

Microbial Genomics

B11

Strategies to Harness the Metabolic Diversity of *Rhodopseudomonas palustris*

Caroline S. Harwood¹ (caroline-harwood@uiowa.edu), Jizhong Zhou², F. Robert Tabita³, Frank Larimer², Liyou Wu², Yasuhiro Oda¹, Federico Rey¹, and Sudip Samanta¹

¹The University of Iowa; ²Oak Ridge National Laboratory; and ³Ohio State University

Rhodopseudomonas palustris is an extremely successful bacterium that can be found in virtually any temperate soil or water sample on earth. It can grow by adopting one of the four major metabolic modes: photoautotrophic growth (energy from light and carbon from C1 compounds), photoheterotrophic growth (energy from light and carbon from organic compounds), chemoheterotrophic growth (carbon and energy from organic compounds) and chemoautotrophic growth (energy from inorganic compounds and carbon from C1 compounds). *Rhodopseudomonas* enjoys exceptional versatility within each of these growth modes. It can grow with or without oxygen and can use many alternative forms of carbon, nitrogen and inorganic electron donors. It degrades plant biomass and chlorinated pollutants and shows promise as a catalyst for biofuel production. *Rhodopseudomonas* has thus become a model to probe how the web of metabolic reactions that operates in the confines of a single cell adjusts in response to subtle changes in environmental conditions. Genes for key metabolic enzymes and regulatory proteins are easily identified in the 5.49 Mb *Rhodopseudomonas* genome. It has a large cluster of photosynthesis genes and a collection of additional genes that encode light-responsive proteins. It has genes for the metabolism of diverse kinds of carbon sources, including lignin monomers, complex fatty acids and dicarboxylic acids. It encodes two different carbon dioxide fixation enzymes and three different nitrogen fixation enzymes, each with a different transition metal at its active site. Each of the three nitrogenases is active in *Rhodopseudomonas* and each catalyzes the conversion of nitrogen gas to ammonia and hydrogen, a biofuel.

Rhodopseudomonas can convert sunlight to ATP and derive electrons by biodegrading plant material. ATP and electrons so generated can, in turn, be used to fix nitrogen, with accompanying hydrogen production. Because multiple systems are involved, hydrogen production is a good starting point for studies or integrative metabolism by *Rhodopseudomonas*. To achieve efficient hydrogen production it is important to understand how expression levels of genes involved in photosynthesis, carbon dioxide fixation, lignin monomer degradation and nitrogen fixation fluctuate in response to variations in conditions. It is also important to identify regulatory bottlenecks that restrict the flow of energy and electrons from plant biomass to hydrogen production. Towards this end we have constructed a whole genome DNA microarray of *Rhodopseudomonas* and we are now using the array to analyze global patterns of gene expression.

B13

Gene Expression Profiles in *Nitrosomonas europaea*, an Obligate Chemolithoautotroph

Dan Arp¹ (arpd@bcc.orst.edu), Xueming Wei¹, Luis Sayavedra-Soto¹, Martin G. Klotz², Jizhong Zhou³, and Tingfen Yan³

¹Oregon State University; ²University of Louisville; and ³Oak Ridge National Laboratory

Nitrosomonas europaea derives energy for growth from the oxidation of ammonia to nitrite. This process contributes to nitrification in soils and waters, often with detrimental effects in croplands, and with beneficial effects in wastewater treatment. Our long-range goal is to understand the molecular underpinnings for the oxidation of ammonia and other cellular processes carried out by these organisms. Towards this goal, the genome of this bacterium was sequenced at the Lawrence Livermore National Laboratory (Jane Lamerdin, Patrick S. Chain). Through a collaborative effort with at Oak Ridge National Lab-

oratory, we are developing microarrays in order to examine whole genome expression profiles for this organism. We are interested in understanding the effects of nutrient shifts, starvation, and other environmental changes on gene expression. The arrays are now constructed and preliminary results will be presented.

