California, San Diego, California

Project Goals: The overall purpose of this project is to develop experimental and computational tools to predictively model the behavior of complex microbial communities involved in microbial processes of interest to the Department of Energy. The five year goal is to deliver in silico models that can predict the behavior of two microbial communities of direct relevance to Department of Energy interests: 1) the microbial community responsible for in situ bioremediation of uranium in contaminated subsurface environments; and 2) the microbial community capable of harvesting electricity from waste organic matter and renewable biomass. The research in this abstract summarizes research under Subproject II. The purpose of Subproject II is to describe gene expression of *Geobacter*aceae under environmental conditions that have relevance for in situ uranium bioremediation and harvesting electricity from waste organic matter and to understand factors that may alter this gene expression.

Geobacter species are typically the predominant microorganisms in uranium-contaminated subsurface environments undergoing *in situ* uranium bioremediation and on the surface of electrodes harvesting electricity from organic wastes. The conditions that *Geobacter* species face during *in situ* uranium bioremediation and on the surface of energy-harvesting anodes are substantially different than the conditions under which *Geobacter* species have evolved for billions of years. For example, electron donors are generally limiting in most subsurface environments, but during *in situ* uranium bioremediation electron donors are supplied in excess. Furthermore, in some instances the electron donors supplied, such as lactate, are compounds that are not the primary electron donors for growth of *Geobacter* species under natural conditions. Microbial fuel cells represent a novel environment for *Geobacter* species because, as far as is known, there has been no previous evolutionary pressure on microorganisms to produce electricity. These considerations suggest that long-term exposure of *Geobacter* species to the environmental conditions that prevail during *in situ* uranium bioremediation and in microbial fuel cells will select for changes in gene expression, and possibly beneficial mutations, that will favor growth under these artificially imposed conditions. If so, this could improve the rate and extent of *in situ* uranium bioremediation and enhance power output from microbial fuel cells.

Recent studies with *E. coli* have demonstrated that microorganisms can rapidly evolve when subjected to new environmental constraints. For example, *E. coli* K-12 initially grew much slower on glycerol than predicted from genome-based *in silico* modeling, but growth rates progressively increased toward the predicted optimum as the organism was repeatedly transferred in a minimal medium containing glycerol as the sole carbon and energy source. It has previously been difficult to link such phenotypic variation directly to changes in genotype because subtle differences (i.e. SNPs) could not readily be detected. However, microarray-based comparative genome sequencing revealed that there were a number of spontaneous mutations including SNPs, indels, duplications, and large-scale rearrangements during the adaptation of *E. coli* to faster growth on glycerol. Mutations in genes encoding glycerol kinase (*glpK*) and large subunits of the RNA polymerase (*rpoC* and *rpoB*) resulted in the greatest fitness for growth on glycerol and these mutations were rapidly fixed in the

* Presenting author

evolving populations. Evolved growth phenotypes could be reconstructed in the wild-type strain by introducing the experimentally determined mutations via site-directed mutagenesis. These studies demonstrated that genome-wide perturbations can be identified during laboratory-scale evolution studies and that causal mutations directly linking evolved genotypes to the resulting phenotypes can be determined. These methods developed with *E. coli* can now be applied to adaptive evolution studies with *Geobacter* species.

In order to understand how *Geobacter* species might evolve during *in situ* uranium bioremediation or harvesting electricity from waste organic matter, adaptive evolution studies were initiated with *Geobacter sulfurreducens*. For example, to determine if *G. sulfurreducens* could be adapted for more rapid extracellular electron transfer, it was repeatedly transferred with Fe(III) oxide as the electron acceptor under conditions which favored rapid growth on Fe(III) oxide. A strain was developed which can transfer electrons to Fe(III) and Mn(IV) oxides 10 times faster than the unadapted strain. Analysis of gene transcript levels with whole-genome DNA microarrays demonstrated that the evolved strain had higher levels of expression of genes for proteins thought to be involved in electron transfer to Fe(III) oxides, such as *c*-type cytochromes, and pili. Transcripts of genes encoding transport proteins, central metabolism enzymes, and several hypothetical proteins were also higher in the adapted strain. Surprisingly, when the gene for the outer-membrane *c*-type cytochrome, OmcS, was deleted in the adapted strain this had no impact on Fe(III) oxide reduction whereas OmcS is required for Fe(III) oxide reduction in the unadapted strain. This suggests that the adapted strain of *G. sulfurreducens* has significant changes in it's extracellular electron transport chain. One key mutation in the adapted strain has already been detected and comparative genome sequencing is in progress.

Other adaptive evolution studies are underway. For example, the lower the potential that microorganisms transfer electrons to the anodes of microbial fuel cells, the greater the power production. A strain of *G. sulfurreducens* has been adapted to transfer electrons to a fuel cell anode at much lower potential than the wild-type strain. Lactate is a convenient electron donor source to add to the subsurface in order to promote *in situ* uranium bioremediation, but lactate is not a common electron donor in natural anaerobic sedimentary environments. Therefore, studies have been initiated to elucidate how *G. sulfurreducens* adapts for enhanced lactate utilization with continual exposure to excess lactate as the sole electron donor. Additional selective pressures that have relevance to *in situ* uranium bioremediation and/or optimizing power output of microbial fuel cells are also in progress.

76 GTL

Proteomic Profiling of the *Caulobacter crescentus* Cell Cycle and Starvation Response

Esteban Toro[1]* (etoro@stanford.edu), Leticia Britos,[1] Samuel O. Purvine,[2] Mary S. Lipton,[2] Tom Taverner,[1] Feng Yang,[2] **Harley H. McAdams**,[1] Richard D. Smith,[2] and Lucy Shapiro[1]*

[1]Department of Developmental Biology, Stanford University, Stanford, California and [2]Pacific Northwest National Laboratory, Richland, Washington

Project Goals: Identification of the overall regulatory and metabolic networks in *Caulobacter crescentus*, largely through gene expression microarray assays and bioinformatic analysis

The gram-negative bacterium, *Caulobacter crescentus*, is closely related to economically important plant symbionts (e.g. *Rhizobium spp*.) and human pathogens (e.g. *Rickettsia spp*.), is ubiquitous, innocuous, and easily manipulated with standard genetic methods. *Caulobacter* cells are asymmetric, dividing into two distinct cell types that can be easily differentiated using light microscopy. Most importantly, *Caulobacter* cultures can be synchronized with a simple procedure, allowing researchers to investigate the modulation of cellular processes during the cell cycle and determine the detailed molecular mechanisms governing cellular operations. We have used liquid chromatography coupled to mass spectrometry (LC-MS and LC-MS/MS) to: i) directly determine the relative levels of all *Caulobacter* proteins during the cell cycle; ii) compare the proteomic profile of exponentially growing cells to stationary cells and cells being starved for carbon; and iii) quantitate absolute levels of key regulatory proteins.

Using 2-dimensional sample fractionation, with strong cation exchange (SCX) and reverse phase liquid chromatography coupled to tandem MS analysis (LC-MS/MS) we have confidently identified 3174 distinct *Caulobacter* proteins (accounting for about 84% of *C. crescentus* predicted genes) in exponentially growing cells. These proteins included inner and outer membrane proteins as well as proteins with extreme pIs that are difficult to resolve using gel electrophoresis-based methodologies. This number includes only proteins identified using a conservative methodology which requires at least two unique peptides and unambiguous MS/MS characterization before a protein hit is called. Comparing our peptide probability results to independent predictions from PeptideProphet shows our false positive rate to be below 5%.

Taking advantage of the ability to synchronize *Caulobacter* cultures, we performed global proteomics measurements as the cells progressed through the cell cycle. Furthermore, in order to obtain a preliminary view of how information from the environment is processed, we have done proteomic experiments to measure *Caulobacter's* response to carbon starvation and stationary phase. Finally, we describe initial results on absolute quantitation of key regulatory proteins throughout the *Caulobacter* cell cycle.

77 GTL

Quantitative Shotgun Proteomics with ProRata: Application to Anaerobic Aromatic Degradation in *Rhodopseudomonas palustris*

C. Pan,[1]* G. Kora,[1] Y. Oda,[2] D. Pelletier,[1] N. C. VerBerkmoes,[1] W. H. McDonald,[1] G. Hurst,[1] C. S. Harwood,[2] R. L. Hettich[1] (hettichrl@ornl.gov), and **N. F. Samatova**[1] (samatovan@ornl.gov)

[1]Oak Ridge National Laboratory, Oak Ridge, Tennessee and [2]University of Washington, Seattle, Washington

Project Goals: Quantitative shotgun proteomics has recently emerged as a high throughput technique for measuring the relative abundances of thousands of proteins between two cellular conditions. The project addressed the following challenges: 1) accurate estimation of protein abundance ratios from noisy mass spectral data; (2) extraction of reliable information from a biological comparison. It then aims to characterize the catabolism of p-coumarate in *Rhodopsedumonas palustris*.

Organisms often respond to environmental or physiological stimuli by adjusting the type and abundance of proteins in their cells. Measurement of the relative abundances of proteins in treatment cells subjected to stimuli, compared to that in the reference cells, provides valuable insights about protein function and regulation. Quantitative shotgun proteomics has recently emerged as a high throughput

* Presenting author

technique for measuring the relative abundances of thousands of proteins between two cellular conditions. The reference and treatment proteomes are labeled with different stable isotope tags and then mixed in equivalent amounts. In such a proteome mixture, each protein has two mass-different isotopic variants: the light isotopologue and the heavy isotopologue. The proteome mixture is digested and then analyzed with liquid chromatography–tandem mass spectrometry (LC–MS/MS). There are two folds of informatics challenges in quantitative proteomics: 1) accurate estimation of protein abundance ratios from noisy mass spectral data; (2) extraction of reliable information from a biological comparison.

ProRata: A quantitative proteomics program for accurate protein abundance ratio estimation with confidence interval evaluation [1, 2].

The abundance ratio between the light and heavy isotopologues of an isotopically labeled peptide can be estimated from their selected ion chromatograms. However, quantitative shotgun proteomics measurements yield selected ion chromatograms at highly variable signal-to-noise ratios for tens of thousands of peptides. This challenge calls for algorithms that not only robustly estimate the abundance ratios of different peptides but also rigorously score each abundance ratio for the expected estimation bias and variability. Scoring of the abundance ratios, much like scoring of sequence assignment for tandem mass spectra by peptide identification algorithms, enables filtering of unreliable peptide quantification and use of formal statistical inference in the subsequent protein abundance ratio estimation. In this study, a parallel paired covariance algorithm is used for robust peak detection in selected ion chromatograms. A peak profile is generated for each peptide, which is a scatter-plot of ion intensities measured for the two isotopologues within their chromatographic peaks. Principal component analysis of the peak profile is proposed to estimate the peptide abundance ratio and to score the estimation with the signal-to-noise ratio of the peak profile (profile signal-to-noise ratio). We demonstrate that the profile signal-to-noise ratio is inversely correlated with the variability and bias of peptide abundance ratio estimation. Then, a profile likelihood algorithm is proposed to infer the abundance ratios of proteins from the abundance ratios of isotopically labeled peptides. Given multiple quantified peptides for a protein, the profile likelihood algorithm probabilistically weighs the peptide abundance ratios by their inferred estimation variability, accounts for their expected estimation bias, and suppresses contribution from outliers. This algorithm yields maximum likelihood point estimation and profile likelihood confidence interval estimation of protein abundance ratios. This point estimator is more accurate than an estimator based on the average of peptide abundance ratios. The confidence interval estimation provides an "error bar" for each protein abundance ratio that reflects its estimation precision and statistical uncertainty. The accuracy of the point estimation and the precision and confidence level of the interval estimation were benchmarked with standard mixtures of isotopically labeled proteomes. The parallel paired covariance algorithm, the principal component analysis algorithm and the profile likelihood algorithm were integrated into a quantitative proteomics program, called ProRata, freely available at www.MSProRata.org.

Characterization of Anaerobic Catabolism of *p*-Coumarate in *Rhodopseudomonas palustris* by Integrating Transcriptomics and Quantitative Proteomics [3].

In this study, the pathway for anaerobic catabolism of *p*-coumarate by a model bacterium, *Rhodopseudomonas palustris*, was characterized by comparing its gene expression profile under *p*-coumarate growth against those under succinate and benzoate growth. Gene expression was quantified at the mRNA level with transcriptomics and at the protein level with quantitative proteomics using ^{15}N metabolic labeling. Both -omics measurements were critical, since the transcriptomics provided near full genome coverage of gene expression profiles and the quantitative proteomics surveyed the expression activities of over 1,500 genes at the protein level. The integrated gene expression data are consistent with the proposal that *p*-coumarate is converted to benzoyl-CoA, which is then degraded

via a known aromatic ring reduction pathway. For the metabolism of *p*-coumarate to benzoyl-CoA, two alternative routes, a ß-oxidation route and a non-ß-oxidation route, are possible. Based on the integrated gene expression data, we suggest that the anaerobic catabolism of *p*-coumarate likely proceeds through the non-ß-oxidation route in *R. palustris*. A putative gene was proposed for every step in the non-ß-oxidation route.

Research sponsored by the DOE Genomics:GTL Program.

References

1. C. Pan, G. Kora, D.L. Tabb, D.A. Pelletier, W.H. McDonald, G.B. Hurst, R.L. Hettich, and N.F. Samatova, Robust estimation of peptide abundance ratios and rigorous scoring of their variability and bias in quantitative shotgun proteomics. Analytical Chemistry, 2006. **78**(20): p. 7110-7120.
2. C. Pan, G. Kora, W.H. McDonald, D.L. Tabb, N.C. VerBerkmoes, G.B. Hurst, D.A. Pelletier, N.F. Samatova, and R.L. Hettich, ProRata: A quantitative proteomics program for accurate protein abundance ratio estimation with confidence interval evaluation. Analytical Chemistry, 2006. **78**(20): p. 7121-7131.
3. C. Pan, Y. Oda, D.A. Pelletier, B. Zhang, P.K. Lankford, N.F. Samatova, C.S. Harwood, R.L. Hettich. Characterization of Anaerobic Catabolism of p-Coumarate in *Rhodopseudomonas palustris* by Integrating Transcriptomics and Quantitative Proteomics. (Submitted)

78 GTL

From Genome to Metabolome: Correlating a System-Wide Response to Environmental Adaptation in a Hyperthermophile

Sunia A. Trauger[1]* (strauger@scripps.edu), Ewa Kalisiak,[1] Jarek Kalisiak,[2] Hiro Morita,[1] Angeli Lal Menon,[3] Michael V. Weinberg,[3] Farris L. Poole,[3] Michael W. W. Adams,[3] and **Gary Siuzdak**[1]

[1]Center for Mass Spectrometry and the [2]Department of Chemistry, The Scripps Research Institute, La Jolla, California; and [3]Department of Biochemistry and Molecular Biology, University of Georgia, Athens, Georgia

Project Goals: Watson and Crick's discovery united structure and genetics in a very powerful way. With the combined effort of the MAGGIE investigators, we seek to achieve a similar type of unification for gene sequence and cell biology. MAGGIE investigators have developed model microbial systems and technologies suitable to derive a prototypical multi-level map of protein modifications and interactions. We believe that the requested level of effort and funding is merited as we are working to solve an extremely challenging recognition problem. For example, it's not obvious from the surfaces of proteins how they interact specifically and how they will functionally switch conformational states. Yet, given the resources of MAGGIE and the establishment of coupled gene-biology cycles for the microbial systems investigated by this Program, we can experimentally define these modifications and interactions promoting protein functions. Long term, MAGGIE seeks to test the underlying hypothesis that the architecture of proteins and their complexes encodes in part how individual steps of pathways are coupled to each other to form coherent pathways; and how these pathways interact without disruptive interference. This idea is built upon both experimental results from the Program members and considerations rooted in first principles about the nature of macromolecular interactions.

Hyperthermophiles represent a unique group of prokaryotic microorganisms that optimally grow at temperatures exceeding those normally observed for other organisms (at or above 80°C).[1] They are

* Presenting author

biochemically and evolutionarily unique organisms that have adapted to the challenges of molecular and structural stability of their higher temperature habitats. Important biomolecules such as DNA, RNA and proteins can undergo deleterious changes at higher temperatures for most organisms.[2] However, these deleterious effects are largely suppressed for hyperthermophiles. This suggests that the unique evolutionary history and adaptation to their geo-thermal habitats has led to the development of novel mechanisms of DNA and protein stabilization. Hyperthermophiles do not perish at cooler temperatures, but undergo an adaptation response that allows them to grow in a sustained manner after an initial cold shock.[3] To better understand the biochemical changes in these unique organisms grown at their optimal 95°C and those that have adapted to 72°C, we undertook a study of the molecular changes through a systems-wide mass spectrometry based approach that involves comparative metabolite and proteomic profiling at the two environmental conditions followed by the identification of molecules that differentiate the two cell populations of *Pyrococcus furiosus*.

Metabolites were extracted from the soluble fraction of cells grown at 95°C and 72°C using cold acetone precipitation. The protein pellet from this extraction procedure was reserved for separate proteomic analysis. The extract was analyzed using LC-MS on an Agilent MSD-TOF mass spectrometer that routinely yields accurate masses (<5 ppm). LC-MS data from cells grown at both temperatures was analyzed using the XCMS program developed in the Siuzdak laboratory.[4] This software package allows for the non-linear alignment of chromatograms and the identification of significant differences between multiple samples. Using this approach, many features differentiated the two samples. Most metabolites were down regulated in samples grown at 72°C compared to 95°C. These included amino acids like arginine, phenylalanine, and leucine. However, there were some metabolites that actually underwent an up-regulation. One of these could be identified through accurate mass measurement and tandem mass. A previously unidentified molecule was discovered to be up-regulated during cold adaptation and was finally identified as N-acetylthermospermine. Thermospermine has previously been identified in other organism in Archaea.[5] The identification of these polyamines as being up-regulated during cold adaptation response is highly significant since these are known to be involved in DNA/RNA stabilization in other organisms.

In order to better understand the overall proteomic changes as furious adapts to the colder 72 °C environment, a proteomic profiling experiment was designed. Trypsin digests were prepared from the protein pellets reserved from the metabolite profiling experiment. Protein concentrations were measured using a Bradford assay and the concentrations were normalized for two samples after re-dissolving the pellets. Since the pyrococcus proteome has low cysteine content, thiol specific quantitative proteomic techniques such as isotope coded affinity tags (ICAT) were not suitable. We decided to use a spectral counting approach to relative protein quantitation using the ESI-TOF, and pursue protein identification using tandem mass spectrometry using a linear ion trap. Nano-LC-MS analyses were performed on the ESI-TOF mass spectrometer, as well as on the Finnigan LTQ using a nano-LC system on a mobile cart to minimize any differences in the chromatographic retention times. LC-MS/MS data on the ion trap was using for protein identification using Mascot (Matrix Science). This resulted in the identification of over 200 proteins. XCMS analysis was used to perform the quantitative proteomic analysis, as well as to identify the major differences between the samples grown under the two different conditions. The top 100 peptide ions observed to undergo significant change were identified through accurate mass and tandem MS experiments performed on the LTQ within narrow windows of retention time. Most proteins showed a down-regulation trend as the organism adapts to 72°C, which is consistent with the gene expression data on *P. furiosus* grown at 95°C and 72°C.[5] However, some proteins like ABC transporter showed an up-regulation.

The identification of metabolites such as spermidine, raises the question of which enzymes are involved with the regulation of these metabolites in hyperthermophiles. An experiment was designed

to do a targeted proteomic analysis using immobilized spermidine as a probe. Affinity columns were prepared by covalently binding spermidine to beads (Microlink kit from Pierce). Control columns were also prepared without spermidine. These were incubated overnight with the cytoplasmic fraction from the sample grown at 72°C. After several washes to remove non-specific binding, the nano-LC-MS/MS analysis of the eluted protein samples after trypsin digestion resulted in the identification of over a hundred proteins when compared to the control. Proteins identified included an putative acetyl transferase, archeal histones and three proteins previously identified as being up-regulated in the DNA microarray experiments.[5]

In this study, we demonstrate a comprehensive approach to metabolite and proteomic profiling using mass spectrometry that not only allows the identification of a novel metabolite, but also the characterization of changes in the proteome. Finally the metabolite immobilization and proteomics approach allows us to discover protein candidates that may help regulate key metabolites in the cold adaptation response.

References

1. Stteter K (1996). Hyperthermophilic prokaryotes. *FEMS Micro. Biol. Rev*. 18: 149-158.
2. Marmur J., and Doty, P. (1962) Determination of the base composition of deoxyribonucleic acid from its thermal denaturation temperature. *J. Mol Biol* 5, 109-118.
3. Dennis W. Grogan (1998). Hyperthermophiles and the problem of DNA instability. *Molecular Microbiology*. 28(6),1043-1049.
4. Smith C.A., Want E.J., O'Maille G., Abagyan R., Siuzdak G. (2006). XCMS: Processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching and identification. *Analytical Chemistry* 78, 779-787.
5. Weinberg, M. V., Schut, G. J., Brehm, S., Datta, S., Adams M. W.W. (2005), Cold shock of a hyperthermophilic archaeon: *Pyrococcus furiosus* exhibits multiple responses to a suboptimal growth temperature with a key role for membrane-bound glycoprotein. *J. Bacteriology* 187 (1), 336-348.

79 GTL

High Throughput Comprehensive and Quantitative Microbial and Community Proteomics

Richard D. Smith* (rds@pnl.gov), Joshua N. Adkins, David J. Anderson, Kenneth J. Auberry, Mikhail E. Belov, Stephen J. Callister, Therese R.W. Clauss, Jim K. Fredrickson, Xiuxia Du, Kim K. Hixson, Navdeep Jaitly, Gary R. Kiebel, Mary S. Lipton, Eric A. Livesay, Anoop Mayampurath, Matthew E. Monroe, Ronald J. Moore, Heather M. Mottaz, Carrie D. Nicora, Angela D. Norbeck, Daniel J. Orton, Ljiljana Paša-Tolić, Kostantinos Petritis, David C. Prior, Samuel O. Purvine, Yufeng Shen, Anil K. Shukla, Aleksey V. Tolmachev, Nikola Tolić, Harold R. Udseth, Rui Zhang, and Rui Zhao

Biological Sciences Division, Pacific Northwest National Laboratory, Richland, Washington

Project Goals: Our primary goals are to significantly increase analysis throughput and data quality, and capitalize on these advances to enable the study of increasingly complex microbial communities.

Significance: Capabilities for quantitative proteomics measurements of steadily increasing throughput and quality have been implemented and are being applied to studies of diverse microbial systems, and more increasingly to microbial communities.

* Presenting author

With recent advances in whole genome sequencing for a growing number of organisms, biological research is increasingly incorporating higher-level "systems" perspectives and approaches. For example, in nature, microbial cells rarely exist as individual colonies, but interact with other microbes in a community and with their environment, thus creating an ecosystem. The challenges of studying these higher-levels systems, such as a microbial ecosystem, are effectively open-ended due to the complexity of microbial communities, the number of possible interactions, and the technology that allows us to more completely observe complex systems. Advancing a systems-level understanding of microbial and other biological research is at the heart of the DOE Genomics:GTL program.

One aspect pertinent to a systems-level understanding is the ability to quantitatively measure the array of proteins (i.e., the proteome) under stable and perturbed conditions and from naturally occurring microbial communities. Among the challenges associated with making useful comprehensive proteomic measurements are identifying and quantifying large sets of proteins whose relative abundances span many orders of magnitude. Additionally, these proteins may vary broadly in chemical and physical properties, have transient and low levels of modifications, and be subject to endogenous proteolytic processing. Ultimately, such measurements and the resulting insight into biochemical processes are expected to enable development of predictive computational models that could profoundly affect environmental clean-up, understandings related to climate, and energy production by e.g., providing a more solid basis for mitigating the impacts of energy production-related activities on the environment.

A "prototype high throughput production" lab established in FY 2002 was an early step towards implementing higher throughput proteomics measurements. Operations within this lab are distinct from technology development efforts, both in laboratory space and staffing. This step was instituted in recognition of the different staff "mind sets" required for success in these different areas, as well as to allow "periodic upgrades" of the technology platform in a manner that does not significantly impact its production operation. The result has been faster implementation of technology advances and more robust automation of technologies that improve overall effectiveness.

The biological applications of the technology and associated activities are the subject of a separate, but interrelated project (J. K. Fredrickson, PI), involving studies of a number of individual microbial systems (e.g., *Shewanella oneidensis*, *Geobacter sulfurreducens, Rhodobacter sphaeroides, Caulobacter crescentus*) and communities (e.g., SAR 11 marine community) in collaboration with leading experts. These studies have demonstrated the capability for automated high-confidence protein identifications, broad proteome coverage, and for exploiting both stable isotope labeling and label-free methods to obtain high precision in protein abundance measurements.

With a paradigm established for high throughput proteomic measurements, our primary goals are to significantly increase analysis throughput and data quality, and capitalize on these advances to enable the study of increasingly complex microbial communities. A significant challenge is how to maximize the information content derived from large and complex data sets to obtain improved understanding of biological systems. Thus, a key component of our program involves developing the informatics tools needed to quantify and define the quality of data, as well as the tools to make the results broadly available and understandable to the researchers. Efforts currently in progress aim to:

- Significantly increase the overall data production by more than an order of magnitude in conjunction with increased data quality, providing data that are quantitative and have statistically-based measures of quality.
- Extend the application to an increasing number of different kinds of post-translation modifications.

- Apply improved data quality and improved sample processing for high throughput measurements of increasingly complex microbial communities.
- Provide the infrastructure and informatics tools required to efficiently manage, use, and disseminate large quantities of data generated by GTL "users."

This presentation will highlight the advances in providing high quality data with statistically-founded measures of quality, while providing increased measurement throughput. The advances will be illustrated in the context of applications to microbes and microbial communities of interest to the GTL program.

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

80 GTL

Exploring the Genome and Proteome of *Desulfitobacterium hafniense* DCB-2 for its Protein Complexes Involved in the Reduction of Selenium and Iron

Christina Harzman,[1] Christi Hemming,[1] Sang-Hoon Kim,[1] David DeWitt,[1] John Davis,[2] Rachel Udelhoven,[3] Kaitlin Duschene,[3] Joan B. Broderick,[3] **James M. Tiedje**,[1] Terence L. Marsh[1]* (marsht@msu.edu)

[1]Michigan State University, East Lansing, Michigan; [2]Columbus State University, Columbus, Georgia; and [3]Montana State University, Bozeman, Montana

Project Goals: The goal of our investigations into the cell biology of *Desulfitobacterium hafniense* is to identify the functional genomics/proteomics of metal reduction and determine the chemistry of relevant enzymatic activities. To this end we have investigated the regulation of growth under environmental conditions and have identified the genes and proteins up and down-regulated when metals were used as terminal electron acceptors. The identification of putative pathways and proteins through gene arrays and proteomics under metal reducing conditions will be confirmed using genetic knock-outs and putative activities explored through cloning and overexpression. In this way, the response of *D. hafniense* to its environment will be better understood and approaches to successfully employ it as an ally in bioremediation can be designed.

Desulfitobacterium hafniense is an anaerobic, low GC Gram-positive, spore-forming rod with remediation capabilities that include chlororespiration and metal reduction. The latter has been of interest to us as we have previously reported that the metal reduction capabilities of the organism include Fe(III), Cu(II), Co(III), Se(VI), and U(VI). We have also demonstrated that *D. hafniense* can form biofilms under metal-reducing and fermentative conditions. Our investigations of selenium reduction have revealed morphological changes to the cell, especially on the surface. In response to 1 mM Se(VI), *D. hafniense* cells elongate and form vesicular blebs on the cell's surface. The vesicles appear to eventually bud off of the cell and can be seen as spherical entities in the milieu with a diameter of approximately 0.2 μm (See Fig.1). These vesicles are bound by both membrane and cell wall and contain high concentrations of selenium as judged by energy dispersive x-ray spectroscopy. Hence, we have postulated that these vesicles detoxify through sequestration of the selenium.

* Presenting author

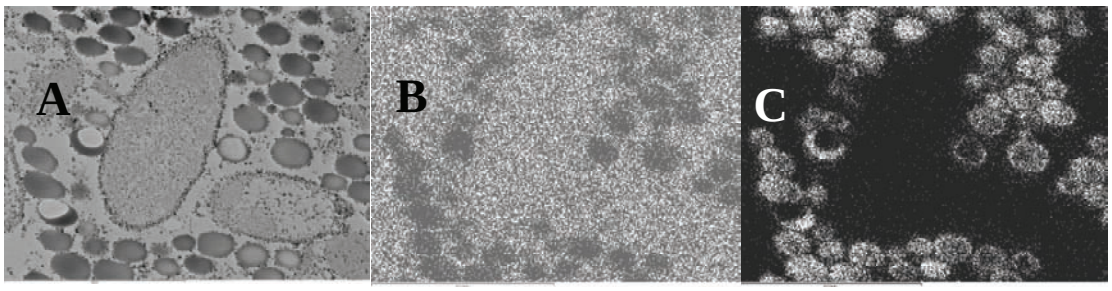


Figure 1. *D. hafniense* grown fermentatively in the presence of 1 mM selenium.

A. negatively stained transmission EM showing two cells surrounded by smaller vesicles, B. same field scanned for carbon; note low carbon content of vesicles, C. same field scanned for selenium;

Genomics

The genome of *D. hafniense* has been sequenced and we have constructed a gene array containing 4,715 targeted ORFs. Towards a more detailed understanding of the components required for metal reduction, we have performed transcriptome and proteome analysis of *D. hafniense* growing fermentatively with and without 1 mM selenium. In the presence of selenium, 27 putative operons were up-regulated and 34 were down-regulated. A total of 304 genes were up-regulated and 376 down-regulated at least 2 fold. Notable up-regulated genes coded for antitoxin RelB, transcriptional regulators, radical SAM, ABC transporters, ferredoxins, inorganic ion transport, cytochromes, DMSO reductases, heavy metal translocation, and two MinD-like proteins similar to cell septation protein MinD. Down-regulated genes included genes for NADH dehydogenase subunits, coenzyme metabolism, flagella biosynthesis, ABC transporters, amino acid transport and metabolism. In addition, many hypothetical genes were down-regulated severely.

The formation of vesicles in response to selenium is consistent with altered regulation of *minD,* which is involved in cell septation and minicell formation. Attempts are being made to overproduce the MinD-like protein in *D. hafniense* in order to test its effect on vesicle formation with and without Se(VI). The machinery capable of reducing Se(VI) is being examined by constructing knock-out mutants defective in the neighboring [Fe-S] protein genes that are likely co-transcribed with the minD-like genes. Finally, a lipoprotein transporter (LolCDE) that is usually found in Gram-negative bacteria has been detected in *D. hafniense* as transcriptionally active with selenium. This is being investigated with gene disruption approaches.

In contrast to fermentative growth, 72 operons were up-regulated and 74 were down-regulated when Fe(III) was used as the terminal electron acceptor. A total of 678 genes were up-regulated and 643 genes were down-regulated relative to fermentative growth. Up-regulated groups include: lactate transport, ferrodoxins, flavoproteins, cytochromes, radical SAM, DMSO reductases, nitrogenases, cell envelope biogenesis, heat shock proteins, heavy metal translocation, antitoxin genes (*relB*), and drug resistance genes.

Proteomics

A MarR transcription factor, a histidine kinase, and multiple proteins from each of three operons have been identified by proteomics analyses as elevated in *D. hafniense* cells grown in 1 mM Se(VI), and are proposed to be involved in selenium respiration by this organism. The MarR transcription factor and histidine kinase were elevated 65- and 25-fold respectively. The role of the MarR during

selenium reduction is not known, but histidine kinases have been implicated in the elevated expression of reductive dehalogenases in *D. hafniense*. A nickel dependent hydrogen oxidase subunit and b-type cytochrome subunit belonging to an operon similar to that shown previously to provide reducing equivalents for dehalogenation in *D. dehalogens* were also elevated. Four proteins in a DMSO reductase operon including a periplasmic (twin-arginine signal) anaerobic selenocysteine containing dehydrogenase were also found to be elevated. Four genes in a third operon, two of which code for the molybdenum cofactors necessary for DMSO reductase activity were elevated on average 5 fold at the protein level and 5-16 fold by microarray analysis. These studies suggest that respiratory reduction of selenium occurs via a DMSO reductase, using electrons provided by hydrogen oxidation, and provide an explanation for the unique function of one of the more than 50 DMSO operons in *D. hafniense*.

***In vitro* biochemical analysis**

Proteins identified by proteomics and genomics analysis as relevant to metal reduction by *D. hafniense* are being cloned for expression in *E. coli*, in order to isolate sufficient protein for detailed functional and biochemical analysis. Initial targets include DMSO reductases and sulfite reductases, as these are in some cases significantly up-regulated and are likely catalysts for selenate reduction. Other targets include several of the putative radical-SAM proteins observed to be up-regulated; some of these are likely involved in biosynthesis of the molybdopterin cofactors found in enzymes such as DMSO reductase and sulfite reductase. Target genes have been amplified from *D. hafniense* genomic DNA by PCR, and have been inserted into expression vectors using the Gateway system. Preliminary characterization of the heterologously expressed proteins will be presented.

Section 2
Metabolic Network Experimentation and Modeling

81 MEWG

Improving the Production of Biotherapeutics using Metabolic Engineering

M. Bauman,[1]* J. Jones,[1] S. Krag,[3] V. Ciccarone,[4] D. Judd,[5] S. Gorfien,[5] Y.C. Lee,[2] N. Tomiya,[2] and **M. Betenbaugh**[1]* (beten@jhu.edu)

[1]Departments of Chemical and Biomolecular Engineering, [2]Biology, and [3]Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland; [4]Macrogenics Corporation, Rockville, Maryland; and [5]Gibco, Invitrogen Corporation, Grand Island, New York

Project Goals: We are investigating the metabolic pathways involved in protein N-glycosylation. Glycoproteins that are expressed in mammalian cell culture systems often have variable N-glycosylation and we are evaluating the potential causes of this phenomenon. Using metabolic engineering strategies, we plan to overcome any bottlenecks in the pathways that may lead to inefficient or improper N-glycosylation, and therefore to improve the yield and quality of biotherapeutics that can be produced in mammalian cell culture systems.

* Presenting author

Many of the valuable commercial biotherapeutic products, such as monoclonal antibodies, growth factors, hormones and clotting factors, are secreted glycoproteins. These biotherapeutics are often produced in mammalian cell expression systems and are made up of oligosaccharide chains attached to polypeptides at specific amino acid residues. The addition of these oligosaccharides to the proteins occurs through a post-translational modification called N-glycosylation. The number, type, and location of the oligosaccharides (glycans) on the protein can affect key biochemical properties of the biotherapeutic, including its clearance rate, immunogenicity, bioactivity, solubility, and stability against proteolysis. Unfortunately, when these therapeutic products are produced by over-expression in mammalian and non-mammalian hosts, the glycosylation processing can generate products with highly variable glycosylation patterns. This N-glycan variability limits the yield and affects the quality of the target secreted glycoproteins and therefore can significantly affect the value of biotherapeutic products.

In an attempt to overcome the problems associated with variable N-glycosylation, we are investigating the metabolic pathways involved and evaluating potential causes of this phenomenon. Specifically, we are examining the N-linked glycosylation process that involves the transfer of a preformed oligosaccharide onto an acceptor Asparagine residue on a nascent polypeptide in the lumen of the endoplasmic reticulum (ER). Unfortunately, these N-glycan acceptor sites are not always fully occupied, leading to site occupancy heterogeneity. Using human transferrin (hTf) as a model protein, we have shown that variable site occupancy occurs when over-expressing certain glycoproteins using two different mammalian cell culture systems. Examination of hTf using SDS-PAGE as well as MECC (Micellar Electrokinetic Chromatography) reveals a difference in the molecular weight profile between the intracellular and secreted fractions. Treatment with tunicamycin (a glycosylation inhibitor) abrogates this difference, implying that N-glycosylation is responsible for the size difference. Immunoprecipitation experiments suggest that, instead of being secreted, under-glycosylated hTf interacts with ER chaperone proteins and accumulates intracellularly. We plan to use metabolic engineering strategies to overcome bottlenecks in the N-glycosylation pathway that lead to the formation of under-glycosylated proteins in mammalian cell culture systems.

82 GTL

Improved Microbial Hydrogen Production by the Engineering of Specific Metabolic Segments of *Escherichia coli*

Zhanmin Fan,[1] Ling Yuan,[1] Yu Wang,[2] and **Ranjini Chatterjee**[2]* (rchatterjee@farasis.com)

[1]Department of Agronomy, University of Kentucky, Lexington, Kentucky and [2]Farasis Energy, Inc., Hayward, California

Project Goals: 1. The development of *Escherichia coli* as a platform microbe for the production two clean fuels, hydrogen and ethanol, and the commodity chemical succinic acid. 2. The construction of a unique high throughput gas chromatography (GC) instrument for detection of hydrogen in micro-scale fermentations.

Hydrogen (H_2) has significant potential as a clean energy source to replace non-renewable and polluting fossil fuels. Biological routes to H_2 production represent environmentally benign processes that utilize renewable raw material derived from agricultural products, and microbes or enzymes as the catalysts for energy generation. Development of *Escherichia coli* (*E. coli*) as a biocatalyst for H_2-production offers solutions to some of the challenges facing biological processes for H_2. An endogenous H_2-evolving activity exists in *E. coli* under fermentative conditions: the formate hydrogenlyase

(FHL) enzyme complex catalyzes the disproportionation of formate to CO_2 and H_2, and is the subject of our research. The production of a functional FHL complex was improved by systematically engineering key regulatory, and amino acid biosynthetic pathways of *E. coli*. The engineered *E. coli* strains produced elevated levels of H_2 compared to the wild-type (WT) strain. The results from this program provide valuable insights into a panel of auxiliary proteins and metabolic pathways that can be engineered to increase H_2 production by microbes.

83

Toward the Automatic Generation of Genome-Scale Metabolic Models in the SEED

Matthew DeJongh[1]* (dejongh@hope.edu) and Aaron Best[2]

[1]Department of Computer Science and [2]Department of Biology, Hope College, Holland, Michigan

Project Goals: We are developing a method for automating the generation of genome-scale metabolic models suitable for analytical techniques such as flux-balance analysis. We are implementing our approach within the SEED, a software environment for comparative genome annotation and analysis (www.theseed.org). Our technology sets the stage for the automatic generation of substantially accurate metabolic reconstructions for over 400 complete genome sequences currently in the SEED.

Current methods for the automatic generation of genome-scale metabolic models focus heavily on genome annotation and preliminary biochemical reaction network assembly, but do not adequately address the process of identifying and filling gaps in the reaction network, and verifying that the network is suitable for systems level analysis. Thus, current methods are only sufficient for generating draft-quality models, and refinement of the reaction network is still largely a manual, labor-intensive process [1].

We have developed a method for automating the generation of genome-scale metabolic models that produces substantially complete reaction networks, suitable for analytical techniques such as flux-balance analysis. Our method partitions the reaction space of central and intermediate metabolism into discrete, interconnected components that can be assembled and verified in isolation from each other, and then integrated and verified at the level of their interconnectivity. We have developed a database of components that are common across organisms, and have created tools for automatically assembling appropriate components for a particular organism based on the metabolic pathways encoded in the organism's genome. This focuses manual efforts on those portions of an organism's metabolism that are not yet represented in the database. We have demonstrated the efficacy of our method by reverse-engineering and automatically regenerating the reaction network from a published genome-scale metabolic model for *Staphylococcus aureus* [2]. Additionally, we have created initial reconstructions of three other published metabolic models (*Escherichia coli* [3], *Helicobacter pylori* [4], and *Lactococcus lactis* [5]) to demonstrate that our approach reduces the manual effort involved in model creation, by building on the common reaction network components already created for the *S. aureus* model. We have implemented our tools and database within the SEED, a software environment for comparative genome annotation and analysis (www.theseed.org) [6].

Our technology sets the stage for the automatic generation of substantially accurate metabolic reconstructions for over 400 complete genome sequences currently in the SEED. With each genome that

* Presenting author

is processed using our tools, the database of common components grows to cover more of the diversity of metabolic pathways, further reducing the manual effort involved in generating subsequent genome-scale metabolic models for other sequenced organisms.

Research sponsored by Argonne National Laboratory and Howard Hughes Medical Institute.

References

1. Francke, C., R.J. Siezen, and B. Teusink (2005). "Reconstructing the metabolic network of a bacterium from its genome." *Trends Microbiol*. **13**(11): p. 550-8.
2. Becker, S.A. and B.O. Palsson (2005). "Genome-scale reconstruction of the metabolic network in *Staphylococcus aureus* N315: an initial draft to the two-dimensional annotation." *BMC Microbiol*. **5**(1): p. 8.
3. Reed, J.L., et al. (2003). "An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR)." *Genome Biol*. **4**(9): p. R54.
4. Schilling, C.H., et al. (2002). "Genome-scale metabolic model of *Helicobacter pylori* 26695." *J Bacteriol*. **184**(16): p. 4582-93.
5. Oliveira, A.P., J. Nielsen, and J. Forster (2005). "Modeling *Lactococcus lactis* using a genome-scale flux model." *BMC Microbiol*. **5**: p. 39.
6. Overbeek, R., et al. (2005). "The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes." *Nucleic Acids Res*. **33**(17): p. 5691-702.