Prior to the initiation of this project, the expression of the genes coding for the enzymes involved in ammonia oxidation were characterized. These genes are *amoCAB*, coding for ammonia mono-oxygenase, and *hao*, coding for hydroxylamine oxidoreductase. The *amo* genes in particular show a strong response to ammonia. In preparation for the microarray experiments, we were interested in learning more about the expression of the genes coding for Rubisco. This autotroph assimilates CO₂ via this enzyme and the Calvin Cycle. Sequence data reveals that *N. europaea* has a type I Rubisco. The Rubisco operon in *N. europaea* likely consists of five open reading frames. Although Rubisco is essential for the autotrophy of this organism, studies on its expression are scant. We analyzed mRNA levels of Rubisco and some other genes by Northern hybridization with specific probes. Rubisco large and small subunits (encoded by *cbbL* and *cbbS*) were highly expressed in growing cells. The message levels appeared higher than those of *amo* and *hao*. In ammonia-deprived and stationary phase cells, Rubisco mRNAs were undetectable. Particularly interesting is that the expression of Rubisco genes was inversely proportional to the carbon levels in the medium. Rubisco mRNAs in cells grown in a medium with only atmospheric CO₂ were several times higher than those in cells in carbonate-containing medium. The higher the carbonate level in the medium, the lower the Rubisco mRNA levels. This result was in contrast to house keeping genes such as *hao* and carbonic anhydrase genes.

We also investigated the induction of the *cbb* operon and its message depletion patterns. After 30 min induction in normal medium, abundant *cbbL* and *cbbS* messages were detected and they were expressed fully after one hr. The estimated halflives of *cbbL* and *cbbS* were 0.5 and 0.75 hr, which were shorter than the half lives of *amo* and *hao*.

The three remaining *cbb* genes in the operon were expressed at much lower levels than *cbbL* and *cbbS*. This result may be explained by the presence of a transcription terminator downstream of *cbbS*, as revealed in the DNA sequence data. These results indicated that Rubisco gene expression was

dependent on ammonia, and that carbon had a negative control on Rubisco transcription.

B15

Genomics of *Thermobifida fusca* Plant Cell Wall Degrading Proteins

David B. Wilson (dbw3@cornell.edu), Yuan-Man Hsu, and Diana Irwin

Department of Molecular Biology & Genetics, Cornell University, Ithaca, NY 14853

Plasmids have been successfully introduced into *Thermobifida fusca* YX by mating to *E. coli* cells containing a transfer plasmid and selecting for thiostrepton resistance. We have constructed a suicide plasmid with a defective *celR* gene containing an insert and are trying to knockout the *celR* gene in *T. fusca*. One puzzling result is that mating does not work in *T. fusca* strain ER₁, which was isolated from *T. fusca* YX by mutagenizing spores with methyl sulfate and selecting for a colony lacking the major extracellular protease. The two strains should only differ by a few point mutations and it seems unlikely that inactivating the protease would interfere with mating or plasmid transfer.

We have continued our study of *T. fusca* XG74, the main *T. fusca* xyloglucanase. Surprisingly, *T. fusca* does not grow on xyloglucan and this appears to be due to the lack of a transport system for taking up the products of xyloglucan hydrolysis since they accumulate in the media of *T. fusca* cells incubated with xyloglucan. We have shown that XG74 does allow a mixture of *T. fusca* cellulases to hydrolyze cellulose coated with xyloglucan, which is not hydrolyzed by the mixture containing only cellulases. The mixture containing XG74 cannot degrade the cellulose in tomato cell walls, but *T. fusca* crude cellulase will it. We will try to identify the additional proteins that are required for hydrolysis. The level of Xg74 in the culture supernatant of *T. fusca* grown on different carbon sources was determined by Western blotting and it was low on glucose or cellobiose slightly higher on xylan and high on corn fiber, Sulka Floc or xyloglucan.