84 GTL

Metabolic Engineering of Light and Dark Biochemical Pathways in Wild-Type and Mutant *Synechocystis* PCC 6803 Strains for Maximal, 24-Hour Production of Hydrogen Gas

P. S. Schrader,[1] E. H. Burrows,[2] F. W. R. Chaplen,[2] and **R. L. Ely**[2]* (ely@engr.orst.edu)

[1]Yale University, New Haven, Connecticut and [2]Oregon State University, Corvallis, Oregon

Project Goals: The objectives of the proposed research are addressed in the following four tasks: 1. Evaluate the effects of various culture conditions (N, S, or P limitation; light/dark; pH; exogenous organic carbon) on H_2 production profiles of WT cells and an NDH-1 mutant; 2. Conduct metabolic flux analyses for enhanced H_2 production profiles using selected culture conditions and inhibitors of specific pathways in WT cells and an NDH-1 mutant; 3. Create PCC 6803 mutant strains with modified H_2ases exhibiting increased O_2 tolerance and greater H_2 production; 4. Integrate enhanced H_2ase mutants and culture and metabolic factor studies to maximize 24-hour H_2 production.

Global power demand is projected to increase from the current 13 terawatts (TW) to 46 TW by the end of this century (US DOE, 2005). Though fossil fuels will certainly continue to dominate the growth of the energy sector in the near term (E.I.A., 2006), a renewable alternative must be pursued now to prevent severe economic disruptions from a transition that may take decades to complete (Hirsch, 2005). Solar energy, harvested directly in the form of hydrogen gas (H_2) from the splitting of water, offers a promising alternative to meet regional, national, and global energy demand in a sustainable, environmentally friendly manner.

We are using the cyanobacterial species *Synechocystis* sp. PCC 6803 to pursue two initial project goals: 1) Optimize H_2 production conditions through a series of H_2 profiling experiments varying the light/

dark ratio, pH, exogenous organic carbon, and nitrogen, sulfur, and phosphorus concentrations, and 2) Conduct metabolic flux analysis to quantify intracellular reductant fluxes and to identify obstructions to H_2 production, specifically to identify factors that decrease intracellular levels of NADPH to support H_2 production. Both of these goals are being addressed with wild-type cells as well as with a high H_2-producing mutant with impaired type I NADPH-dehydrogenase (NDH-1) function.

Results from the **H_2 profiling experiments** show that pH appears to have a dramatic effect on H_2 production by *Synechocystis* sp. PCC 6803. Over pH values ranging from 5 to 10, we have observed significantly higher H_2 production at higher pH, both by WT cells and by the NDH-1 mutant. With WT cells, this effect seems especially pronounced when N is high (18 mM or 35 mM), whereas high P (180 μM and 360 μM) seems to dampen this effect. In contrast, with the NDH-1 mutant, the effect is pronounced when N is limiting, and we have seen mixed results with P. At higher pH, there may be a trend toward higher H_2 production at higher P concentrations, but the data we have generated so far have not shown it to be statistically significant. At neutral pH, our preliminary results suggest that slightly higher H_2 production may be occurring under N-limited conditions in WT cells, which would be consistent with the observations of Schutz et al. (2004), who found that nitrogen-limitation (1mM N) increased H_2 photoproduction in *Synechocystis* sp. PCC 6803 over 48 hours. Results of screening tests conducted so far have not shown a similar trend in the NDH-1 mutant. Also, any tendencies toward higher H_2 production caused by N limitation seem to be overwhelmed by pH effects at higher pH values, as described above. Our initial results suggest that the optimal concentration of $NaHCO_3$ for H_2 production, from the concentrations tested, is 80 mM. Tsygankov et al. (2002) found that sulfur deprivation increased H_2 photoproduction in *Chlamydomonas rheinhardtii*, so we have anticipated that sulfur deprivation may have a similar effect in *Synechocystis* PCC 6803. Preliminary results regarding the effects of S concentration on H_2 production have been mixed, and we are looking into this question further. In evaluating light/dark effects, we have tested 12 and 24 light-dark cycles per day so far, with the dark/light ratio varying from 8:1 to 29:1 based on published values of the respiration/photosynthesis ratio. Initial results suggest that a regimen consisting of 24 cycles per day with a dark/light ratio of 8:1 produces the highest H_2 production. We are working on an advanced optimization algorithm that will predict the optimal conditions for hydrogen production, based on the data collected.

Metabolic flux analysis is a linear algebra technique used to determine the fluxes among a network of intracellular metabolites, based on measured inflows and outflows from the cells, and based on the assumption that the system is at steady-state. To achieve true metabolic steady-state, we are conducting the metabolic flux analysis experiments in a chemostat. We are working to achieve a sinusoidal steady-state because the chemostat is exposed to light/dark cycling, which is needed for H_2 production. The network chosen for photoautotrophic growth consists of 20 constraints and 24 reactions, thus requiring at least 4 measurements. The inflows and outflows measured are H_2, O_2, CO_2, glucose, glycogen, ammonium, and biomass production/consumption. Intracellular NADPH/NADP+ concentrations and light intensity are also measured.

References

1. E.I.A. 2006. International Energy Outlook 2006, Energy Information Administration, U.S. Department of Energy.
2. Hirsch, R. L. 2005. Testimony on Peak Oil, House Subcommittee on Energy and Air Quality, U.S. House of Representatives.
3. Schutz, K., T. Happe, O. Troshina, P. Lindblad, E. Leitao, P. Oliveira, & P. Tamagnini. 2004. Cyanobacterial H_2 production - a comparative analysis. Planta. **218**(3):350-359

* Presenting author

4. Tsygankov, A., S. Kosourov, M. Seibert, & M.L. Ghirardi. 2002. Hydrogen photoproduction under continuous illumination by sulfur-deprived, synchronous *Chlamydomonas reinhardtii* cultures. Int. J. Hydrogen Energy. **27**(11-12):1239-1244

5. U.S. DOE. 2005. Basic Research Needs for Solar Energy Utilization. Report of the Basic Energy Sciences Workshop on Solar Energy Utilization. US Department of Energy. Argonne National Laboratory Publication 2005.

85 GTL

Pathway Tools + MetaCyc = Comprehensive Pathway Modeling

Ron Caspi,[1] Hartmut Foerster,[2] Carol Fulcher,[1] Michelle Green,[1] Pallavi Kaipa,[1] Markus Krummenacker,[1] Mario Latendresse,[1] Suzanne Paley,[1] Chris Tissier,[2] Peifen Zhang,[2] Sue Rhee,[2] and **Peter D. Karp**[1]* (pkarp@ai.sri.com)

[1]Bioinformatics Research Group, SRI International, Menlo Park, California and [2]Carnegie Institution, Department of Plant Biology, Stanford, California

Project Goals: The goal of the MetaCyc project is to develop a comprehensive, universal database of experimentally derived information on metabolic pathways and enzymes from many organisms. The goal of the Pathway Tools project is to facilitate understanding of metabolic and regulatory networks, and their relationship to the genome, by developing computational tools for inference, querying, visualization, and analysis of databases that integrate pathway and genome information.

Metabolic engineering demands an accurate model of the metabolic network of a target organism and the relationship of that network to the genome, plus powerful analysis tools for constructing, refining, and analyzing that model.

The MetaCyc multiorganism pathway database [1,2] describes experimentally elucidated metabolic pathways and enzymes reported in the experimental literature. MetaCyc is both an online reference source on metabolic pathways and enzymes for metabolic design, and a solid foundation of experimentally proven pathways for use in computational pathway prediction. MetaCyc version 10.6 describes 890 pathways from more than 900 organisms. The 6100 biochemical reactions in MetaCyc reference 6000 chemical substrates, most of which contain chemical structure information. MetaCyc describes the properties of 3500 enzymes, such as their subunit structure, cofactors, activators, inhibitors, and in some cases their kinetic parameters. The information in MetaCyc was obtained from 12,000 research articles, and emphasizes pathways and enzymes from microbes and plants, although it also contains animal pathways.

Pathway Tools [3,4] constructs a metabolic model of an organism from its annotated genome using the following computational inference tools. The model is in the form of a Pathway/Genome Database (PGDB).

- It predicts the metabolic pathways of the organism by recognizing known pathways from the MetaCyc database
- It predicts which genes fill holes in those metabolic pathways (pathway holes are pathway steps for which no enzyme has been identified in the genome)
- It predicts operons for prokaryotic genomes

- It infers the presence of transport reactions from the names of transport proteins in the genome annotation
- The software automatically generates a one-screen cellular overview diagram containing the metabolic and transport networks of the cell

A set of graphical editors within Pathway Tools allows scientists to refine a PGDB by adding, or modifying metabolic pathways, gene annotations, reactions, substrates, and regulatory information. The existence of an accurate knowledge base of the metabolic network is a critical resource for metabolic engineering.

The software provides a large number of operations for querying, visualization, web publishing, and analysis of PGDBs. New network debugging tools allow the user to find errors or incompleteness in the metabolic model by identifying dead-end metabolites, and mismatches between the transport and metabolic subsystems. These tools can speed the identification of errors in the genome annotation and in the metabolic model.

A new metabolite tracing tool supports graphical exploration of the path that a substrate follows through the metabolic network, in either the forward or backward direction. The user interactively guides the software in selecting which branches of metabolism to follow, and metabolic paths are highlighted on the cellular overview diagram. A new tool for graphical construction of complex database queries provides a quantum leap in the power of database queries that a user can construct without knowledge of SQL.

A family of omics viewers support systems-level visualization of large-scale datasets onto cellular networks. The first omics viewer paints omics datasets onto the cellular overview of metabolic and transport networks. The second (new) omics viewer paints omics datasets onto a diagram of the transcriptional regulatory network. The third (new) omics viewer paints omics datasets onto the genome. These tools provide complementary perspectives for interpreting omics data.

A set of new comparative genomics tools supports many comparisons across the genomes and metabolic networks of a set of organism's PGDBs. For example, the pathway complements of selected PGDBs can be compared, with the results ordered by pathway ontology.

Other visualization tools include automated display of metabolic pathways, reactions, enzymes, genes, and operons, and a genome browser.

More than 75 groups are using Pathway Tools and MetaCyc to produce PGDBs for more than 150 organisms, including the major model organisms for biomedical research (yeast, worm, fly, *Dictyostelium*), pathogens of biodefense interest, GTL organisms, many other bacteria and archaea, and plants (including *Arabidopsis*, *Medicago*, Rice, Tomato, and Potato).

References

1. R. Caspi et al, "MetaCyc: A multiorganism database of metabolic pathways and enzymes," *Nucleic Acids Research* 34:D511-6 2006.
2. P. Zhang et al, "MetaCyc and AraCyc. Metabolic Pathway Databases for Plant Research," *Plant Physiol* 138:27-37 2005.
3. P.D. Karp et al, "The Pathway Tools Software," *Bioinformatics* 18:S225-32 2002.
4. Paley, S.M. et al, "The Pathway Tools Cellular Overview Diagram and Omics Viewer," *Nucleic Acids Research* 34:3771-8, 2006.
5. Paley, S.M. et al, "Creating Fungal Pathway/Genome Databases Using Pathway Tools," *Applied Mycology and Biotechnology* 6:209-26 2006.

* Presenting author

86 GTL

Constraint-Based Modeling of Central Metabolism in the Family *Geobacteraceae*

Jun Sun[1]* (jsun@genomatica.com), Steve Van Dien,[1] Radhakrishnan Mahadevan,[2] Maddalena Coppi,[3] Laurie DiDonato,[3] Carla Risso,[3] Mounir Izallalen,[3] Bradley Postier,[3] Raymond DiDonato,[3] Kai Zhuang,[2] Priti Pharkya,[1] Tom Fahland,[1] Olivia Bui,[1] Iman Famili,[1] Christophe Schilling,[1] and **Derek Lovley**[3]

[1]Genomatica Inc., San Diego, California; [2]University of Toronto, Toronto, Ontario; and [3]University of Massachusetts, Amherst, Massachusetts

Project Goals: The ultimate goal of the *Geobacter* Project is to develop genome-based in silico models that can be used both to interpret environmental gene expression data in environments in which *Geobacter*ace*ae* predominate and to predict the growth and metabolism of *Geobacteraceae* in situ using routine geochemical measurements as input. Such models will also enable prediction of the outcome of various potential manipulations that might be made to optimize processes of interest, such as in situ uranium bioremediation and harvesting electricity from waste organic matter, prior to conducting costly and labor-intensive field experiments.

As part of an effort to predictively model the behavior of *Geobacteraceae* involved in bioremediation and electricity-generation *in situ*, *in silico* reconstructions of the metabolic networks of three *Geobacter* species were generated by the constraints-based approach.

The genome-scale metabolic model of *Geobacter sulfurreducens,* the genetically tractable and acetate-oxidizing species, was the first to be generated. In the past year, ^{13}C isotopic labeling studies were performed to quantitate actual flux distributions and to further refine the *G. sulfurreducens* model. Based on the labeling patterns from acetate:fumarate chemostat cultures, overall amino acid labeling patterns were consistent with flux distributions generated by the model. Computational flux analysis on the levels of flux through the various phosphoenol pyruvate (PEP) synthesizing pathways indicated that PEP was derived from both acetate and oxaloacetate, despite the energetic differences. One surprising prediction by modeling was an ATP-consuming futile cycle that involved three enzymes catalyzing the interconversion of pyruvate, oxaloacetate and PEP. The futile cycle was confirmed by the labeling data. This futile cycle may only be active during growth on fumarate, which is not a natural electron acceptor. The possibility that adaptive evolution on fumarate will lead to the inactivation of this futile cycle is currently under investigation.

^{13}C-labeling studies also led to the discovery of an alternate pathway for the biosynthesis of isoleucine in *G. sulfurreducens*. The metabolic model predicted that isoleucine was synthesized from aspartate and pyruvate, but the labeling pattern of isoleucine did not match the prediction. Further flux analysis suggested that the majority of isoleucine was synthesized exclusively from pyruvate and/or acetyl-CoA potentially via a citramalate pathway that was found in several species of methanogens and *Leptospira interrogans*. The citramalate synthase activity was indeed detected in soluble *G. sulfurreducens* extracts and a candidate citramalate synthase gene was identified. Optimizations performed to assess the relative contributions of the citramalate and the aspartate/pyruvate pathways to isoleucine biosynthesis indicated that 68-78% of the isoleucine was synthesized via the citramalate pathway. In addition, simulations indicated that the use of the citramalate pathway instead of the aspartate/pyruvate pathway significantly increased the efficiency of isoleucine biosynthesis during growth on acetate.

As part of an ongoing investigation of the effects of nutrient limitation and other environmentally relevant stresses on central metabolism, the metabolic model was used to estimate the energetic cost of nitrogen fixation and to predict changes in flux distribution as a result of nitrogen fixation. Studies comparing the effects of nitrogen fixation on global gene expression to predicted changes in flux distribution are underway, and a similar approach is being applied to phosphate limitation. The effect of growth rate on central metabolism is also being investigated, as the growth rates of *Geobacteraceae in situ* are much lower than those typically observed in laboratory cultures. *G. sulfurreducens* was cultivated in chemostats at a variety of growth rates spanning both the low and high end of the spectrum and changes in gene expression in response to changes in growth rate are being compared to predicted changes in flux distribution.

The engineering of *Geobacter* species to achieve increased respiration rates as a strategy for increasing electricity production is another area of investigation. Respiration rates were elevated in *G. sulfurreducens* by inducing expression of an ATP-consuming enzyme, the peripheral subunits of the F1F0-ATPase. *In silico* analysis of the metabolic network, using experimentally derived organic acid measurements as input, indicated that introduction of this ATP-drain should result in diversion of acetate from biosynthetic reactions to the TCA cycle and ATP generation. In order to assess the accuracy of these *in silico* predictions, microarray analysis was performed following induction of the ATP drain. There were many examples of changes in gene expression which were consistent with *in silico* predictions. A variety of genes involved in energy metabolism, including multiple cytochromes and electron transport proteins, TCA cycle enzymes, and subunits of the NADH dehydrogenase, were up-regulated, as was the gene encoding the dicarboxylic acid exchanger involved in fumarate uptake. In contrast, the gene encoding acetate kinase, which activates acetate for gluconeogenesis was down-regulated.

Considerable progress has been made towards the development and refinement of metabolic models for two other *Geobacteraceae*: *G. metallireducens* and *P. carbinolicus,* both of which have metabolic capabilities that are not shared by *G. sulfurreducens*. The *G. metallireducens* and *P. carbinolicus* models were initially created using an automated model reconstruction procedure, the Automodel pipelineTM and were manually curated over the past year. It was estimated that the use of the Automodel pipelineTM accelerated model development by 3.7 fold. The Automodel pipeline was also used to rapidly incorporate discoveries made during development of the two newer models into the *G. sulfurreducens* model, resulting in an increase in the total number of reactions and a decrease in the number of non-gene associated reactions. Currently, the *G. sulfurreducens*, *G. metallireducens* and *P. carbinolicus* models contain 649, 606 and 700 reactions, respectively. More than half (64% to 74%) of the reactions present in each individual model are shared among all three models. The unique reactions in the *P. carbinolicus* model include those involved in fermentation and proline biosynthesis, whereas unique reactions in the *G. metallireducens* model include those involved in the metabolism of monoaromatic compounds. In addition, the *G. metallireducens* and *P. carbinolicus* models contain a key alchohol dehydrogenase for ethanol utilization and the oxidative branch of the pentose phosphate pathway, which are not present in the *G. sulfurreducens* model. *In silico* deletion analysis was performed on all three models and revealed that ca. 200 reactions were essential for fermentative growth of *P. carbinolicus* on acetoin and respiratory growth of either *G. metallireducens* and *G. sulfurreducens* on acetate and Fe(III) citrate. Further computational and experimental analyses using these models will provide insight into the metabolism of these and other species of *Geobacteraceae*.

Finally, as microorganisms rarely exist in isolation in the environment, computational analysis of microbial communities has also been initiated. Microbial communities in which Fe(III) serves as the terminal electron acceptor typically include fermentative organisms, which break down complex organic matter, and *Geobacteraceae* which utilize fermentation byproducts as electron donors for

* Presenting author

Fe(III) reduction. In order to model a simplified version of such a community, a coculture consisting of *Escherichia coli* and *G. sulfurreducens,* which together can couple glucose fermentation to iron reduction, has been established. A dynamic model of this co-culture has been developed, and predictions generated by this model have been reconciled with physiological data.

87 GTL

Analysis of Degree of Genetic Redundancy in Prokaryotic Metabolic Networks

R. Mahadevan[1]* (mahadevan@chem-eng.utoronto.ca) and **D.R. Lovley**[2]

[1]University of Toronto, Ontario and [2]University of Massachusetts, Amherst, Massachusetts

Project Goals: Metabolic networks can be robust to environmental and genetic perturbations through genetic redundancy or alternate biochemical pathways. The role of these factors has been analyzed extensively in eukaryotes such as *Saccharomyces cerevisiae* and recent studies indicate a stronger role for gene duplicates in accentuating enzymatic flux rather than as a back-up function. Here, we analyze the extent of genetic and biochemical redundancy in prokaryotic metabolic networks using genome-scale metabolic models. Surprisingly, we find that the extent of genetic redundancy appears to be enriched in *Geobacter sulfurreducens* and *Methanosarcina barkeri* as compared to other organisms. Based on these findings, we suggest that the environmental niche, an organism inhabits might have a role in determining the mechanism of attaining robustness to genetic perturbations.

Robustness to perturbations is almost an intrinsic and essential component of biological systems. Several studies have extensively analyzed the robustness of biological systems and have provided insights on the mechanism of adaptation to genetic and environmental perturbations. Robustness of metabolism after the loss of a gene product can occur due to either the presence of a gene duplicate that has the same function ("gene family buffering") or due to the presence of alternate pathways that can achieve similar function ("pathway buffering"). The roles of genetic redundancy and biochemical buffering and the mechanistic principles arising from molecular interactions is of great interest to further understand robustness in biological systems. Such an understanding of the factors favoring the maintenance of duplicate genes in microbial genomes is essential for developing models of microbial evolution.

A genome-scale flux-balance analysis of the metabolic network of *Saccharomyces cerevisiae* has suggested that gene duplications primarily provide increased enzyme dosage to enhance metabolic flux because the incidence of gene duplications in essential genes is no higher than that in non-essential genes. However, *S. cerevisiae* represents just one example of a wide spectrum of microbial metabolism and is a eukaryote. Therefore, we used genome-scale metabolic models of *Escherichia coli, Bacillus subtilis, Geobacter sulfurreducens* and *Methanosarcina barkeri* to analyze the extent of genetic and biochemical redundancy in prokaryotes that are either specialists, with one major mode of energy generation, or generalists, which have multiple metabolic strategies for conservation of energy.

Genome-scale metabolic models represent the majority of the biological information ranging from genome sequence, biochemical and high-throughput physiological data and have been shown to be successful in predicting the experimentally determined deletion phenotypes. Surprisingly, the results suggest that although generalists, such as *E. coli* and *B. subtilis*, are similar to the eukaryotic general-

ist, *S. cerevisiae*, in having a low percentage (< 10 %) of essential genes and few duplications of these essential genes, metabolic specialists, such as *G. sulfurreducens* and *M. barkeri*, have a high percentage (> 30 %) of essential genes and a high degree of genetic redundancy in these genes compared to non-essential genes. The analysis of flux through the reactions with the gene duplicates reveal that they are no more likely to have a higher rate of flux than the rest of reactions further suggesting a different role for gene duplicate in specialists such as *G. sulfurreducens*. Therefore, the specialist organisms appear to rely more on gene duplications rather than alternative-but-equivalent metabolic pathways to provide resilience to gene loss. Generalists rely more on alternative pathways. Thus, the concept that the role of gene duplications is to boost enzymatic flux rather than provide metabolic resilience, may not be universal. Rather, the degree of gene duplication in microorganisms may be linked to mode of metabolism and environmental niche.

88 GTL

Mechanisms of Sulfur Reduction by *Shewanella*

Edward J. Crane III[1]* (EJ.Crane@pomona.edu), Evan T. Hall,[1] and **Ken Nealson**[2]

[1]Pomona College, Claremont, California and [2]University of Southern California, Los Angeles, California

Project Goals: This project is a component of the *Shewanella* Federation and as such contributes to the overall goal of applying the tools of genomics, leveraging the availability of genome sequence for 18 additional strains of *Shewanella*, to better understand the ecophysiology and speciation of respiratory-versatile members of this important genus. To understand these systems the SF is using genome-based approaches to investigate *Shewanella* as a system of integrated networks; first describing key cellular subsystems those involved in signal transduction, regulation, and metabolism—then building towards understanding the function of whole cells and, eventually, cells within populations. As a general approach, the SF is collectively employing complimentary top-down bioinformatics-based genome functional predictions, high-throughput expression analyses, and functional genomics approaches to uncover key genes as well as metabolic and regulatory networks. The bottom-up component employs more traditional approaches including genetics, physiology and biochemistry to test or verify predictions. This information will ultimately be linked to analyses of signal transduction and transcriptional regulatory systems and used to develop a linked model that will contribute to understanding the ecophysiology of *Shewanella* in redox stratified environments.

Shewanella is famed for its ability to respire a wide range of substrates, and sulfur is one of the substrates that many species of this genus are able to use as an electron acceptor. We are interested in characterizing the mechanisms by which *S. oneidensis* MR-1 reduces sulfur, and are characterizing the sulfur reductase complex of this organism. We will determine growth yields and electron budgets for lactate oxidized/S^0 reduced (with an N_2 atmosphere, as we have shown that H_2 is not required for growth on S^0), lactate/S^0 with H_2 (as an additional electron donor), and with electron donors such as N-acetylglucosamine, α-ketobutyrate and α-ketoglutarate. We will also characterize the mechanism of the putative polysulfide reductase complex, using a range of spectroscopic techniques. One of the difficulties associated with the *in vitro* growth of *Shewanella* with sulfur as an electron acceptor is the inhibition of the organism by the sulfide product. Simply trapping sulfide is not an option, as sulfide is also necessary to begin the reduction process by reducing insoluble sulfur to soluble polysulfide (which is believed to be the actually substrate for the sulfur reductase). We have determined

* Presenting author

that optimal sulfide concentrations for anaerobic growth are in the range of 1 mM sulfide. We have isolated membranes from MR-1 growing with sulfur, and are in the process of developing a sulfur reductase assay, as well as a method for suspending the sulfur reductase complex in polyanionic films in order to facilitate spectroscopic characterization. While comparison of the genome of several *Shewanella* species to those of other sulfur reducers has revealed the enzymes most likely to be involved in the direct reduction of S^0 and/or polysulfide, it is much less clear which pathways feed electrons to these enzymes and how the enzymes and complexes within the system interact. Extensive screening of mutants of the c-type cytochromes has revealed the majority of these proteins, including the Mtr and Omc heme proteins involved in metal reduction, as well as ΔSO4144, which has been shown to be essential for the reduction of tetrathionate, are not essential for sulfur reduction. Of these proteins, only ΔSO2930 and ΔSO2931 appear to show any decrease in ability to reduce sulfur.

89 GTL

Carbon and Energy Metabolism Strategies in *Shewanella*

G. Pinchuk[1]* (Grigoriy.Pinchuk@pnl.gov), A. Beliaev,[1] O. Geydebrekht,[1] D. Kennedy,[1] I. Famili,[2] J. Reed,[2] J. Scott,[3] S. Reed,[1] M. Romine,[1] and J. Fredrickson[1]

[1]Pacific Northwest National Laboratory, Richland, Washington; [2]Genomatica, Inc., San Diego, California; and [3]Earth Sciences, Dartmouth College, Hanover, New Hampshire

Project Goals: This project is a component of the *Shewanella* Federation and as such contributes to the overall goal of applying the tools of genomics, leveraging the availability of genome sequence for 18 additional strains of *Shewanella*, to better understand the ecophysiology and speciation of respiratory-versatile members of this important genus. To understand these systems the SF is using genome-based approaches to investigate *Shewanella* as a system of integrated networks; first describing key cellular subsystems those involved in signal transduction, regulation, and metabolism—then building towards understanding the function of whole cells and, eventually, cells within populations. As a general approach, the SF is collectively employing complimentary top-down bioinformatics-based genome functional predictions, high-throughput expression analyses, and functional genomics approaches to uncover key genes as well as metabolic and regulatory networks. The bottom-up component employs more traditional approaches including genetics, physiology and biochemistry to test or verify predictions. This information will ultimately be linked to analyses of signal transduction and transcriptional regulatory systems and used to develop a linked model that will contribute to understanding the ecophysiology of *Shewanella* in redox stratified environments.

The genus *Shewanella* is unusually well-adapted to chemically (redox) stratified environments as reflected in the ability to utilize a broad range of electron acceptors via a highly diversified electron transport system. Occupying such niches requires the ability to adapt rapidly to changes in electron donor/acceptor type and availability; hence the ability to compete and thrive in such environments must ultimately be reflected in the organization and flexibility of the electron transfer networks as well as central carbon metabolism pathways. Although MR-1 is typically considered to have a relatively restricted substrate range for carbon and energy sources, genome-based analyses revealed multiple pathways for $C_{2\text{-}6}$ compounds, amino acids, and fatty acids, reflecting its ecological role as a consumer of organic matter breakdown products in relatively carbon-rich environments that support diverse anaerobic microbial communities. Using controlled cultivation, biochemical, genetic and

genomic approaches in conjunction with pathway modeling, we showed that (i) metabolic pathways in MR-1 expressed under different redox conditions utilize pyruvate as a key metabolite and (ii) the pathways involved in ATP production under aerobic and anaerobic conditions fundamentally differ reflecting the amount of energy this organism can generate by oxidative phosphorylation.

Aerobic utilization of C_{2-5} compounds through pyruvate as a central intermediate. When grown on lactate or pyruvate (which is the first product of lactate oxidation) *S. oneidensis* MR-1 displayed the highest growth rates in comparison to other substrates tested. To determine which metabolic pathways are used to metabolize different organic compounds, several MR-1 deletion mutants with genes predicted to be involved in central carbon metabolism were generated. Initially, a pyruvate dehydrogenase complex (PDHc) mutant carrying a deletion of the E1 subunit (SO0424) was tested for aerobic growth. We determined that this mutant was unable to grow on minimal medium supplemented with any single C_{2-5} compounds tested including acetate. Results obtained with other mutants tested for their ability to grow with lactate or acetate revealed that isocitrate lyase (SO1484) and malate synthase (SO1483) are not essential for growth with lactate but are indispensable for acetate metabolism. Malic enzyme (SO3855) was not required for aerobic growth with any compound tested. Acetyl-CoA synthase, but not the combined action of phosphotransacetylase and acetate kinase, was necessary for exogenous acetate utilization. Taken together, these results suggest that pyruvate is the central metabolic intermediate involved in aerobic utilization of C_{2-5} compounds by *S. oneidensis* MR-1. Additionally, our experiments strongly suggest that *S. oneidensis* does not oxidize acetate using the TCA cycle under aerobic conditions, and a new pathway responsible for aerobic acetate oxidation in *S. oneidensis* cells is proposed.

The role of substrate-level phosphorylation under anaerobic and O_2-limited growth of *S. oneidensis* MR-1.

One of the fundamental characteristics of *S. oneidensis* metabolism is its inability to use acetate as an electron donor under anaerobic conditions. We extended these previous observations by demonstrating that acetate cannot be used as carbon and energy source by *S. oneidensis* MR-1 under Fe(III)- and fumarate-reducing as well as under O_2-limited conditions. Anaerobic or O_2-limited growth with lactate as electron donor is accompanied by acetate excretion (with 80-90% lactate converted to acetate). These results implied that acetate excretion may be coupled to ATP production catalyzed by acetate kinase (SO2915). Indeed, an MR-1 acetate kinase deletion mutant did not grow anaerobically with either Fe(III)-NTA or fumarate when lactate served as the carbon and energy source. Chemostat experiments also showed that the amount of acetate produced was in inverse proportion to the O_2 flux. These results strongly suggest that under anaerobic or O_2-limited growth *S. oneidensis* MR-1 depends solely on substrate level phosphorylation for energy generation.

We also have generated several lines of evidence, including analysis using a flux balance model of *S. oneidensis* metabolism, which show that under conditions of O_2 limitation and fumarate reduction most of ATP is produced from lactate or pyruvate on the level of substrate phosphorylation (from acetyl phosphate). In contrast, the redox chain functions mostly to re-oxidize electron carriers and, in case of fumarate, does not couple electron acceptor reduction to oxidative phosphorylation. Our results suggest that for *S. oneidensis* the rate of electron transfer to a terminal electron acceptor determines the growth rate and part of energy spent on maintenance needs, whereas efficiency of electron transport coupling to phosphorylation partially determine biomass growth yield. Such flexibility of central carbon metabolism allows *Shewanella* to survive during periods of nutrient-limitation and proliferate rapidly when both electron acceptor(s) and donor(s) are available.

* Presenting author

90 GTL

Metabolic Reconstruction of *Shewanella oneidensis*: A Community Resource

Jennifer L. Reed,[1] Iman Famili,[2] **Sharon J. Wiback**,[2] Christophe H. Schilling,[2] Grigoriy Pinchuk,[3] Margaret R. Romine,[3] Johannes C. Scholten[3*] (johannes.scholten@pnl.gov), Joel Klappenbach,[4] and James K. Fredrickson[3]

[1] Department of Bioengineering, University of California, San Diego, La Jolla, California; [2]Genomatica, Inc., San Diego, California; [3]Biological Sciences Division, Pacific Northwest National Laboratory, Richland, Washington; and [4]Center for Microbial Ecology, Michigan State University, East Lansing, Michigan

Project Goals: This project is a component of the *Shewanella* Federation and as such contributes to the overall goal of applying the tools of genomics, leveraging the availability of genome sequence for 18 additional strains of *Shewanella*, to better understand the ecophysiology and speciation of respiratory-versatile members of this important genus. To understand these systems the SF is using genome-based approaches to investigate *Shewanella* as a system of integrated networks; first describing key cellular subsystems - those involved in signal transduction, regulation, and metabolism—then building towards understanding the function of whole cells and, eventually, cells within populations. As a general approach, the SF is collectively employing complimentary "top-down" - bioinformatics-based genome functional predictions, high-throughput expression analyses, and functional genomics approaches to uncover key genes as well as metabolic and regulatory networks. The "bottom-up"component employs more traditional approaches including genetics, physiology and biochemistry to test or verify predictions. This information will ultimately be linked to analyses of signal transduction and transcriptional regulatory systems and used to develop a linked model that will contribute to understanding the ecophysiology of *Shewanella* in redox stratified environments.

Genome-scale network reconstructions account for components and component interactions in biological networks, and are a way in which to collate and analyze data from a variety of sources. Here we report a metabolic reconstruction for *Shewanella oneidensis* MR-1 based on the current genome annotation and primary literature. The reconstruction includes 769 reactions, 779 genes, and 660 metabolites. The reconstruction was used to build a flux balance model that was used in a variety of computational analyses, including: assessment of growth phenotypes, evaluation of metabolite usage (as substrates or by-products), and prediction of knock-out phenotypes. The model correctly predicted growth on a variety of carbon and nitrogen sources. In addition, quantitative evaluation of alternative electron acceptors led to the identification of 7 classes of electron acceptors, with differing biomass yields (g D.W. produced per mmol electron acceptor consumed). Gene deletion simulations across 10 different environmental conditions with various carbon sources and electron acceptors found that a large fraction of genes were never essential (535 out of 779), while a smaller fraction were always essential (202 out of 779) for growth on these 10 conditions. Together this work provides a resource that can be used by *Shewanella* researchers and illustrates how reconstructions can serve as a means to evaluate experimental data and generate testable hypotheses to better understand its ecophysiology.

* Presenting author

91 GTL

The Challenge of Incorporating Regulatory Effect in Genome-Scale Networks

C.L. Barrett*, M.J. Herrgard*, B.K. Cho, E.M. Knight, J. Elkins, and **B.O. Palsson (palsson@ucsd.edu)**

Department of Bioengineering, University of California, San Diego, California

Project Goals: Genomic and bibliomic data has been used to reconstruct a number of genome-scale metabolic networks. The stoichiometric structure of these networks has enabled a series of basic and applied studies that address both proximal and distal causation in biology. One of the challenges going forward with computational models at the genome- scale is to account for regulatory effects. Regulation occurs primarily at two levels; 1) the transcriptional level, and 2) the gene product activity level. Significant progress is being made with the former issue, while the latter is still at a conceptual stage. The current state and future challenges of both issues will be discussed.

Genomic and bibliomic data has been used to reconstruct a number of genome-scale metabolic networks. The stoichiometric structure of these networks has enabled a series of basic and applied studies that address both proximal and distal causation in biology. One of the challenges going forward with computational models at the genome-scale is to account for regulatory effects. Regulation of metabolic enzymes occurs primarily at two levels; 1) the transcriptional level, and 2) the post-transcriptional (protein expression and activity) level. Significant progress is being made with characterizing, reconstructing and modeling transcriptional regulatory networks regulating metabolism. However, the development of the corresponding methods for incorporating post-transcriptional regulation into genome-scale models is still at an early stage. The current state and future challenges of incorporating both transcriptional and post-transcriptional regulation in genome-scale networks will be discussed.

92 GTL

Acclimation of *Chlamydomonas reinhardtii* to Anoxic Conditions: Gene Expression, Hydrogenase Induction and Metabolic Pathways

Michael Seibert[1]* (mike_seibert@nrel.gov), Florence Mus,[2] Alexandra Dubini,[1,3] Maria L. Ghirardi,[1] Matthew C. Posewitz,[3] and Arthur R. Grossman[2]

[1]National Renewable Energy Laboratory, Golden, Colorado; [2]Carnegie Institution of Washington, Stanford, California; and [3]Colorado School of Mines, Golden, Colorado

Project Goals: Past research has shown that photosynthesis, respiration, and fermentation are all required to sustain H_2 photoproduction from water in algae. These microbes utilize [Fe]-hydrogenases, which are the most efficient H_2-generating biocatalysts known. The long-term objective of our project is to identify the suite of genes facilitating and/or limiting H_2 photoproduction in the alga, *Chlamydomonas reinhardtii*, by conducting global gene expression and cell metabolism studies using algal cells acclimated to conditions known to induce H_2-production activity. A detailed understanding of the influences of metabolism and other environmental factors on

* Presenting author

the coordinated expression of genes and biochemical pathways associated with H_2-production activity will ultimately be required to increase the yields of renewable H_2 production for potential future applications. To accomplish this we will examine WT cells and a number of NRELs H_2-production mutants under a number of experimental conditions using *Chlamydomonas* gene microarrays along with extensive biochemical assays. Algal H_2 production requires the synergies of multiple redox proteins, sensors, biochemical pathways and regulatory processes. Knowledge gained by deconvoluting these interactions will help us identify specific targets for future strain engineering aimed at enhancing H_2 production in *C. reinhardtii*.

The unicellular green alga *Chlamydomonas reinhardtii* has emerged as a prototype organism for investigating processes such as photosynthesis, nutrient deprivation, flagellar function, and H_2 production. Although previous physiological studies have linked fermentation and photosynthetic electron transport to H_2 production in *C. reinhardtii*, a more precise knowledge of the metabolic and regulatory context required for H_2 production will be necessary in order to understand current limitations in H_2 yields. We have combined molecular and physiological approaches to examine the acclimation of *C. reinhardtii* strain CC-425 during a shift from oxic to anoxic conditions, which leads to H_2 evolution. The levels of transcripts involved in fermentative metabolism were monitored to determine whether the accumulation of these transcripts reflects the abundance of specific metabolites that accumulate in the cultures during anoxic adaptation. We also used high-density, oligonucleotide-based microarrays to obtain insights into the genome-wide responses initiated by anoxia, as monitored by changes in the relative abundance of ~10,000 unique transcripts. While transcripts from a number of genes associated with fermentation metabolism increased, as expected, several genes encoding proteins involved in transcriptional/translational regulation, post-translational modifications, and stress responses also increased as the cell cultures became anoxic.

Microarray and qPCR analyses were used to examine the pathways associated with fermentation at the RNA level. Several transcripts encoding proteins critical for fermentation increase as the cultures became anoxic. Indeed, anoxia leads to the upregulation of transcripts encoding pyruvate formate lyase (PFL), pyruvate:ferredoxin oxidoreductase (PFOR), alcohol dehydrogenase (ADHE), phosphotransacetylase (PTA2) and acetate kinase (ACK1), as well as some cognate proteins. These results imply that upon exposing cells to dark, anoxic conditions, *C. reinhardtii* can switch very rapidly to fermentative metabolism. The fermentation products synthesized by *C. reinhardtii* following the imposition of anoxia include malate, formate, acetate, and ethanol. Formate, acetate and ethanol, in the ratio 1:1:0.5, were the major fermentative products formed over a 24-h period of anoxia; malate was observed only at minor levels. The observed ratio of fermentation products confirms that both the PFL and PFOR pathways are activated as O_2 in the cultures declines. Moreover, during fermentation, starch, the principal carbon-storage compound in this alga, is degraded primarily to glucose-1-phosphate, which is subsequently oxidized to pyruvate during glycolysis. Accordingly, the levels of transcripts encoding amylase and **β**-amylase (involved in starch degradation) also increased.

The array data hints at regulatory processes that accompany the acclimation of the cells to dark, anoxic conditions. Many transcripts that increase correspond to chloroplast regulatory elements, including ppGpp synthetase/degradase, Mbb1 factor, translation initiation factor IF2, Tab2, and Tbc2 proteins. Elevated levels of transcript for these polypeptides may indicate the need to control translational and post-translational processes that occur in the chloroplast as the environment becomes anoxic (perhaps reflecting both structural and functional changes that occur in the chloroplast). Transcripts, encoding several proteases and kinases, also increase in cells following exposure to dark, anoxic conditions. These results suggest the possible activation of specific signalling pathways, the initiation of specific protease-dependent regulatory processes, and/or the need to redistribute amino acid resources of the cell. Notably, transcripts encoding putative O_2-sensing proteins are

upregulated, indicating a possible mechanism by which algae sense and respond to the presence or absence of O_2. Increased levels of several transcripts encoding proteins associated with anaerobic respiration are also observed. These proteins are potentially involved in pathways that compete with hydrogenase for reductant and could be the focus of future engineering efforts. A significant number of transcripts encoding proteins of unknown function are also observed to be differentially expressed.

Array data from distinct *C. reinhardtii* mutants, which are (1) unable to synthesize an active [FeFe]-hydrogenase or (2) defective in starch synthesis, have been obtained and will be analyzed. Wild-type and mutant strains will also be examined using cultures deprived of sulfate, a physiological condition that induces anaerobiosis in the light and results in sustained H_2 photoproduction. Finally, insights obtained from the cellular-metabolism and gene-expression data are being integrated into a larger systems framework that is focused on understanding the flexibility of whole-cell metabolism under rapidly changing environmental conditions. Moreover, this information may potentially be leveraged into metabolic-engineering strategies designed to optimize the production of fermentative products including H_2 and/or ethanol.

93 MEWG

Perspectives in Metabolic Flux Mapping

Jacqueline V. Shanks* (jshanks@iastate.edu)

Department of Chemical and Biological Engineering, Iowa State University, Ames, Iowa

Project Goals: Experimentally determine flux maps for plant systems (NSF) and microbes (Interagency Program on Metabolic Engineering).