B17

The *Rhodospseudomonas palustris* Microbial Cell Project

F. Robert Tabita¹ (tabita.1@osu.edu), Janet L. Gibson¹, Caroline S. Harwood², Frank Larimer³, Thomas Beatty⁴, James C. Liao⁵, Jizhong (Joe) Zhou³, and Richard Smith⁶

¹Ohio State University; ²University of Iowa; ³Oak Ridge National Laboratory; ⁴University of British Columbia; ⁵University of California at Los Angeles; and ⁶Pacific Northwest National Laboratory

The long-range objective of this interdisciplinary study is to examine how processes of global carbon sequestration (CO₂ fixation), nitrogen fixation, sulfur oxidation, energy generation from light, biofuel (hydrogen) production, plus organic carbon catabolism and metal reduction operate in a single microbial cell. The recently sequenced *Rhodospseudomonas palustris* genome serves as the raw material for these studies since the metabolic versatility of this organism makes such studies both amenable and highly feasible. Bioinformatics analysis has allowed for a reasonable approximation of many of the metabolic schemes that are utilized by this organism to catalyze the above processes. However, it appears from recent studies that several of the above processes are coordinately controlled and interdependent; thus during the first part of this project we have focused at identifying key regulatory genes and proteins. From the genomic sequence a number of likely target genes of opportunity were also revealed. These have been systematically knocked out to produce a battery of useful mutant strains that are employed in a variety of studies to examine the regulation of metabolism. In addition, we have taken advantage of a transposon library in which virtually all the open reading frames of the genome have been interrupted. Screening this transposon library under diverse growth conditions has enabled us to identify several additional unique and previously unappreciated genetic loci that are important for the above processes. The latter mutant strains are currently being studied to reveal exactly how these newly identified genes and their products influence the processes under study.

In addition to these more traditional approaches, we have also undertaken genomics, proteomics, and metabolomics oriented experiments to further analyze the integrative control of metabolism, with the rationale that these approaches will help

direct our future investigations and focus our efforts. For example, whole genome microarrays have been prepared for *R. palustris* and initial studies have commenced that are assisting our efforts to identify additional genes that are up and down regulated under growth conditions of interest, with examination of both wild-type and mutant strains underway or contemplated in the near future. Furthermore, a Bioinformatics method was developed to deduce operon structure using microarray data and gene distance. This method allows the refinement of operon prediction based on genomic information. In contrast to the theoretical prediction based only on genomic information, this method incorporates microarray data and considers the noise level in the microarray experiments. In parallel with microarray experiments, examination of the proteome has commenced under selected growth conditions, using both whole cells and isolated intracytoplasmic membranes (these are intracellular structures that are employed for photochemical energy generation by this and related organisms). Proteomic analysis provides a real advantage as one can examine the end result of transcription and translation under different physiological growth conditions and use this protein data to relate back to the regulation of gene expression and/or potential posttranslational events.

The above experimental approaches will allow us to reach the eventual goal of this project; i.e., to generate the knowledge base to model metabolism for the subsequent construction of strains in which carbon sequestration and hydrogen production are maximized.

B19

Lateral Gene Transfer and the History of Bacterial Genomes

Scott R. Santos and **Howard Ochman**

Department of Biochemistry and Molecular Biophysics, University of Arizona, Tucson, AZ 85721

Comparative analyses of complete microbial sequences have brought many new insights into the evolution of genomes and the genetic relationships among microorganisms. For the vast majority of microbial species, molecular phylogenetic relationships have been on a single gene, i.e., small subunit ribosomal DNA (16S rDNA). Because 16S rDNA is highly conserved – both in

terms of its function and distribution in all lifeforms and its rate of change – it is particularly well-suited for resolving the relationships among very divergent organisms. However, recent studies provide clear evidence that lateral gene transfer is common among bacteria with the result that genomes are chimeric, such that different regions will have very different histories.

The long-range objectives of our research is to employ nucleotide sequences of a large set of universally distributed genes among eubacteria of differing degrees of genetic relatedness in order to address questions relating to the role of gene transfer in shaping bacterial genomes. We are using existing databases as well as newly determined nucleotide sequences in order to design conserved primers for PCR amplification and sequencing of nucleotide sequences across taxa. These primer sets were initially derived from alignments of 143 protein coding genes common to the majority of eubacteria. Of these, conserved primer sets, i.e., those of limited degeneracy that yield products 400-2000 bp in length, could be obtained for 21 of these genes. Primer sets were then tested against DNA templates from representatives of diverse bacteria phyla, and the initial screening identified nine genes that could be amplified reliably from over 50% of the isolates.