Metabolic flux maps provide a quantitative depiction of carbon flow through competing metabolic pathways, thus providing: analysis of substrate utilization and product formation; flexibility or rigidity of carbon flow at network nodes; the rate of a given enzymatic reaction in vivo; and inferred availability of NADPH or ATP. Thus, metabolic fluxes are an important physiological characteristic complementary to levels of transcripts, proteins, and metabolites. The system-wide quantification of intracellular fluxes in an organism is called metabolic flux analysis (MFA). The most basic approach to MFA is stoichiometric MFA, which involves writing balances for intracellular metabolites based on the stoichiometry of the biochemical reactions in the metabolic network. This results in a system of linear equations, which are solved by employing extracellular and biomass synthesis flux measurements to resolve some or all degrees of freedom. Genome-wide or in silico flux models provide the solution space of feasible fluxes resulting from optimization of the balances to a global cellular goal, such as maximum growth rate. Recently, constraints to the in silico models provided by data from ^{13}C labeling experiments, have narrowed the solution space.

^{13}C metabolic flux analysis (^{13}C MFA), with isotope detection via GC/MS or NMR of metabolites (e.g. amino acids from hydrolyzed protein), quantifies intracellular metabolic fluxes for smaller reaction networks, where the fluxes are completely determined (in constrast to the in silico models). ^{13}C MFA provides redundant measurements for flux quantification, as well as testing the consistency of the network topology for the physiological conditions. Isotopomers, which are isomers of a metabolite that differ in the labeling state (^{13}C or ^{12}C) of their individual carbon atoms, are a central concept in the analysis and mathematical modeling of ^{13}C MFA.

* Presenting author

Increasing levels of information can be obtained from ^{13}C labeling data when coupled with a stoichiometric model of the biochemical pathways and computational methods to solve for flux data in the smaller network. More rigorous analysis is indicated as one moves from analytical (a few flux ratios at metabolic branchpoints) to ^{13}C constrained flux analysis (stoichiometric model with a few flux ratios as constraints) to fully integrated determination of fluxes from all the experimental data and the stoichiometric and isotopomer balances. Iterative methods have been used to solve the full relationship of isotopomer balances and the NMR or GC measurements to provide consistency, and routines to minimize error from the overdetermined data sets are required.

^{13}C MFA studies of aerobic glycolysis in microorganisms have become "higher throughput" since simplifications to the metabolic network can be made, and ^{13}C constrained flux analysis can be used. For alternative physiological conditions, for example where anapleurotic pathways are active, reversibility of reactions are indicated, or substrates other than glucose are used, the development of a consistent network topology and the strategy for the choice of the label to obtain identifiable fluxes are not as straight-forward. Furthermore, due to compartmentation and the existence of parallel pathways in plants, more experimental measurements are needed than in microbial systems, and the number of isotopomer balances increases, further increasing the computational burden. As a note, to date in silico models for plants have not been developed. Thus, at this point, these more challenging systems are not yet ready for "high-throughput" measurements. However, a growing knowledge base in ^{13}C MFA in these systems should enable movement towards more genome-wide flux estimation. This presentation will summarize these points with approaches from our laboratory in determining ^{13}C-based metabolic flux maps in plants and microbes.

94 GTL

High-Resolution Functional Assignments of Genes through Mapping KEGG Pathways to Bacterial Genomes

Fenglou Mao, Hongwei Wu, and **Ying Xu*** (xyn@bmb.uga.edu)

Department of Biochemistry and Molecular Biology and Institute of Bioinformatics, University of Georgia, Athens, Georgia

Project Goals: Modeling of biological pathways and networks.

We have developed a computational capability for mapping KEGG metabolic pathways to sequenced bacterial genomes. This capability assigns genes of a bacterial genome to specific enzymatic roles of a given KEGG pathway using a two-level strategy: (a) initial assignment is based on the premise that a bacterial metabolic pathway is in general encoded by a number of (in general transcriptionally co-regulated) operons and based on predicted functions of individual genes possibly at a low-resolution level; and (b) filling the gaps, the unassigned enzymes, in a partially-assigned pathway based on the detected co-evolutionary, co-occurrence and co-regulated relationships and predicted protein-protein interactions between un-annotated genes and genes already assigned to the pathway. To facilitate automated functional assignment of genes, we have developed a number of supporting computational tools, including prediction of operons [1], uber-operons [2] and regulons (unpublished results).

A. Initial KEGG pathway mapping: We have developed a computational algorithm for mapping a KEGG pathway to a specified bacterial genome. The algorithm starts by searching each gene in the target genome against gene databases with annotated functions such as the *nr* database and making

functional predictions, possibly at a low-resolution level, based on identified homology relationship. Then a number of genes (possibly zero) with annotated functions will be predicted as possible candidates for each enzyme in the KEGG pathway, based on the match between the predicted gene functions and the enzyme. We then assign at most one candidate gene to each enzyme of the KEGG pathway, using the following criteria: the overall consistency between the predicted gene functions and their assigned enzymatic roles should be as high as possible, and the selected genes should be clustered as much as possible as the predicted operons. This problem is formulated as a constrained optimization problem, specifically a linear integer programming problem, and solved using a commercial linear integer programming solver COIN. This overall prediction capability has been implemented as a computer program, called PMAP-KEGG. Using this capability, we have mapped over 140 KEGG pathways to 300+ sequenced bacterial genomes, including *E. coli*, for which detailed validation has been done using pieces of information from multiple sources. For every sequenced bacterial genome, our mapping results cover a substantial fraction of all the genes in that genome. Detailed data will be reported in an extended version of this abstract.

It should be noted that the operon prediction for each target bacterial genome is made using three prediction programs, JPOP [1], OFS [3] and VIMSS [4]. A simple majority-vote scheme is used for the final operon prediction. In the actual formulation of the problem, we have also taken into consideration the predicted uber-operon information using our own prediction program [2], where a *uber-operon* represents a group of opreons whose union is conserved across multiple genomes, which gives a higher prediction coverage than using operons alone.

B. Filling gaps in a partially assigned pathway: The mapped KEGG pathways often contain "gaps", unassigned enzymatic roles, due to various reasons. We have developed a computational procedure attempting to fill in these gaps, using three types of information: (a) co-evolutionary and co-occurrence information between assigned genes and unassigned & un-annotated genes, (b) predicted regulon information (i.e., transcriptionally co-regulated operons), and (c) protein-protein interaction information derived using various techniques such as the gene fusion method. It has been generally known that co-evolutionary, co-occurrence and co-regulation information of genes can help to predict functional relatedness among genes, even when functions of some of the genes are unknown. By employing this idea, we have recently developed a computational technique for predicting genes that are possibly working closely together in the same biological process [5,6]. Using this capability, we have predicted an initial set of candidate genes for each "gap" in a partially assigned KEGG pathway. We have then predicted protein-protein interaction relationships between the candidate genes for each "gap" with the genes already assigned to the network neighborhoods of the gap. Our final prediction for each gap is selected, using a trained neural network, based on the predicted functional relatedness and protein-protein interaction. We found that we were able to make correct gene assignments (as top assignments), for about 30% of the gaps, on a large test set using well characterized *E coli.* pathways after manually removing some of the assigned genes (1-3 genes are randomly removed from each assigned pathway). Detailed results will be reported in an extended version of this abstract.

Concluding remarks: By assigning genes of a bacterial genome to KEGG pathways, we can provide functional prediction of genes at a high-resolution level (knowing exactly the functional role in a well understood metabolic pathway), compared to the low-resolution functional annotation typically provided by a genome annotation system, e.g., gene A encodes a protease, and also can assign un-annotated genes to possible functional roles in a metabolic pathway. Our computational prediction program consists of a number of prediction and analysis tools, which are pipelined together to facilitate large-scale applications. A database containing all mapped KEGG pathways to each of the 300+ sequenced bacterial genomes is currently being developed, and will be made publicly available within a few months. This collection of mapped pathways has provided a very rich set of information

* Presenting author

for studies of bacterial metabolic pathways and their evolution. For example, by comparing the same mapped KEGG pathways across multiple genomes, we can derive information about how a pathway has evolved in adaptation to an organism's living environments, leading to general information about pathway evolution and adaptation.

Acknowledgement: This work was supported in part by the by National Science Foundation (NSF/DBI-0354771, NSF/ITR-IIS-0407204, NSF/CCF-0621700, NSF/DBI-0542119) and the US Department of Energy's Genomics: GTL Program.

References

1. X. Chen, Z. Su, P. Dam, B. Palenik, Ying Xu and T. Jiang, Operon prediction by comparative genomics: an application to the *Synechococcus WH8102* genome, *Nucleic Acids Research*, 32 (7), 2147 – 2157, 2004.
2. D. Che, G Li, F. Mao, H Wu, and Ying Xu, "Detecting uber-operons in microbial genomes", *Nucleic Acids Research*, 2006 (in press).
3. BP. Westover, JD. Buhler, JL Sonnenburg, and J.I. Gordon, Operon prediction without a training set, *Bioinformatics*, 21, 880-888, 2005.
4. MN. Price, KH. Huang, EJ. Alm and AP. Arkin, A.P. A novel method for accurate operon predictions in all sequenced prokaryotes. *Nucleic Acids Res*, 33, 880-892, 2005.
5. F. Mao, Z. Su, V. Olman, P. Dam, Z. Liu and Ying Xu, Mapping of Orthologous Genes in the Context of Biological Pathways: an Application of Integer Programming, *Proc Natl Acad Sci USA*, 103, 129-134, 2006.
6. H. Wu, Z. Su, V Olamn, Ying Xu, Prediction of functional modules through comparative genome analysis and application of gene ontology, *Nucleic Acids Research*, 33, 2822-2837, 2005.

Section 3
Regulatory Processes

95 GTL

A Systems Approach to Characterizing Evolutionarily Conserved Transcriptional Complexes Elucidates the Architecture of a Global Regulatory Network in Archaea

Marc T. Facciotti*, David J. Reiss, Min Pan, Amardeep Kaur, Madhavi Vuthoori, Richard Bonneau, Paul Shannon, Alok Srivastava, Samuel M. Donahoe, Leroy Hood, and **Nitin S. Baliga** (nbaliga@systemsbiology.org)

Institute for Systems Biology, Seattle, Washington

Project Goals: MAGGIE: To characterize evolutionarily conserved protein complexes from a systems perspective

MAGGIE Component 3

Cells responding to dramatic environmental changes or undergoing a developmental switch typically change the expression of numerous genes. In bacteria sigma factors regulate much of this process

while in eukaryotes four RNA polymerases and a multiplicity of generalized transcription factors (GTFs) are required. Here, using a systems approach, we demonstrate how archaeal organisms accomplish similar large scale transcriptional segregation and modulation of related physiological functions with an expanded family of Transcription Factor B (TFB) proteins. Further, our data suggest that a gene regulatory circuit assembled through an evolution of protein-protein and protein-promoter interactions among the seven TFBs might mediate coordination of their regulatory functions. The findings reported here represent a significant contribution towards closing the gap in our understanding of gene regulation by GTFs for all three domains of life.

96 GTL

CRP and cAMP Regulatory Networks of *Shewanella oneidensis* MR-1 Involved in Anaerobic Energy Metabolism

Daad A. Saffarini,[1]* Sheetal Shirodkar,[1] Yang Zhang,[2] and **Alexander S. Beliaev**[2]

[1]University of Wisconsin, Milwaukee, Wisconsin and [2]Pacific Northwest National Laboratory, Richland, Washington

Project Goals: This project is a component of the *Shewanella* Federation and as such contributes to the overall goal of applying the tools of genomics, leveraging the availability of genome sequence for 18 additional strains of *Shewanella*, to better understand the ecophysiology and speciation of respiratory-versatile members of this important genus. To understand these systems the SF is using genome-based approaches to investigate *Shewanella* as a system of integrated networks; first describing key cellular subsystems those involved in signal transduction, regulation, and metabolism—then building towards understanding the function of whole cells and, eventually, cells within populations. As a general approach, the SF is collectively employing complimentary top-down bioinformatics-based genome functional predictions, high-throughput expression analyses, and functional genomics approaches to uncover key genes as well as metabolic and regulatory networks. The bottom-up component employs more traditional approaches including genetics, physiology and biochemistry to test or verify predictions. This information will ultimately be linked to analyses of signal transduction and transcriptional regulatory systems and used to develop a linked model that will contribute to understanding the ecophysiology of *Shewanella* in redox stratified environments.

Shewanella oneidensis MR-1 is a facultative metal-reducing bacterium with extensive respiratory versatility. Unlike many bacteria studied to date, the ability of *S. oneidensis* to grow anaerobically with several electron acceptors is regulated by the cAMP-receptor protein (CRP) and not the fumarate-nitrate reduction (FNR) regulator. CRP-deficient mutants of MR-1 are impaired in anaerobic respiration and growth with Fe(III), Mn(IV), fumarate, nitrate, and DMSO due to the loss of anaerobic terminal reductases as well as heme and porphyrin biosynthesis deficiencies.

Genetic and biochemical evidence confirms that similarly to other bacteria, CRP in *S. oneidensis* is activated upon binding to cAMP. The genome sequence of *S. oneidensis* contains three genes that are predicted to encode adenylate cyclases. This prediction was confirmed by complementation of *E. coli* mutants that lack the adenylate cyclase gene *cyaA*. An *S. oneidensis* triple mutant that lacks the adenylate cyclase genes (*cyaA*, *cyaB*, and *cyaC*) was generated and found to be deficient in anaerobic respiration similar to the *crp* mutant. To further elucidate the role of CRP and to understand the

* Presenting author

mechanisms of cAMP-dependent gene expression under anaerobic conditions in *S. oneidensis*, a combination of experimental and computational approaches have been applied.

Expression profiling of mRNA levels suggests that CRP regulates gene expression directly and indirectly. Global transcriptome comparisons of the wild-type *vs.* the Δ*crp* mutant indicate that CRP positively regulates the expression of genes involved in anaerobic energy generation and transcriptional regulation. These include the anaerobic DMSO reductase (*dmsAB*), the decaheme *c*-type cytochrome cluster (*omcA, mtrCAB*), the anaerobic formate reductase (*fdhABC*), the pyruvate-formate lyase (*pflAB*), and genes encoding the two-component signal transduction involved in anaerobic respiration of sulfur compounds. Mobility shift assays using purified CRP and DNA upstream of the DMSO reductase operon, indicate that this protein directly regulates the expression of the *dms* genes. Regulation of other anaerobic reductase genes appears to involve additional proteins that are under CRP control. Recent experiments identified a two-component signal transduction system (SO4155 and SO4157) that is upregulated by CRP and that regulates the expression of thiosulfate reductase genes. Additionally, we identified a regulatory protein (SO0490) that appears to negatively regulate sulfur reduction. SO0490 is regulated by CRP as suggested by microarray analysis and transcriptional *lacZ* fusions. Our results suggest that a complex regulatory network, with CRP as the global regulator, controls anaerobic respiration in *S. oneidensis*. Further work is underway to further elucidate the mechanisms of anaerobic gene regulation in *S. oneidensis* MR-1.

97 GTL

Mapping the Genome-Scale Regulatory Network of *Shewanella oneidensis* MR-1: Identification of Metal-Respiratory Regulation

M.E. Driscoll, F.S. Juhn, J.J. Faith, B. Hayete, J.J. Collins, and **T.S. Gardner*** (tgardner@bu.edu)

Department of Biomedical Engineering, Boston University, Boston, Massachusetts

Project Goals: To map the transcriptional regulatory pathways underlying metal respiration and carbon source utilization in *Shewanella oneidensis* MR-1

Shewanella's respiratory versatility reflects its diverse environmental ecology. In recent years, *Shewanella* species have been isolated from fresh water lake sediments, surface waters of the Sargasso sea, hydrothermal vents of the deep Pacific, marine sediments from around the globe, mollusks and spoiling fish. To successfully compete across these distinct niches, many of which represent dynamically shifting redox environments, *Shewanella* species must be respiratory generalists. In particular, *Shewanella*'s capacity for driving respiration with metals as electron acceptors – including arsenic and uranium – has made it a candidate for use in microbial fuel cells and environmental remediation applications.

Though multiple *Shewanella* genomes have been sequenced and many of the enzymes involved in electron transport have been identified, little is known about how this metabolic machinery is regulated. To this end, we have designed the first high-density oligonucleotide array for *S. oneidensis MR-1* to observe and model its global gene expression. We have profiled gene expression in more than one hundred environmental conditions which vary carbon sources, electron acceptors, and environmental factors within physiological ranges. These conditions represent the first phase of more than 300 planned conditions.

Using this initial expression data, we have predicted a regulatory network of more than 200 transcriptional interactions for *S. oneidensis* using the CLR algorithm we recently developed. The CLR algorithm, a novel extension of the relevance networks class of algorithms, has been successfully validated in *Escherichia coli* for mapping global regulatory networks.[1] In the *E. coli* study, 741 novel regulatory interactions were identified at a 60% true positive rate.

The predicted *S. oneidensis* regulatory map suggests several novel relationships between as-yet uncharacterized transcription factors and genes governing heme synthesis and cytochromes implicated in iron and manganese reduction. Analysis of our expression profiles also suggests that *S. oneidensis* possesses a broader capacity for carbon source utilization than has been previously observed.

While electron acceptor pathways have been a dominant focus of study for *S.oneidensis* to date, bacterial respiration involves a complex interplay between both electron donor and acceptor pathways. A deeper knowledge of both electron donor and acceptor metabolism is relevant not only to understanding the role of *S.oneidensis* in its natural environments, but also towards the optimization of dissimilatory metal reducing bacteria for multiple applications.

Reference

1. JJ Faith, B Hayete, JT Thaden, I Mogno, J Wierzbowski, G Cottarel, S Kasif, JJ Collins, TS Gardner. (2007) Large-scale mapping and validation of *Escherichia coli* transcriptional regulatory interactions from a compendium of expression profiles. *PLoS Biology*. 5(1): e8.

98 GTL

A Web-Based Tool for Visualizing *Shewanella* Gene Expression Profiles in Their Chromosomal Context

J.J. Faith,[1]* R. Sachidanandam,[2] and **T.S. Gardner**[1] (tgardner@bu.edu)

[1]Department of Biomedical Engineering, Boston University, Boston, Massachusetts and [2]Cold Spring Harbor Laboratory, Cold Spring Harbor, New York

Project Goals: To map the transcriptional regulatory pathways underlying metal respiration and carbon source utilization in *Shewanella oneidensis* MR-1

Common microarray analysis procedures yield lists of genes, whose expression changes significantly in response to an environmental or genetic perturbation. The functional role for most of these expression changes is typically unknown, and the often large number of changed genes hinders human interpretation of their role. In bacteria, genes with similar functional roles often exhibit chromosomal proximity and therefore operate as a coexpressed module, even when part of distinct operons. Moreover, the expression of an RNA in intergenic regions can also suggest a probable role as noncoding regulatory RNA. To facilitate the sharing, discovery and analysis of *Shewanella oneidensis* expression data and gene function, we developed a web-based genome browser where users can dynamically choose any two sets of experiments from the current *Shewanella oneidensis* Affymetrix microarray compendium and view expression levels or changes of genes in their chromosomal context. This capability is built into the M3D Database of Affymetrix microarray compendia. M3D includes compendia of several hundred expression profiles for multiple microbes including *S. oneidensis, E. coli,* and *S. cerevisiae*, and provides viewing and raw data download tools.

* Presenting author

Reference

1. Faith JJ, Fusaro V, Driscoll M, Juhn F, Cosgrove E, Hayete B, Gardner TS. Many Microbes Microarray Database (M3D). http://m3d.bu.edu

99 GTL

Comparative Genomics of Signal Transduction in *Shewanella*

Luke E. Ulrich and **Igor B. Zhulin*** (joulineib@ornl.gov)

Joint Institute for Computational Sciences, University of Tennessee, Knoxville, Tennessee and Oak Ridge National Laboratory, Oak Ridge, Tennessee

Project Goals: This project is a component of the *Shewanella* Federation and as such contributes to the overall goal of applying the tools of genomics, leveraging the availability of genome sequence for 18 additional strains of *Shewanella*, to better understand the ecophysiology and speciation of respiratory-versatile members of this important genus. To understand these systems the SF is using genome-based approaches to investigate *Shewanella* as a system of integrated networks; first describing key cellular subsystems—those involved in signal transduction, regulation, and metabolism—then building towards understanding the function of whole cells and, eventually, cells within populations. As a general approach, the SF is collectively employing complimentary "top-down" -bioinformatics-based genome functional predictions, high-throughput expression analyses, and functional genomics approaches to uncover key genes as well as metabolic and regulatory networks. The "bottom-up" component employs more traditional approaches including genetics, physiology and biochemistry to test or verify predictions. This information will ultimately be linked to analyses of signal transduction and transcriptional regulatory systems and used to develop a linked model that will contribute to understanding the ecophysiology of *Shewanella* in redox stratified environments

The availability of genome sequence of 13 *Shewanella* strains provided us with a unique opportunity to unravel the evolutionary trends of signal transduction systems on a single genus scale. We used recently developed MiST database to analyze signal transduction profiles in *Shewanella* [1, 2]. Results obtained allowed us to link the overall signal transduction profile of a given strain to its metabolic potential. We also found that the distinct natural history of signal transduction proteins in *Shewanella* species provides useful markers for improving their taxonomy. Overall, all *Shewanella* strains maintain a similar profile of signal transduction with respect to protein families that constitute regulatory networks; however, the number of proteins in each family varies significant from strain to strain. The most dramatic changes in the overall composition of signal transduction is observed in *S. denitrificans* OS217. This strain has only 148 transcriptional regulators, whereas the strain *S. putrefaciens* CN-32, which has a comparable genome size, has 181 transcriptional regulators. Adjusted to the genome size, *S. denitrificans* OS217 is the outlier in each major category of signal transduction. Such reduction in this important functional category correlates with the lack of most of the anaerobic respiratory machinery that is present in other strains. This finding strongly supports the hypothesis that many regulatory pathways in *Shewanella* control its branched electron transport system.