The target genes for which we have developed conserved primer sets are involved in a variety of cellular functions, including translation, ribosomal structure and biogenesis (*fusA*, *ileS*, *leuS*, *rplB*, *valS*), DNA replication, recombination and repair (*gyrB*), cell motility and secretion (*lepA*), nucleotide transport and metabolism (*pyrG*) and transcription (*rpoB*). This diversity makes them ideal candidates to test if genes subject to lateral transfer are functionally or genetically linked and, to test the phylogenetic limits to lateral gene transfer. We are also developing primer sets for proteins specific to members of the specific phyla (e.g., *Proteobacteria*, *Spirochaetes*) to explore the ancestry of taxa-specific genes.

B21

Environmental Sensing, Metabolic Response, and Regulatory Networks in the Respiratory Versatile Bacterium *Shewanella oneidensis* MR-1

James K. Fredrickson¹

(jim.fredrickson@pnl.gov), Margie F. Romine², William Cannon², Yuri A. Gorby², Mary S. Weir-Lipton², H. Peter Lu², Richard D. Smith², Harold E. Trease², and Shimon Weiss²

¹Pacific Northwest National Laboratory; and ²University of California at Los Angeles

Shewanella oneidensis MR-1 is a motile facultative bacterium with remarkable metabolic versatility in regards to electron acceptor utilization; it can utilize O₂, nitrate, fumarate, Mn, Fe, and S⁰ as terminal electron acceptors during respiration. This versatility allows MR-1 to efficiently compete for resources in environments where electron acceptor type and concentration fluctuate in space and time. The ability to effectively reduce polyvalent metals and radionuclides, including solid phase Fe and Mn oxides, has generated considerable interest in the potential role of this organism in biogeochemical cycling and in the bioremediation of contaminant metals and radionuclides. In spite of considerable effort, the details of MR-1's electron transport system and the mechanisms by which it reduces metals and radionuclides remain unclear. Even less is known regarding the molecular networks in this organism that allow it to respond to compete efficiently in a changing environment. The entire genome sequence of MR-1 has been determined and high throughput methods for measuring gene expression are being developed and applied.

DOE recognized that in order to achieve the goals of the Genomes to Life Program, obtaining a comprehensive systems-level level understanding of the components and functions of the cell that give it life, a united effort integrating the capabilities and talents of many would be required. To this end, the *Shewanella* Federation was formed to probe in detail the functions of *Shewanella oneidensis* MR1 cells. The Federation consists of teams of scientists from academia, national laboratories, and private industry (shewanella.org) that are working together in a collaborative, coordinated mode to jointly achieve a comprehensive understanding of the biology of this remarkably versatile organism. This project is contributing to

the collaborative experiments of the Federation by providing, among other things, characterized chemostat cultures of MR-1 and high resolution separation and high mass accuracy and sensitivity Fourier transform ion cyclotron resonance (FTICR) mass spectrometry for global proteome analyses of these cultures. To date, >50% of the predicted 5000+ proteins in MR-1 have been identified by accurate mass tags (AMTs) with an average of 3 AMTs per protein.

To date, we have established procedures for growing MR-1 in continuous culture under both aerobic and anaerobic (with fumarate) conditions with lactate as the growth-limiting nutrient for the initial collaborative *Shewanella* experiments. The initial results from 2-D PAGE analysis of proteins by C. Giometti (ANL) revealed that there were substantial variations in the proteome in the aerobic vs. anaerobic cultures and that biological replicates of chemostat samples were in excellent agreement. Microarray based analyses of mRNA expression are underway in collaboration with J. Zhou (ORNL) as is MS-based proteome analyses at PNNL. An unanticipated result was the formation of flocs in the aerobic chemostat. Floc formation was due to the production of exopolymeric substance (EPS) by MR-1 and was hypothesized to be due to a defense mechanism against O₂ stress (i.e., induced by O radicals). In the absence of Ca, flocs are unable to form likely due to a lack of cross-linking of the EPS. We are currently generating additional MR-1 continuous cultures in the absence of Ca under aerobic and anaerobic conditions as well as under O₂ limiting conditions. The major goal of these experiments is to provide insights into gene and protein expression patterns under these conditions as a baseline for future experiments with other types of electron acceptors, such as metals and radionuclides or nitrate, and for identifying genes that are potential targets for mutagenesis.

Another approach for characterizing differential gene expression can be provided by analysis of reporter activity mediated by transcriptional fusions. Because reporter activity can be measured in living cells in real-time, the use of transcriptional fusions is more amenable than are microarrays to dynamic measurements of gene expression analysis under many growth conditions and measurements can be made at the level of individual cells as opposed to a bulk average of the entire population. We describe the construction of a small targeted reporter library (62 constructs) in MR-1, whereby promoter-containing DNA sequences upstream to genes associated

with electron transport, adhesion, and cell signaling were cloned in a broad-host range plasmid upstream to the green fluorescent protein (GFP). Using MR-1 bearing GFP reporter constructs, we have demonstrated that 1) the vector, pProbe-NT, utilized is stable after several passages in the absence of antibiotic, 2) GFP is stable for at least a week, and 3) very little time is needed for fluorescence development, even where cells are grown anaerobically prior to making fluorescence measurements. Initial testing using a crude assay developed to measure the effect of growth in suspension versus on solid surfaces, suggest that expression of the promoter upstream to *mtrDEF* is significantly higher on surfaces than in suspension. Similar, but not so dramatic, effects were observed for other promoter constructs. Proposed methods for high throughput analyses with these constructs and the utility of these assays for design of complementary microarray analyses are underway.

Outer membrane vesicles (MVs) are unique to Gram-negative bacteria, are initiated by the formation of “blebs” in the outer membrane, and are released from the cell surface during growth, trapping some of the underlying periplasmic contents in the process. Membrane vesicles provide an excellent means to identify proteins that are localized to the outer portions of the MR-1 cell envelope without disturbing cellular integrity or the need to further fractionate cells. Mass spectrometric analyses of vesicles isolated from MR-1 cells grown on LB supplemented with fumarate and lactate revealed the presence of 18 outer membrane and 12 periplasmic proteins. Proteins that were identified include electron transport pathway components (OmcA, OmcB, MtrB, CymA, fumarate reductase, and formate dehydrogenase alpha and Fe-S subunits), five putative porins, three proteases, proteins involved in protein maturation (PpiD and DsbA), and two transport proteins (long-chain fatty acids and tungstate). In addition, these samples contained FlaA flagellin proteins and the MshA pilin protein, head and tail proteins from prophage LambdaSo and MuSo2 which, along with several other putative inner membrane and cytoplasmic proteins, probably co-purified with vesicles. The presence of phage coat proteins in these samples suggests that a fraction of cells within MR-1 cultures are undergoing lysis during culture and may explain why proteins predicted to be associated with the inner membrane or cytoplasm were also detected in MV preparations. The presence of electron transport proteins shown *in vitro* to be capable of reducing Fe(III) is consistent with related findings in *S.*

putrefaciens CN32, a close relative of MR-1, where vesicles have been shown to mediate Fe(III), U(VI) and Tc(VII) reduction.

A64

Interdisciplinary Study of *Shewanella oneidensis* MR-1's Metabolism and Metal Reduction

Eugene Kolker (ekolker@biatech.org)

BIATECH (www.biatech.org), 19310 N. Creek Parkway, Suite 115, Bothell, WA 98011, 425.481.7200 x100, Fax: 425.481.5384

Since our project became part of the Shewanella Federation, we focused our work mostly on analysis of different types of data produced by global high-throughput technologies to characterize gene and protein expression as well as getting a better understanding of the cellular metabolism. Specifically, first year activities include development of:

1. New labeling technique for quantitative proteomics, so called methyl esterification labeling approach, complementary to currently available methods;
2. New algorithm for de novo protein sequencing;
3. New statistical model for spectral analysis of arbitrary shape data;
4. One of the first analyses of the transcriptome of the entire microorganism;
5. New approach to predict operon structures and transcripts within untranslated regions;
6. The first control protein experimental mixtures with known physico-chemical characteristics for high-throughput proteomics experiments;
7. The first statistical models for peptide and protein identifications for high-throughput proteomics analysis;
8. *Shewanella* metabolic capability experiments with minimal media on aerobically & anaerobically grown cells and transformation experiments.