Recent divergence of the 13 *Shewanella* strains allowed us to identify distinct cases of horizontal gene transfer and strain-specific gene loss that are most common trends in the evolution of regulatory systems. These findings lead to a better resolution of *Shewanella* taxonomy

References

1. Ulrich, L.E., E.V. Koonin, and I.B. Zhulin (2005) One-component systems dominate signal transduction in prokaryotes. *Trends Micrbiol* 13: 52-56.
2. Ulrich, L.E., and I.B. Zhulin (2006) MiST: a microbial signal transduction database. *Nucleic Acids Res.* Nov 28, [Epub ahead of print].

100 GTL

Comparative Genomics of Transcriptional Regulation of Metabolic Pathways in *Shewanella* Species

Dmitry Rodionov[1]* (rodionov@burnham.org), Mikhail Gelfand,[2] Margaret Romine,[3] and **Andrei Osterman**[1,4]

[1]Burnham Institute for Medical Research, La Jolla, California; [2]Institute for Information Transmission Problems RAS, Moscow, Russia; [3]Pacific Northwest National Laboratory, Richland, Washington; and [4]Fellowship for Interpretation of Genomes, Burr Ridge, Illinois

Project Goals: This project is a component of the *Shewanella* Federation and as such contributes to the overall goal of applying the tools of genomics, leveraging the availability of genome sequence for 18 additional strains of *Shewanella*, to better understand the ecophysiology and speciation of respiratory-versatile members of this important genus. To understand these systems the SF is using genome-based approaches to investigate *Shewanella* as a system of integrated networks; first describing key cellular subsystems—those involved in signal transduction, regulation, and metabolism—then building towards understanding the function of whole cells and, eventually, cells within populations. As a general approach, the SF is collectively employing complimentary "top-down" - bioinformatics-based genome functional predictions, high-throughput expression analyses, and functional genomics approaches to uncover key genes as well as metabolic and regulatory networks. The "bottom-up" component employs more traditional approaches including genetics, physiology and biochemistry to test or verify predictions. This information will ultimately be linked to analyses of signal transduction and transcriptional regulatory systems and used to develop a linked model that will contribute to understanding the ecophysiology of *Shewanella* in redox stratified environments.

Genomics-based reconstruction and comparative analysis of regulons have been utilized to predict transcriptional regulatory subnetworks in a number of model organisms (e.g. *E. coli*) providing an extremely useful resource for interpreting results from microarray datasets and deriving genome-wide regulatory models. Our approach combines the detection of transcription factors binding sites and cross-genome comparison with the analysis of the genomic and functional context inferred by metabolic reconstruction. The recent availability of genome sequences for 11 *Shewanella* species allowed us to perform, for the first time, a detailed comparative analysis of transcriptional regulons for the key pathways involved in central metabolism, production of energy and biomass, metal ion homeostasis and stress response. This analysis provides insights not only into regulatory networks in *S. oneidensis* MR-1 (our model organism) but also in the *Shewanella* lineage as a whole. In addition, this approach allows us to improve gene function and metabolic pathway assignments, as well as to accurately predict functions of previously uncharacterized gene families. Results of these studies will serve as an important shared resource of the ***Shewanella* Federation** research team who together seek to develop a systems-level understanding of *Shewanella*e metabolic and regulatory networks.

* Presenting author

Using this approach, we identified candidate binding sites for more then 20 transcriptional factors of known specificity, including global regulators (CRP, FNR, ArcA, Fur) and specialized regulators of the metabolism of nitrogen (NarP, IscR, NsrR, DNR, NorR), vitamins and amino acids (BirA, ArgR, MetJ, TyrR, HutC), fatty acids (FadR, FabR), and sugars (NagR, SdaR, ScrR, GntR), as well as of the central carbon metabolism (PdhR, HexR), Fe-S cluster assembly (IscR) and ribonucleotide reduction (NrdR). In addition, we identified candidate binding sites for previously uncharacterized regulators, termed NagR (SO3516), IlvR (SO1898) and FadQ (SO2493) tentatively implicated in the control of the metabolism of N-acetylglucosamine, branch chain amino acids, and fatty acids, respectively.

Cross-validation of these predictions with the results of microarray analysis combined with targeted gene knock-outs are currently under way. The first results reveal substantial consistency between our predictions and gene expression profiles, as will be illustrated for the arginine and N-acetylglucosamine regulons. We have also mapped the genes and operons controlled by five types of metabolite-binding riboswitches (*B12, LYS, RFN, THI, GLY*), and six translational attenuators of amino acid biosynthesis pathways (*ilv, leu, his, thr, trp, phe* operons). Although some diversity of the predicted regulons is observed within the collection of *Shewanella* spp., the most striking difference in the overall regulatory strategy is revealed by comparison with *E. coli* and other gamma-proteobacteria. Multiple examples of divergence and adaptive evolution of regulatory networks were detected and include regulon "shrinking" (as in case of FadR), "expansion" (as in case of PdhR and HexR), "mergers", and "split-ups", as well as multiple cases of using nonorthologous regulators to control equivalent pathways or orthologous regulators to control distinct pathways. Among the most notable are the differences in the regulon content and a respective role of global regulators, such as CRP. These and other observations, indicate that many aspects of metabolic regulation in *Shewanella* species., are substantially different from regulatory network models that were largely derived from studies in *E. coli*.

101 MEWG

Biological Aspects of Deciphering and Engineering Regulatory Networks

George N. Bennett* (gbennett@bioc.rice.edu)

Department of Biochemistry and Cell Biology, Rice University, Houston, Texas

Project Goals: Studies to enhance chemical production in microbes.

In efforts to detect and modify regulation of pathways for a particular goal there are strategies that can be implemented in two circumstances, one the synthetic approach where major components of the pathway are known and the appropriate regulation of the various enzymes can be adjusted by taking advantage of modeling and combinatorial assembly methods; and the other if the proteins to be altered are not obvious and thus the modification must take a more empirical approach with selection or screening methods being the issues.

In known expression systems, it is still needed to detect levels of regulation, for example determination of the level of functioning of a protein (enzyme for metabolic process) more than knowledge of the level of gene expression is needed to understand how the activity varies under the physiological condition contemplated for use. This is the situation if genes from various sources are placed together in a new way to form a non-endogenous pathway as often is the case in metabolic engineering. Detection of proteins, metabolites etc from systems biology approach of measurements in different

genetically engineered cells under various conditions can help in this endeavor. Due to the large number of possible combinations of mutations and conditions a way is needed to minimize or focus the experimental measurements on the most appropriate ones to examine. Such experiments should make an effort to take into account the effects of intercellular conditions produced by introduced changes on related protein activities (other enzymes of pathway, regulatory factors, functional state of activity of the enzyme or regulatory proteins) and models that include this interaction information would be more comprehensive.

In order to carry out appropriate modification of regulation there are relatively straightforward approaches in the case of desired changes in specific known expression controls such as through modification of transcriptional events and to a lesser extent, modification of more general aspects of cellular physiology (redox and air, enzyme stability, enzyme parameters, osmotic conditions). To target regulation to specific pathway, designed regulation of small units can be employed. These involve the use of known regulated promoters that are varied to adjust the level of constitutive expression, and can be combined with terminators or RNA structural elements to afford variation of level of expression. In the case of less known processes found in many industrial organisms that have less genetic and biochemical literature, efforts are more a matter of perturbing a somewhat more global functioning system and screening or selection for those altered cells that perform better, then analyzing and combining the most promising. The idea of eliminating complicating or undesired processes that may obscure the regulation you would like to enhance is an useful experimental strategy. For these wider scope effects, or regulation of unknown factors with less obvious connections to the metabolically engineered process, modification of transcription factors such as sigma factors, Zn-finger motif factors, general or global transcription factors may be used in combination with powerful selection or screening systems for the desired property.

102 GTL

Characterization of Behavioral Responses in *Shewanella oneidensis*

Jun Li,[1] Margie Romine,[2] and **Mandy Ward**[1]* (mjward@jhu.edu)

[1]Department of Geography and Environmental Engineering, Johns Hopkins University, Baltimore, Maryland and [2]Pacific Northwest National Laboratory, Richland, Washington

Project Goals: This project is a component of the *Shewanella* Federation and as such contributes to the overall goal of applying the tools of genomics, leveraging the availability of genome sequence for 18 additional strains of *Shewanella*, to better understand the ecophysiology and speciation of respiratory-versatile members of this important genus. To understand these systems the SF is using genome-based approaches to investigate *Shewanella* as a system of integrated networks; first describing key cellular subsystems—those involved in signal transduction, regulation, and metabolism—then building towards understanding the function of whole cells and, eventually, cells within populations. As a general approach, the SF is collectively employing complimentary "top-down" - bioinformatics-based genome functional predictions, high-throughput expression analyses, and functional genomics approaches to uncover key genes as well as metabolic and regulatory networks. The "bottom-up" component employs more traditional approaches including genetics, physiology and biochemistry to test or verify predictions. This information will ultimately be linked to analyses of signal transduction and transcriptional regulatory systems and used to develop a linked model that will contribute to understanding the ecophysiology of *Shewanella* in redox stratified environments.

* Presenting author

Shewanella species are often found in redox stratified environments. This, and the ability of the genus to utilize many different terminal electron acceptors for anaerobic respiration, suggests that sensing, and responding behaviorally to, different redox conditions may be integral to the lifestyle of these microorganisms. The first studies on chemotaxis by *Shewanella oneidensis* MR-1 were conducted over ten years ago. These studies revealed that, unlike *Escherichia coli*, MR-1 does not respond to a number of carbon sources including methanol, ethanol, acetate, lactate, propionate, pyruvate, citrate, several amino acids (both individually and in combinations with vitamins), malate, glucose, yeast extract, and peptone. Formate did produce a weak response, although only under anaerobic conditions. Instead, chemotaxis to several respiratory electron acceptors (nitrate, nitrite, fumarate, TMAO, DMSO, and to a lesser extent thiosulfate) was observed. The background presence of certain energetically favorable and energetically less favorable electron acceptors (O_2, nitrate, nitrite, DMSO, and elemental sulfur) inhibited these responses, although they did not necessarily inhibit reduction of the test electron acceptor. The genome sequence of MR-1 was released seven years later, providing us with a valuable new resource for revisiting this earlier data and designing new experiments targeted at developing a more complete understanding of tactic responses in MR-1. The sequence of MR-1, as well as those soon to become available for 18 other *Shewanella* genomes, allows us to compare domain architecture of the signal tranduction proteins in this genus (presented by Dr. Jhoulin) and to conduct genome neighborhood analysis (presented by Dr. Osterman) to identify candidate metabolisms that may illicit chemotactic responses. Furthermore, high throughput analyses (global proteomics and microarrays) available from other collaborators within the *Shewanella* Federation will provide additional clues for designing experiments to investigate behavioral responses. In this presentation, we will provide preliminary findings derived from characterization of various mutants defective in tactic or respiratory functions as well as preliminary insights gleaned from surveying available microarray and proteome data from MR-1.

The genome sequence revealed three clusters of chemotaxis genes, each of which could potentially encode a complete chemotactic signal transduction pathway. Each locus encodes a CheY, CheA, CheW, CheR, and CheB protein. However, two genes (*cheA-2* and an *mcp*) in the predicted Cluster II operon are interrupted by a transposase, indicating that this entire signal transduction pathway may be degenerate. The cluster III locus encodes the sole CheZ protein and the entire subsystem required for biosynthesis and assembly of the flagellum, as well as the sigma factors σ^{54} and σ^{27}, and an anti-σ^{28} factor. Examination of our proteome data revealed that the proteins associated with cluster III are by far the most frequently observed across all our datasets. Consequently, the cluster III-encoded proteins probably constitute the main chemotactic signal transduction system in MR-1.

In-frame deletions of the *cheA-1* and *cheA-3* genes have been constructed and the phenotypes analyzed in swarm plate assays. These experiments showed that the Δ*cheA-3* mutant is non-chemotactic to a range of anaerobic electron acceptors, supporting the previous prediction that the cluster III genes encode the main chemotactic signal transduction pathway in *S. oneidensis* MR-1. Motion analysis of this mutant showed that it has the smooth swimming phenotype associated with non-chemotactic mutants. The Δ*cheA-1* mutant showed a reduced-swarming phenotype, suggesting that the cluster I-encoded signal transduction pathway is not essential for chemotaxis to anaerobic electron acceptors, but may play a role in optimizing responses.

A total of 29 different methyl-accepting chemotaxis proteins (MCPs) are encoded by the genome (2 by pseudogenes) suggesting that the variety of sensory inputs detected by *S. oneidensis* MR-1 is far greater than has presently been discovered. Data from DNA microarray and proteomic analyses of MR-1 have provided insights into the range of environmental information that these MCPs may sense. For example, induction of specific *mcp* genes in response to chromium and uranium (1), acidic pH (2), and iron and manganese (3) suggest that behavioral responses to these stimuli warrant inves-

tigation. Mutants with insertions in several *mcp* genes have been constructed and tested for loss of behavioral responses to anaerobic electron acceptors and metals. To date, the only mutants that show swarming defects to electron acceptors have insertions in the *mcp* genes that encode receptors with redox sensing PAS domains (SO0584, SO1385, SO2123, and SO3404). This result suggests that several MCPs are involved in energy taxis rather than chemotaxis. Two *c*-type cytochrome mutants display the same swarming defects as the aforementioned CheA-3 and PAS-MCP mutants, suggesting that energy taxis in *S. oneidensis* may involve respiratory electron transport sensing. Additional mutants are being constructed based on predictions from the DNA microarray and proteomics data, and new assays will be designed to test for loss of responses to specific stimuli.

The integration of predictions, based on large scale genomic / proteomic analyses, with single gene mutational approaches, has so far proven successful in determining roles for some of the Che and MCP proteins in *S. oneidensis* MR-1. Consequently, this approach will be continued and combined with comparative analyses that will make use of information available for other *Shewanella* species. The ability to design experiments based on genomics will facilitate the rapid generation of results that will be shared with other researchers in the *Shewanella* Federation by incorporation into the Integrated Knowledge Source (ORNL).

References

1. Bencheikh-Latmani *et al*. (2005) Global transcriptional profiling of *Shewanella oneidensis* MR-1 during Cr(VI) and U(VI) reduction. Appl. Environ. Microbiol., **71**:7453-7460.
2. Leaphart *et al*. (2006). Transcriptome profiling of *Shewanella oneidensis* gene expression following exposure to acidic and alkaline pH. J. Bacteriol., **188**:1633-1642.
3. Beliaev *et al*. (2006) Global transcriptome analysis of *Shewanella oneidensis* MR-1 exposed to different terminal electron acceptors. J. Bacteriol., **187**:7138-7145.

103 GTL

Development of *in vitro* Transcription System using Recombinant *Shewanella oneidensis* RNA Polymerase

Younggyu Kim* (ykim@chem.ucla.edu), Sam On Ho, Natalie Gassman, and **Shimon Weiss** (sweiss@chem.ucla.edu)

Department of Chemistry and Biochemistry, University of California, Los Angeles, California

Project Goals: Overall objective: Characterization of gene regulation in *Shewanella oneidensis* MR-1 using single molecule fluorescence spectroscopy.
1. In vitro reconstitution of MR-1 transcription machinery
2. Structural and functional characterization of the transcription system using single molecule fluorescence spectroscopy.

Shewanella oneidensis MR-1 grows aerobically and anaerobically using a variety of electron acceptors including fumarate, nitrate, nitrite, thiosulfate, elemental sulfur, trimethylamine N-oxide (TMAO), dimethyl sulfoxide (DMSO) and metal pollutants[1]. It has gained significant research attention as a potential bioremediation tool due to its ability to process biohazardous pollutants. The complete genome was sequenced by the Institute for Genomic Research (TIGR; http://www.tigr.org/) under the support of the U.S. Department of Energy (Microbial Genome Program and NABIR Program)[2], enabling us to study the biology of the bacterium at a system-wide level.

* Presenting author

One of our main goals is the elucidation of gene regulation in MR-1. It is a complex process accompanied with a network of biomolecular interactions that affect the level of gene expression. Transcription is the first step in gene expression and is the step at which most of the gene regulation takes place. RNA polymerase (RNAP) is responsible for the transcription and is, directly or indirectly, a target for the most of gene regulation. Therefore, understanding the basic transcription mechanism is an important step for the study. The core enzyme of bacterial RNA polymerase consists of five subunits: **α** dimer (**α**2), **β**, **β'**, and **ω**. When the core enzyme binds to one of its initiation factors (**σ**s), a holoenzyme is formed and it is then capable of initiating a promoter-specific transcription. Although the mechanism has been very well studied in *E. coli*, it is poorly understood in the *Shewanella* species.

In order to characterize the basic transcription machinery of the MR-1, we designed an *E. coli* expression system, co-overexpression and *in vivo* assembly of MR-1 RNA polymerase subunits in *E. coli*. We created an *E. coli* expression construct that produces polycistronic mRNA containing all MR-1 RNAP subunit coding sequences (**α**, **β**, **β'**, and **ω**) from a single T7 promoter; the C-terminus of α subunit is fused to a 6 histidine tag. The RNAP was purified by IMAC (Immobilized Metal Affinity Chromatography) and further purified by anion exchange chromatography (MonoQ). The purified multisubunit RNAP core (~400 kDa) shows a correct stoichiometric subunit ratio. Although there is a significant similarity (80-90%) in the subunit coding sequences between MR-1 and *E. coli*, we found that the subunits are not interchangeable between the two species, indicating the presence of lineage-specific coding sequences.

The recombinant RNAP is functional: (i) the core RNAP drives promoter-independent transcription, a characteristic of most core RNAPs. (ii) it forms a holoenzyme with the initiation factor, $\boldsymbol{\sigma}^{70}$, responsible for house-keeping gene transcription; (iii) electrophoretic mobility shift assay shows that the holo RNAP binds to a target promoter specifically and forms an open complex; and (iv) the open complex initiates promoter-specific transcription. It is also capable of interacting with transcriptional activators. For example, MR-1 CAP (catabolite activator protein), one of transcription activators, interacts with the RNAP on the DMSO reductase promoter[3] containing potential CAP binding site and increases transcription activity. The initial characterization was conducted on the basis of well-established *E. coli* system because the amino acid sequences of the RNAP are highly conserved throughout evolution. Nevertheless, we found another remarkable difference between two RNAPs. While *E. coli* RNAP has the highest transcriptional activity at 37°C, MR-1 RNAP exhibits the optimal activity below 30°C. It suggests that the difference in lineage-specific sequences accounts for the trait which matches with the MR-1's natural habitat.

This system can be utilized for biochemical and biophysical studies in gene regulation. Those include studies in: core RNAP with different initiation factors, RNAP-transcriptional regulator interaction, structural studies of the transcription machinery using a single molecule fluorescence spectroscopy[4], and mutational studies of the enzymes having alterations in lineage-specific sequences or in transcription activation (or repression)-responsive sequences. Therefore, the newly established in vitro transcription system will serve as an important tool in order to study the gene regulation in MR-1.

References

1. Venkateswaran, K., et al., *Polyphasic taxonomy of the genus Shewanella and description of Shewanella oneidensis sp. nov.* International journal of systematic bacteriology, 1999. **49 Pt 2**: p. 705-24.
2. Heidelberg, J.F., et al., *Genome sequence of the dissimilatory metal ion-reducing bacterium Shewanella oneidensis.* Nature biotechnology, 2002. **20**(11): p. 1118-23.
3. Saffarini, D.A., R. Schultz, and A. Beliaev, *Involvement of cyclic AMP (cAMP) and cAMP receptor protein in anaerobic respiration of Shewanella oneidensis.* Journal of bacteriology, 2003. **185**(12): p. 3668-71.
4. Kapanidis, A.N., E. Margeat, S.O. Ho, E. Kortkhonjia, S. Weiss, and R.H. Ebright, *Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism.* Science, 2006. **314**(5802): p. 1144-7.