Several collaborations have been established within the *Shewanella* Federation with PNNL, USC, ORNL, and MSU. The first year of this project, supported by DOE's Offices of Biological and Environmental Research and Advanced Scientific Computing Research, also resulted in 6 published papers.

This is a joint work of A. Keller, A. Nesvizhskii, A. Picone, B. Tjaden, D. Goodlett, S. Purvine, S. Stolyar, and T. Cherny done at BIATECH and ISB.

B23

Integrated Analysis of Protein Complexes and Regulatory Networks Involved in Anaerobic Energy Metabolism of *Shewanella oneidensis* MR-1

Jizhong Zhou¹ (zhouj@ornl.gov), Dorothea K. Thompson¹, Matthew W. Fields¹, Adam Leaphart¹, Dawn Stanek¹, Timothy Palzkill², Frank Larimer¹, James M. Tiedje³, Kenneth H. Nealson⁴, Alex S. Beliaev⁵, Richard Smith⁵, Bernhard O. Palsson⁶, Carol Giometti⁷, Dong Xu¹, Ying Xu¹, Mary Lipton⁵, James R. Cole³, and Joel Klappenbach³

¹Oak Ridge National Laboratory; ²Baylor College of Medicine; ³Michigan State University; ⁴University of Southern California; ⁵Pacific Northwest National Laboratory; ⁶University of California at San Diego; and ⁷Argonne National Laboratory

Large-scale sequencing of entire genomes represents a new age in biology, but the greatest challenge is to define cellular responses, gene functions, and regulatory networks at the whole-genome/proteome level. The key goal of this project is to explore whole-genome sequence information for understanding the genetic structure, function, regulatory networks, and mechanisms of anaerobic energy metabolism in the metal-reducing bacterium *Shewanella oneidensis* MR-1. To define the repertoire of MR-1 genes responding to different terminal electron acceptors, transcriptome profiles were examined in batch cultures grown with fumarate, nitrate, thiosulfate, DMSO, TMAO, ferric citrate, ferric oxide, manganese dioxide, colloidal manganese, and cobalt using DNA microarrays covering ~99% of the total predicted protein-encoding open reading frames in *S. oneidensis*. Total RNA was isolated from cells exposed to different electron acceptors for 3.5 h under anoxic conditions and compared to RNA extracted from cells under fumarate-reducing conditions. Microarray analy-

ses revealed significant differences in global expression patterns in response to different anaerobic respiratory conditions. The data indicated a number of genes that displayed preferential induction in response to specific terminal electron acceptors. This work represents an important step towards the goal of characterizing the anaerobic respiratory system of *S. oneidensis* MR-1 on a genomic scale.

To understand the molecular basis of anaerobic energy metabolism in MR-1, more than 20 genes with putative functions in global gene regulation, energy metabolism and adaptive cellular responses to stress were inactivated by deletion mutagenesis. Three double mutants defective in two global regulatory genes were also obtained. Genetic, biochemical and physiological characterization of these mutants are currently underway. Also, a random *Shewanella* phage display library is being constructed, and this library will contain 10 million unique inserts with insert sizes ranging from 300 bp to about 1.2 kb. Thus, the library should have a fusion point for approximately every base pair in the genome. In addition, the conditions for cloning individual open reading frames from *Shewanella* were optimized, and now the design and synthesis of primers for amplifying individual genes are underway.