* Presenting author

104 GTL

Genetic Analysis of Anaerobic Respiration in *Shewanella oneidensis* MR-1

Jizhong Zhou,[1] Haichun Gao,[1,3]* Xiaohu Wang,[5] Soumitra Barua,[1] Yunfeng Yang,[2] Samantha B. Reed,[4] Dave Culley,[4] Zamin Yang,[2] Christopher Hemme,[1] Zhili He,[1] Margaret Romine,[4] Kenneth Nealson,[6] James M. Tiejde,[3] Timothy Palzkill,[5] and James K. Fredrickson[4]

[1]University of Oklahoma, Norman, Oklahoma; [2]Oak Ridge National Laboratory, Oak Ridge, Tennessee; [3]Michigan State University, East Lansing, Michigan; [4]Pacific Northwest National Laboratory, Richland, Washington; [5]Baylor College of Medicine, Houston, Texas; and [6]University of Southern California, Los Angeles, California

Project Goals: This project is a component of the *Shewanella* Federation and as such contributes to the overall goal of applying the tools of genomics, leveraging the availability of genome sequence for 18 additional strains of *Shewanella*, to better understand the ecophysiology and speciation of respiratory-versatile members of this important genus. To understand these systems the SF is using genome-based approaches to investigate *Shewanella* as a system of integrated networks; first describing key cellular subsystems—those involved in signal transduction, regulation, and metabolism—then building towards understanding the function of whole cells and, eventually, cells within populations. As a general approach, the SF is collectively employing complimentary "top-down" -bioinformatics-based genome functional predictions, high-throughput expression analyses, and functional genomics approaches to uncover key genes as well as metabolic and regulatory networks. The "bottom-up" component employs more traditional approaches including genetics, physiology and biochemistry to test or verify predictions. This information will ultimately be linked to analyses of signal transduction and transcriptional regulatory systems and used to develop a linked model that will contribute to understanding the ecophysiology of *Shewanella* in redox stratified environments.

Shewanella oneidensis MR-1, a facultative γ-*proteobacterium*, possesses remarkably diverse respiratory capacities. In addition to aerobic respiration, *S. oneidensis* can anaerobically respire various organic and inorganic substrates, including fumarate, nitrate, nitrite, thiosulfate, elemental sulfur, trimethylamine N-oxide (TMAO), dimethyl sulfoxide (DMSO), Fe(III), Mn(III) and (IV), Cr(VI), and U(VI). However, the molecular mechanisms underlying the anaerobic respiratory versatility of MR-1 remain poorly understood. As a part of the *Shewanella* Federation efforts, we have used integrated genomic, proteomic and computational technologies to study the regulation of energy metabolism of this bacterium from a systems-level perspective.

ArcA. The Arc two-component system is a major control system for the regulation of many genes involved in aerobic/anaerobic respiration and fermentative metabolism in *Escherichia coli*. MR-1 genome contains a gene encoding a putative *ArcA* homolog with ~81% amino acid sequence identity to the *E. coli ArcA* protein but no full-length *arcB* gene homolog. Results from physiological, microarray and computational analyses of an *arcA* deletion mutant revealed that the regulon of *S. oneidensis ArcA* differs significantly from that of *E. coli*. For instance, *ArcA* does not appear to be involved in regulation of the TCA cycle in *S. oneidensis*. Among the 50 operons controlled by *S. oneidensis ArcA* regulon, only 6 operons were shared by *E. coli ArcA*.

To further probe the role of *ArcA* in *Shewanella*, electrophoretic motility shift assay (EMSA), DNase I footprinting, and *LacZ* reporter gene assays were conducted. As predicted, *Shewanella* and *E. coli ArcA* bind a similar 15bp DNA motif and phosphorylation of *Asp*54 residue is still essential for

* Presenting author

Shewanella ArcA activation. Surprisingly, however, our experimental results suggest *Shewanella ArcA* is constitutively activated by phosphorylation under both aerobic and anaerobic growth conditions and that activated *ArcA* regulates genes involved in the anaplerotic *sfcA* shunt, H_2 metabolism and terminal DMSO (dimethylsulfoxide) reduction. These findings indicate that *ArcA* is not a major redox response regulator in *Shewanella*. The protein-promoter binding interactions were also examined using the promoter binding microarray (PBM) for the *ArcA* protein. Using this array, the contribution of each nucleotide to the binding of *ArcA* within the *ArcA* binding site was evaluated by point mutations. Results indicate that several nucleotides are essential for *ArcA* binding and that the flanking sequence in addition to the 15bp motif appears to play a role in *ArcA* binding to the promoter.

Nitrate respiration. *S. oneidensis* MR-1 is able to reduce nitrate to ammonium under anaerobic conditions. *napDAGHB* gene cluster encoding periplasmic nitrate reductase (*NapA* of the *Nap* system) and accessory proteins and an *nrfA* gene encoding periplasmic nitrite reductase (*NrfA* of the *Nrf* system) were identified from the *S. oneidensis* genome. The Nap system catalyzes respiratory reduction of nitrate to nitrite and the *Nrf* system converts nitrite to ammonium. However, the genome lacks both *napC* and *nrfH*, which are essential for reduction of nitrate to nitrite and nitrite to ammonium, respectively. Mutation in *napA* renders the cells incapable of reducing nitrate to nitrite under either aerobic or anaerobic conditions. Similarly, mutation in *nrfA* eliminates reduction of nitrite to ammonium. Furthermore, a strain carrying the deleted *napB* gene exhibits significant differences in nitrate reduction compared to both MR-1 and the *napA* mutant. A further analysis reveals that the mutation causes reduction of nitrate to ammonium without nitrite accumulation. Mutational analysis of a *naB-nrfA* double mutant indicates that the cells missing *napB* also employ *Nap* and *Nrf* systems for nitrate and nitrite reduction. In an attempt to identify candidate replacements for *NapC* and *NrfH* functions, both microarray and mutational analyses have been performed. The results suggest that *CymA* is likely to be the functional replacement of both *NapC* and *NrfH* and a conceptual model is proposed. It appears that *NapB* is the preferred electron transferring protein and has an absolute priority in accepting electron from C*ymA*. In the presence of *NapB, NrfA* could not function efficiently due to the lack of electrons from *CymA*.

Functional analysis of *c*-type cytochromes. To investigate the role of *c*-type cytochrome genes in anaerobic respiration, targeted deletions of 37 out of 44 predicted intact *c*-type cytochrome encoding genes have been generated by either homologous cross-over using host-encoded recombinases or by introduced phage *cre-loxP* recombinases. Each mutant was tagged with unique bar codes (i.e. short synthetic oligonucleotides) to facilitate tracking of the individual strains in planned competitive growth studies to determine the fitness of each mutant under different electron acceptor conditions. Growth studies revealed significant effects of these mutations with different electron acceptors compared to wild type MR-1, suggesting a complex network of electron transfer reactions. In agreement with previous observations, a key *c*-type cytochrome, *CymA*, showed decreased growth dynamics in five different terminal reductases except TMAO. In addition the *mtrDEF* gene cluster, which is similar to the metal reduction gene cluster (*mtrC/omcB*, *mtrAB*) were also partially defective in growth with Mn(IV) and Cr(IV). Mutants in the high affinity *cbb3* cytochrome oxidase components exhibit a defect both aerobically and anaerobically with TMAO and Cr(VI), suggesting a role for this complex in both suboxic and anaerobic respiratory processes. In addition, an insertional mutant was obtained for a periplasmic tetraheme flavacytochrome gene, *SO3056*. Genetic analyses indicate that this mutant is defective in Fe(III) reduction and growth with thiosulfate, nitrate, fumarate, DMSO but not TMAO and oxygen. The phenotypes of *SO3056* mutant are largely similar to what have been known for *CymA* mutant. Thus, *SO3056* might function downstream of cytoplasmic membrane protein *CymA* and likewise control multiple branches of electron transfer chain. It is clear from this overview of the current data available that many of these *c*-type cytochromes participate in respiratory metabolism, detoxification or sensing processes that have not yet been explored in

detail. Therefore, the availability of these *c*-type cytochrome mutants and of additional sequenced *Shewanella* strains provides an excellent resource for comparative physiology studies and will greatly facilitate our goal of characterizing respiratory networks in *Shewanella*.

105 GTL

A Phylogenetic Gibbs Sampler for High-Resolution Comparative Genomics Studies of Transcription Regulation

William A. Thompson,[1] Sean P. Conlan,[2] Thomas M. Smith,[2,3] Lee A. Newberg,[2,3] Lee Ann McCue[2,4]* (leeann.mccue@pnl.gov), and **Charles E. Lawrence**[1]

[1]Center for Computational Molecular Biology and the Division of Applied Mathematics, Brown University, Providence, Rhode Island; [2]The Wadsworth Center, New York State Department of Health, Albany, New York; [3]Department of Computer Science, Rensselaer Polytechnic Institute, Troy, New York; and [4]Pacific Northwest National Laboratory, Richland, Washington

Project Goals: The two major components of a prokaryotic cell's transcription regulation network are the transcription factors (TFs) and the transcription factor binding sites (TFBS); these components are connected by the binding of TFs to their cognate TFBS under appropriate environmental conditions. Comparative genomics has proven to be a powerful bioinformatics method with which to study transcription regulation on a genome-wide level. We will further extend comparative genomics technologies that we have introduced over the last several years, developing and applying statistical approaches to analysis of correlated sequence data (i.e. sequences from closely related species). We also plan to combine functional genomic and proteomic data with sequence data from multiple species; combining these complementary data types promises to improve our ability to predict regulatory sites of small or genus-specific regulons.

High-throughput sequencing initiatives are enabling ambitious comparative genomics projects, such as the detection of functionally conserved regions in protein and DNA sequences. Of particular interest is the identification of transcription factor binding sites (TFBS) and *cis*-regulatory modules in the promoters of genes - a critical step for delineating the transcription regulatory network of an organism. While closely related species are most likely to share common transcription factors (and therefore common *cis*-regulatory elements), the recent speciation of closely related genomes results in correlation among the sequences which confounds the detection of functionally conserved motifs.

To facilitate high-resolution comparative genomics studies, we have developed a version of the Gibbs recursive sampler that incorporates phylogeny of the input sequences through the use of an evolutionary model, and calculates an ensemble centroid motif solution. This phylogenetic Gibbs sampler accepts aligned as well as unaligned orthologous sequence data; these may be orthologous sequences from a single gene or orthologous sequences for a group of co-regulated genes. The algorithm also requires a user-supplied tree describing the phylogenetic relationship of the orthologous sequences within each multiple-sequence alignment. For each alignment, the algorithm traverses this phylogenetic tree and calculates the joint probability of each nucleotide at each position, ultimately describing a motif as a product phylogeny model. We also employ an "ensemble centroid" motif estimator, *i.e.*, the solution that is the set of sites that has the minimum total distance to the set of sites sampled from the posterior weighted ensemble of sites. The results below indicate that these two extensions provide significant additional power to the existing capabilities of the Gibbs recursive sampler.

* Presenting author

We demonstrate the advanced features of the phylogenetic Gibbs sampler on the challenging problem of predicting motifs in orthologous data from a single gene. Using data that simulate this scenario, we show that false positive predictions, caused by correlation among the sequences, are dramatically reduced by the features described here. Specifically, we show that using a phylogenetic model and ensemble centroid solutions yields improved positive predictive values (PPV), an improved ability to avoid false positives, over a non-phylogenetic version of the Gibbs recursive sampler and a phylogenetic version of the Gibbs sampler that predicts maximum-likelihood alignments. We further demonstrate the ability of the phylogenetic Gibbs sampler to detect transcription factor binding sites in real orthologous sequence data from eight proteobacterial genomes, the majority of which are closely related enterobacteria.

106 MEWG

Challenges in Predictive Modeling for Engineering/Deciphering the Regulatory Networks

James C. Liao*

Chemical and Biomolecular Engineering, University of California, Los Angeles, California

Project Goals: Network analysis and prediction

The ultimate goal of systems biology and metabolic engineering is the prediction of cellular behavior, either for wild-type strains or the engineered strains. This ambitious task involves two levels: 1) prediction of regulatory networks, and 2) prediction of the behavior of the regulatory networks. We will discuss challenges in these problems and suggest some possible ways to tackle them. In particular, we will address how mathematical tools can interact with experimental data to maximize information output.

107 GTL

Molecular Mechanisms Regulating Gene Expression in *Geobacter sulfurreducens* under Environmentally Relevant Conditions

Toshiyuki Ueki[1]* (tueki@microbio.umass.edu), Ching Leang,[1] Byoung-Chan Kim,[1] Richard Glaven,[1] Haiping Ke,[1] Katy Juárez,[2] and **Derek R. Lovley**[1] (dlovley@microbio.umass.edu)

[1]Department of Microbiology, University of Massachusetts, Amherst, Massachusetts and [2]Universidad Nacional Autónoma de México, Mexico

Project Goals: Our project goal is to predictively model how *Geobacter* species respond to natural environmental conditions or conditions that may be artificially imposed to promote in situ uranium bioremediation and electricity harvesting from waste organic matter and renewable biomass. Especially in this subproject, the molecular mechanisms for regulation of gene expression in *Geobacter sulfurreducens* are being investigated in order to better understand physiological functions of *Geobacter* species during in situ bioremediation and on the surface of electricity harvesting electrodes. These studies continually increase our knowledge of regulatory mechanisms

at the molecular level which can provide important insights into the physiology and ecology of *Geobacter* species in these environments.

The mechanisms for regulation of gene expression in *Geobacter sulfurreducens* are being investigated in order to better understand the response of *Geobacter* species to various environmental conditions that *Geobacter* species face during *in situ* uranium bioremediation and on the surface of electrodes harvesting electricity from waste organic matter and renewable biomass. Previous studies have demonstrated that these studies on regulatory mechanisms can provide important insights into the physiology and ecology of *Geobacter* species in these environments.

It is well known that sigma factors play an important role in regulating gene expression by recognizing promoter elements and initiating transcription. *G. sulfurreducens* has six genes encoding homologues of the sigma factors, RpoS, RpoN, RpoH, RpoD, RpoE, RpoN, and FliA. We have previously shown that RpoS is the stationary-phase sigma factor and is required for environmentally significant processes such as Fe(III) reduction and growth with oxygen. We are currently focusing on the alternative sigma factors, RpoN and RpoH. RpoN was constitutively expressed under all growth conditions examined. The *rpoN* gene appeared to be essential for growth, because a deletion mutant strain of the *rpoN* gene could not be obtained. When *rpoN* was overexpressed, growth was inhibited under a variety of environmentally relevant conditions, including growth in the absence of fixed nitrogen. These results suggest that proper expression of *rpoN* is important for optimum growth in the environment. A whole-genome DNA microarray comparison of gene expression between the wild-type strain and the *rpoN*-overexpressing strain revealed that RpoN regulated many genes involved in various cellular activities including pili and flagella biosyntheses. This is significant because of the important role of pili and flagella in growth in the subsurface and on the surface of electrodes.

Expression of the *rpoH* gene was induced, when the growth temperature was increased from 30°C to 42°C. A mutant in which *rpoH* was deleted grew normally at 30°C, but could not adapt to growth at 42°C. Furthermore, the expression of a number of heat-shock genes was undetectable or drastically decreased in the *rpoH*-deficient mutant. These results demonstrate that RpoH is the heat-shock sigma factor in *G. sulfurreducens* and essential for adaptation to growth at higher temperatures.

Two-component systems are an important strategy for adaptation to changes in environmental conditions. They typically consist of a sensor kinase, which senses an environmental signal, and a response regulator, which regulates gene expression to adapt to an environmental change. *G. sulfurreducens* contains an unusually large number of two-component systems, which may reflect the need to adapt to a myriad of different conditions in subsurface environments. A response regulator designated PilR was found to be important for the expression of *pilA*, which encodes the structural protein of the *G. sulfurreducens* pili. These pili are also referred to as microbial nanowires, because they are electrically conductive and are required for Fe(III) oxide reduction as well as optimal current production in microbial fuel cells. When *pilR* was deleted, it yielded a phenotype similar to the *pilA* mutant.

Fe(III) reduction as well as redox sensing were regulated by a two-component system consisting of a sensor kinase, which contains a unique sensor domain with *c*-type heme binding motifs and a region with similarity to the "redox box" of RegB kinase, and a response regulator in the enhancer binding protein family. Another response regulator with a unique output domain was found to be involved in a variety of cellular processes including cell division, biofilm formation, and hydrogen-dependent growth. Furthermore, a sensor kinase, which microarray analysis indicated was up-regulated during nitrogen fixation, was necessary for optimum growth in the absence of fixed nitrogen. These results demonstrate the important role of the two-component systems in *G. sulfurreducens* for adapting to environmental changes.

* Presenting author

G. sulfurreducens contains more than a hundred *c*-type cytochromes. Previous studies have demonstrated that deletion of some cytochrome genes inhibits Fe(III) reduction. However, it has recently become clear that some of these cytochromes are not directly involved in electron transfer, but rather are required for expression of other cytochromes that are key in electron transfer. Several more examples of this phenomenon have been discovered in the last year. For example, a mutant in which the gene for the cytochrome MacA was deleted could not express OmcB, an outer-membrane cytochrome that is required for Fe(III) reduction. When *omcB* was expressed with an exogenous constitutive promoter, the *macA*-deficient mutant reduced Fe(III) as well as the wild-type. A mutant in which the gene for the outer-membrane cytochrome OmcF was deleted was defective in Fe(III) reduction and electricity production. Microarray analysis revealed that genes involved in Fe(III) reduction and electricity production were down-regulated in the OmcF mutant.

Previous analysis of *in situ* gene transcript levels during *in situ* uranium bioremediation has demonstrated that levels of transcripts for the citrate synthase gene can be used to monitor the rates of *Geobacter* metabolism in the subsurface. Deletion of the citrate synthase gene eliminated the capacity for *G. sulfurreducens* to grow not only on acetate, but also with hydrogen as the electron donor when Fe(III) was the electron acceptor. Expression of the citrate synthase gene was regulated mainly at the level of transcription by a repressor. The binding site for the repressor was determined in the promoter of the citrate synthase gene and genome sequence analyses further identified sequences similar to the binding site within the promoter regions of other genes, most of which are likely involved in metabolic pathways including the TCA cycle, in *G. sulfurreducens* as well as other *Geobacter* species.

Overall, these studies have significantly added to the understanding of key modes of regulation in *Geobacter* species that are necessary for important physiological functions in the subsurface and on the surface of energy-harvesting electrodes. This permits us to continually increase the sophistication of our models for predicting how *Geobacter* species will respond to different natural environmental conditions or conditions that may be artificially imposed to promote bioremediation or energy harvesting.

108 GTL

Computational Analysis of Transcription Regulation of *Geobacter sulfurreducens*

Julia Krushkal[1]* (jkrushka@utmem.edu), Marko Puljic,[1] Ronald M. Adkins,[1] Jeanette Peeples,[1] Bin Yan,[1,2] Ching Leang,[3] Laurie N. DiDonato,[3] Cinthia E. Núñez,[3,4] Toshiyuki Ueki,[3] Radhakrishnan Mahadevan,[5] Brad Postier,[3] Barbara Methé,[6] and **Derek R. Lovley**[3] (dlovley@microbio.umass.edu)

[1]University of Tennessee Health Science Center, Memphis, Tennessee; [2]NIDCD, National Institutes of Health, Bethesda, Maryland; [3]University of Massachusetts, Amherst, Massachusetts; [4]Instituto de Biotecnología/UNAM, Cuernavaca, México; [5]University of Toronto, Toronto, Canada; and [6]The Institute for Genomic Research, Rockville, Maryland

Project Goals: The overall purpose of this project is to develop experimental and computational tools to predictively model the behavior of complex microbial communities involved in microbial processes of interest to the Department of Energy. The five year goal is to deliver in silico models that can predict the behavior of two microbial communities of direct relevance to Department of Energy interests: 1) the microbial community responsible for in situ bioremediation of uranium in contaminated subsurface environments; and 2) the microbial community capable of harvesting electricity from waste organic matter and renewable biomass. The research in this abstract

summarizes research under Subproject IV. The purpose of this subtask is to use computational methods to better understand transcriptional regulation of the expression of environmentally relevant genes in *Geobacter* species.

Geobacter species are of interest because of their role in *in situ* bioremediation of uranium and harvesting electricity from waste organic matter and renewable biomass. As part of our involvement in the Genomics:GTL *Geobacter* Project, we are investigating networks of regulatory interactions in this versatile group of microorganisms in order to elucidate molecular mechanisms of their regulatory response to environmental changes.

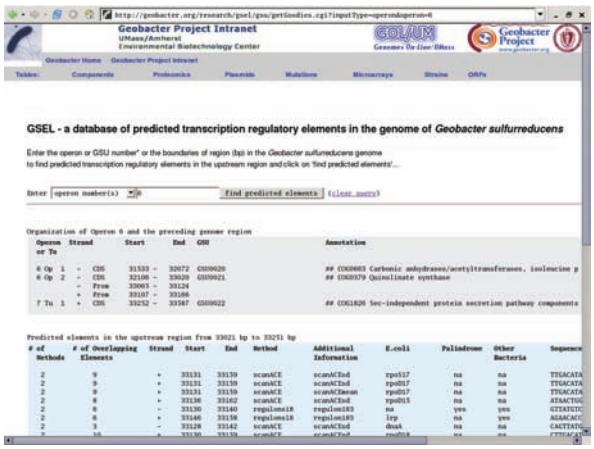


Figure 1. An example of GSEL query output listing predicted regulatory sites and operon structure in a region of the *G. sulfurreducens* genome.

To unveil transcriptional regulatory interactions affecting *Geobacter sulfurreducens* gene expression, we are employing a variety of computational strategies and utilizing a vast array of genome sequence data and gene expression information obtained in this project. In order to understand the complex interplay of multiple regulatory mechanisms, we not only catalogue individual genome sequence elements predicted to be involved in regulatory processes, but also apply data mining tools to those elements that appear to be involved in multiple regulatory pathways. This approach also allows us to identify those target genes that may be involved in important regulatory response mechanisms in a variety of conditions. In order to address these important questions, we have developed a database and an accompanying online query system, GSEL (*Geobacter* Sequence Elements) that compile information on putative transcription regulatory elements in the genome of *G. sulfurreducens* predicted by 10 different computational approaches based on pure *in silico* predictions and analysis of empirical data.

This online system allows users to query the genome of *G. sulfurreducens* using a specified genome region, operon number, or gene identifier (GSU). The output provides the predicted operon organization and the list of regulatory elements in the respective genome region ranked by the number of methods that predicted the site.

Individual regulatory sequence elements are predicted using analyses of individual microarray data sets or sequence data alone. For example, *G. sulfurreducens* has molybdate-responsive transcription factor, ModE (GSU2964). We identified 80 likely ModE binding sites in the genome of *G. sulfurreducens,* including likely functional sites in the upstream regions of (1) the *modABC* operon; (2) an operon containing gene NP_954455, a distant homolog of the *moaA* gene that encodes the molybdenum cofactor biosynthesis protein A in *Archaeoglobus fulgidus* and *Sulfolobus tokodaii*; (3) an

* Presenting author

operon that encodes a putative membrane protein (NP_954447) with homology to permeases of the drug/metabolite transporter (DMT) family (COG0697); and (4) glycine (CCC) tRNA.

Similarly, *G. sulfurreducens* contains RpoS (σS), a global regulator of gene expression in *G. sulfurreducens*, in addition to major housekeeping sigma factor RpoD (σ70). Our analysis of conservation and divergence of possible functional RpoD amino acid residues suggested that most of them are substantially conserved between *E. coli* and *G. sulfurreducens*. However, some degree of sequence variation between the two species was observed in several amino acid sites of RpoS proteins that might be important for promoter recognition. In addition, the similarity or identity of a number of residues between *G. sulfurreducens* RpoS and RpoD that might be functionally important suggests that the differences between the promoters recognized by RpoS and RpoD in *G. sulfurreducens* may be subtle. Using microarray gene expression information, we have been able to suggest a number of RpoS-regulated promoter elements as well as elements regulated by other sigma factors including RpoD. Experimental analysis of several promoters predicted to be RpoS-regulated and those predicted to be RpoD-regulated fully validated computational predictions.

In another example of analysis, we predicted gene regulatory interactions using information from the RelGsu regulon. RelGsu is the single *Geobacter sulfurreducens* homolog of RelA and SpoT proteins found in other organisms to be involved in regulation of levels of guanosine 3', 5' bispyrophosphate, ppGpp, a molecule that signals slow growth and stress response under nutrient limitation in bacteria. We used information obtained from genome wide expression profiling of the rel_{Gsu} deletion mutant to identify putative regulatory sites involved in transcription networks modulated by RelGsu or ppGpp. We identified likely sites regulated by Fur (ferric uptake repressor) in the upstream regions of upregulated operons and RpoS-regulated promoters in the upstream regions of the downregulated operons of the rel_{Gsu} deletion mutant. These findings suggest that Fur- and RpoS-dependent gene expression in *G. sulfurreducens* is affected by ppGpp-mediated signaling.

Among multiple other sequence and gene expression analyses by our group of regulatory interactions influenced by specific transcription factors, our most recent analyses involve prediction of the promoter elements regulated by alternative sigma factor RpoN (σ54). We predicted 467 RpoN-regulated promoter elements that had the same orientation with their downstream target genes or operons, including 110 such elements in the noncoding regions. We identified those promoters for which the expression of their target genes was significantly altered in the RpoN gene overexpression microarrays. Further analyses focus on the function of the specific genes whose regulation may be significantly affected by RpoN and on their possible role in different environmental conditions.

Data mining of predicted regulatory interactions allowed us to identify genome regulatory regions and their target operons that are involved in a variety of regulatory pathways. This powerful approach allows us to identify gene products that may be central to *G. sulfurreducens* response to a variety of trigger conditions, to find genes and operons whose expression may be altered in response to very specific sets of conditions, and to suggest the molecular mechanisms of their regulation.

109 GTL

Identification of Small Non-Coding RNAs and Acceptance Rate Studies in Members of the *Geobacter*aceae

Barbara Methé[1]* (bmethe@tigr.org), Robert DeBoy,[1] Sean Daugherty,[1] Ty Arrington,[1] Kelly Nevin,[2] Jonathan Badger,[1] and **Derek Lovley**[2]

[1]The Institute for Genomic Research, Rockville, Maryland and [2]University of Massachusetts, Amherst, Massachusetts

Project Goals: This abstract submission is a part of the project "Genome-Based Models to Optimize In Situ Bioremediation of Uranium and Harvesting Electrical Energy from Waste Organic Matter" (PI- Lovley) and is focused on the application of comparative genomic and functional genomic techniques to improve our understanding of the evolution and regulation of members of the *Geobacteraceae*.

Members of the *Geobacteraceae* are the subject of intense study as they are the dominant dissimilatory metal-reducing microorganisms in subsurface environments in which organic contaminants are being degraded with the reduction of Fe(III) and in aquatic sediments where dissimilatory metal reduction is important. Further they are of great interest for the practical roles that they can play as agents of bioremediation and in energy production. The completion of genome sequence from multiple members of the *Geobacteraceae* provides the opportunity to apply genome-level analyses to obtain fundamentally new insights into their evolution and regulation.

Predicting the presence and function of protein coding genes has traditionally been an important area of focus in microbial genomics. More recently, a significant interest has developed in the study of chromosomally located small non-coding (sRNAs) which is being aided by the completion of genome sequence from related organisms. Increasing evidence suggests that sRNAs exist in numerous organisms where they play important regulatory roles including responses and adaptations to different stresses.

To identify novel sRNAs in *Geobacter sulfurreducens* and other *Geobacteraceae* we are adapting the computer program, sRNAPredict2, which uses an integrative computational approach to identify sRNAs in bacterial genomes. This program relies on sequence conservation outside of protein-coding genes and the locations of predicted Rho-independent terminators. Several supporting programs which provide input files for sRNAPredict2 are also being used to improve sRNA prediction. These supporting programs provide the locations of protein-encoding genes, rRNAs and tRNAs, a small number of previously predicted sRNAs, predicted terminator sequences and putative conservation of RNA secondary structure in intergenic regions. Currently a total of 80 putative sRNA genes have been identified in *G. sulfurreducens*, of which more than 30 were predicted in both sets of pairwise comparisons with the genomes of *G. metallireducens* and *G. uraniumreducens* and the sequences of six of these are closely related to one another based on primary and predicted secondary sequence structure. The longer term goal of this effort is to experimentally characterize these predictions using a combination of traditional molecular biological techniques such as Northern blots and primer extensions and to use microarray technology in the form of an oligonucleotide tiling array of the complete *G. sulfurreducens* genome.

Microbial genomes are not static entities. The rate and degree of genomic plasticity can best be understood in the framework of evolutionary processes. Rates of nucleotide substitution in genome sequence are a composite of the occurrence of mutations, random genetic drift of neutral or nearly

* Presenting author

neutral alleles and purifying selection against deleterious alleles. In a small number of cases, directional selection (positive selection) occurs when natural selection favors a single allele and therefore allele frequency continuously shifts in that direction. Nonsynonymous mutations result in amino acid replacement while synonymous or silent mutations cause no change in the specified amino acid. Since advantageous mutations undergo fixation in a population more rapidly than neutral mutations the rate of nonsynonymous substitution will exceed that of synonymous substitution if advantageous selection plays a role in the evolution of the protein in question. One way to detect positive Darwinian selection is to determine if the number of substitutions per nonsynonymous site is significantly greater than the number of substitutions per synonymous site.

We are studying the rates of nonsynonymous to synonymous (d_N/d_s) substitutions) across gene families of interest in members of the *Geobacteraceae* and other selected genomes. In a study examining d_N/d_s ratios in c-type cytochromes for example, 26 orthologous gene clusters were identified in at least 5 of the following 10 genomes: *Geobacter sulfurreducens*, *G.* sp. FRC-32, *G. metallireducens*, *G. uraniumreducens*, *Pelobacter carbinolicus*, *P. propionicus*, *Desulfuromonas acetoxidans*, *Desulfovibrio desulfuricans*, *D. vulgaris* and *Rhodoferax ferrireducens*. All 26 clusters were determined to have d_N/d_s ratios of less than one signifying that members of these clusters are undergoing purifying selection. However, examination of amino acid sites within the protein sequences revealed sites with statistically significant evidence of positive selection. For example, four sites have been determined to be undergoing positive selection within a c-type cytochrome for which a three-dimensional structure has been solved (10S6). These sites are located at positions 22, 25, 28 (near ligand to Heme III at His20) and 45 (near Heme IV at His47). A further systematic evaluation of sequences from the *Geobacter*aceae is currently underway. These investigations are of relevance to protein engineering as understanding which residues have changed historically may suggest sites that can be mutated to change the specificity and metabolic characteristics of the protein in question.

110 GTL

Bacterial Cell Cycle Control System and a Control System Simulation Model

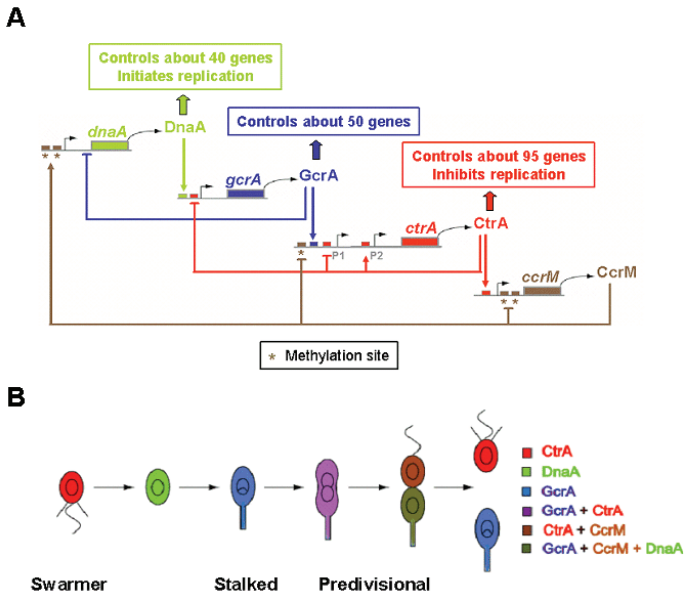
J. Collier,[1] X. Shen,[2] M. Horowitz,[2] L. Shapiro,[1] and **H. H. McAdams**[1]* (hmcadams@stanford.edu)

[1]Department of Developmental Biology and [2]Department of Electrical Engineering, Stanford University, Stanford, California

Project Goals: Identification of the overall regulatory and metabolic networks in *Caulobacter crescentus*, largely through gene expression microarray assays and bioinformatic analysis

A core cell cycle engine involving interlinked master regulator proteins drives cell cycle progression in bacterial cells as in eukaryotic cells. The cell cycle control system in the model bacteria *Caulobacter crescentus* is hierarchically organized, and it activates functional genetic modules 'just in time' when needed, including two processive subsystems, chromosome replication and cytokinesis. The core engine and modular cell cycle subfunctions are repeatedly re-synchronized by non-genetic mechanisms, and checkpoint signaling between modules also acts to assure proper ordering of subfunctions. Each top-level regulatory protein is regulated at multiple levels, e.g., transcription, stability, and activity. In *Caulobacter*, three essential oscillating transcriptional regulators, DnaA, GcrA, and CtrA, control over 200 temporally controlled genes. CtrA also blocks the initiation of chromosome replication, while DnaA promotes it, by directly binding to the origin of replication (*Cori*) and allowing

replisome formation. Elimination of DnaA halts the cell cycle, and control of DnaA stability is a mechanism used by stress sensors to halt the cell cycle. DnaA activates the synthesis of GcrA, which in turn activates the synthesis of CtrA that activates the synthesis of the CcrM DNA methyltransferase. This cascade of master regulator proteins drives the forward progression of the *Caulobacter* cell cycle as shown in Figure 1.



Cyclical oscillator drives the cell cycle. We have shown that a critical element of the *Caulobacter* cell cycle regulatory circuit is the link between *dnaA* expression and the progression of DNA replication. We have demonstrated that the synthesis of DnaA is coordinated with its function by an epigenetic mechanism of regulation, through the methylation of the *dnaA* promoter by the cell cycle-regulated CcrM protein. The *dnaA* promoter contains two DNA methylation sites. We have also shown that these two methylation sites are essential for efficient *dnaA* transcription. When the two methylation sites in the *dnaA* promoter are removed by targeted mutagenesis at the *dnaA* locus, the level of DnaA becomes limiting in cells, resulting in abnormal cell phenotypes. Furthermore, the transcription of *dnaA*, which resides near the *Cori*, is efficient when the *dnaA* promoter is in the fully-methylated state prior to the initiation of DNA replication, but inefficient when the *dnaA* promoter becomes hemimethylated upon passage of the replication fork, soon after DNA replication initiation. Accordingly, the chromosomal location of *dnaA* near the *Cori* is an important component of this regulatory mechanism, which contributes to the changes in DnaA cellular content during the cell cycle. We conclude that the induction of the master regulatory cascade is tied to the replication status of the chromosome by CcrM, which is the last element of this cascade. One major consequence of this finding is that the DnaA/GcrA/CtrA/CcrM master regulatory cascade is a cyclic process, where CcrM activates DnaA at the end of the cell cycle to start a new cell cycle (Figure 1).

Engineering simulation of the *Caulobacter* cell cycle control system. We have approached analysis of the circuitry from an engineering perspective, using computer tools and analysis paradigms drawn from electrical engineering. The results demonstrate both how the *Caulobacter* cell cycle regulatory mechanisms functions as a system and how the circuit is designed for reliable or robust operation in spite of noisy components and highly uncertain operating environments. From the EE circuit design perspective, the cell cycle control circuit approximates a fundamental mode asynchronous state machine with design features long known to electrical engineers as necessary for robust operation. Engineers exploit this FASM class of circuits in situations that are parallel to the cell control problem. For example, there are unclocked electrical systems with concurrent modular subsystems and wide variability in component performance as in the biological cell cycle circuitry. The *Caulobacter* cell cycle circuit design is robust to parameter variation, and it can provide reliable operation over a wide range of growth rates (i.e., generation times) consistent with the requirements of an organism adapted to

* Presenting author

low and uncertain nutrient levels. Additional reliability is provided by the elaborate and redundant regulatory mechanisms that control the most important regulatory proteins, such as CtrA and DnaA.

Since cell cycle regulation involves decisively turning modular functions on and off, the regulatory mechanisms tend to function as binary switches as in many areas of developmental regulation in all organisms. Rapid switching between qualitatively different stable cell states can result from bistable regulatory circuits, that is, circuits that exhibit hysteresis and require changes in input signals to initiate a transition between states. A central element of the design of the *Caulobacter* cell cycle regulatory circuit is a bistable switching element resulting from positive autoregulation of the *ctrA* gene encoding the CtrA master regulator protein. Changes in a phosphosignal originating from the CckA histidine kinase initiate switching between high and low CtrA activity.

Since *Caulobacter* cells divide asymmetrically, producing siblings with significantly different morphology and cell fates, its cell cycle control circuitry has to initialize each daughter cell's control system differently consistent with each cell's individual regulatory program. Dynamic localization of regulatory proteins and proteolytic subsystems to the cell poles is essential to asymmetric cell division. The distinctive identity of the subsequent daughter cells, each containing one of the chromosomes of the predivisional cell, begins at the instant of cytoplasmic compartmentalization about 18 minutes before daughter cell separation. Immediately upon compartmentalization, differentiation begins owing to isolation of key phosphorylation-dependent regulatory proteins from their cognate kinases or perhaps to differential sequestering of a phosphatase. Large differences in binding affinity between the phosphorylated and unphosphorylated response regulators causes gene expression profiles in the compartments to diverge yielding different development programs thereafter with profound consequences for the fates of the two daughter cells.

We have developed a system-level computer simulation of the *Caulobacter* cell cycle and asymmetric cell division control system using the Matlab control system simulation tools, Simulink and Stateflow, that are widely used in engineering analysis. The simulation model provides verification of the operation of the cell cycle control system design by comparing a molecular level simulation of the system with experimental observations of changing protein and mRNA levels over the cell cycle. We find excellent correlation between protein and mRNA levels predicted by the simulation model and experimental results. The simulation model of the *Caulobacter* cell cycle is inherently extensible. In the current version, we include detailed models of the chromosome replication and cytokinesis submodules because progress of these two subsystems is so essentially integrated with progress of the cell cycle engine, but future extensions can add regulation of phosphosignaling, polar organelle development, and blocking of cell cycle progress under limited resource and stress conditions. This top-down incremental development paradigm is a promising avenue for development of whole cell behavioral simulation models that can both emulate the observed behavior of the cell and predict the outcome of genetic changes.

Modeling biological regulatory circuits entirely as systems of ordinary differential equations is inherently self-limiting. Biological control circuits have elements that are more readily simulated with hybrid models that combine differential equation-based models of subsystems with state machine models and *ad hoc* behavioral models. Our results suggest that the attempts to model regulatory networks assuming a transcriptional regulation paradigm without consideration of both dynamic localization of regulatory proteins and of epigenetic chromosomal modifications are unlikely to approximate biological reality.

111 GTL

Automated Accurate, Concise and Consistent Product Description Assignment for Microbial Regulatory Proteins

Loren Hauser* (hauserlj@ornl.gov), Frank Larimer, and **Miriam Land**

Life Sciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee

Project Goals: The project is to develop and continually improve an automated microbial genome annotation pipeline. The current project is the newest bioinformatic tool to be added to the pipeline. This tool is designed to identify, categorize and assign consistent product descriptions for most of the regulatory proteins in every annotated microbial genome.

US DOE's Joint Genome Institute (JGI), which consists of groups at the Production Genome Facility in Walnut Creek CA, LBL, LLNL, LANL, ORNL and Stanford, is scheduled to sequence over 120 microbial genomes in 2007. ORNL has developed and implemented an automated annotation pipeline for the initial annotation of these genomes. The pipeline's multiple tools and database queries are used to create a reference web site for each genome that is used as the basis for both automated and manual annotation. Although most consider manual annotation optimal, it varies in quality, is time consuming, and does not scale with the increase in genomic sequencing. Therefore, it is not feasible for the majority of genomes. However, accurate automated annotation of most of an organism's gene products which contain characterized protein domains is feasible. Complex multidomain proteins, such as signal transduction histidine kinases and other regulatory proteins represent a unique challenge due to the number and variety of domains that are possible in a single protein. A rule based annotation system combining the output from searches of Interpro, COGs, and Swissprot-TREMBL, as well as, TMHMM predictions and the domain architecture of individual proteins has been developed in order to provide accurate, concise and consistent product descriptions for most microbial regulatory proteins. This multidimensional approach increases both the accuracy and sensitivity when compared to the use of single systems such as SMART or simple blast searches. The system provides a list of all the regulatory proteins it has identified, the product descriptions that will be assigned, and all the information used in their identification. The product description nomenclature scheme is being applied to all genomes in IMG so that searches will find all members of gene families regardless of the quality or age of the original annotation. In addition, it will provide a numerical synopsis in the form of a table of the regulatory proteins organized by the effector or output domain, which can be directly used in genome publications. This system has recently been incorporated into the automated annotation pipeline. Similar annotation tools are under development here and at other JGI sites.

* Presenting author