B25

Global Regulation in the Methanogenic Archaeon *Methanococcus maripaludis*

John Leigh¹ (leighj@u.washington.edu), Murray Hackett¹, Roger Bumgarner¹, Ram Samudrala¹, **William Whitman**², Jon Amster², and Dieter Söll³

¹University of Washington; ²University of Georgia; and ³Yale University

Methanococcus maripaludis is a model hydrogenotrophic methanogenic archaeon. Growth on hydrogen and carbon dioxide results in the production of methane as a waste product. *M. maripaludis* stands out among methanogenic archaea as an ideal model species because of fast reproducible growth, a genome sequence, and effective genetic tools. Genetic manipulations in *M. maripaludis* are increasingly facile. In our collaboration under the Department of Energy's Microbial Cell Program we are studying global regulation by hydrogen, amino acid starvation,

growth rate, and other conditions. Little is known of these regulatory systems in Archaea, especially at the global level. Our approaches include continuous culture of *M. maripaludis*, expression arrays, proteomics, measurement of metabolite levels, determination of tRNA charging, and genetic manipulation.

To date our team has installed dual chemostats running with anaerobic gas sources and has established reproducible growth conditions over a range of growth rates. We are in the process of calibrating the continuous culture system for growth under various nutrient limitations including hydrogen. Besides having well controlled steady-state conditions, a particular virtue of the continuous culture approach is that it will allow us to distinguish the specific effects of nutrient limitation from the general effects of growth rate. As work preliminary to our global analysis of hydrogen regulation we have produced *lacZ* fusions to *mtd* (encoding the fourth step in methanogenesis) and two *filh* genes (encoding formate dehydrogenases) and have demonstrated differential regulation under high- and low hydrogen regimes. In preparation for our study of the response to amino acid starvation we have constructed two amino acid-auxotrophic mutants, including a mutant in the biosynthesis of leucine (isopropyl malate synthase, *leuA*) and a mutant in the common pathway of aromatic amino acids biosynthesis (3-dehydroquinate dehydratase, *aroD*). The genotypes of these mutants have been confirmed, and the mutants possess the expected auxotrophic phenotypes. We are also preparing to study other regulatory systems, and for this purpose we have constructed null mutations in genes encoding potential transcriptional regulatory proteins, including members of the AsnC, MerR and GntR families. For the analytical aspects of our global regulation studies we are implementing an approach to proteomics that in preliminary tests appears well suited to quantitative global analysis of the proteomes of prokaryotic organisms with relatively small genomes. In this approach, pools of proteolytic peptides are fractionated by several stages of liquid chromatography, analyzed by tandem mass spectrometry, and computationally matched to open reading frames in the genome. For expression studies at the mRNA level we have purchased a DNA array for *M. maripaludis* and are in the process of spotting glass slides. The expression array center at the University of Washington is continuing to develop data analysis tools that will facilitate our manipulation and integration of expression array data, proteomic data, and annotation information.

B27

Identification of Regions of Lateral Gene Transfer Across the Thermotogales

Karen E. Nelson (kenelson@tigr.org), Emmanuel Mongodin, Ioana Hance, and Steven R. Gill

The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, Maryland, Telephone: 301-838-3565, Fax: 301-838-0208

The genome of *Thermotoga maritima* MSB8 was completely sequenced in 1999. Whole genome analysis of this bacterium suggested that 24% of the DNA sequence was most similar to that of archaeal species, primarily to *Pyrococcus* sp. Many of these open reading frames (ORFs) that were archaeal-like were clustered together in large contiguous pieces that stretched from 4 to 21 kb in size, were of atypical composition when compared to the rest of the genome, and shared gene order with the archaeal species that they were most similar to. The analysis of the genome suggested that this organism had undergone extensive lateral gene transfer (LGT) with archaeal species. Independent biochemical analyses by Doolittle and workers using degenerative PCR and subtractive hybridization techniques have also revealed gene transfer and extensive genomic diversity across different strains of *Thermotoga*. Genes involved in sugar transport, polysaccharide degradation as well as subunits ATPases were found to be variable.

We have created a whole-genome microarray based on the completed genome sequence that has been used to do comparative genome hybridizations (CGH) with 12 different *Thermotoga* strains/species that include *Thermotoga* sp. RQ2, *Thermotoga neopolitana* NS-ET and *Thermotoga thermarum* LA3. PCR products representing the 1879 total *T. maritima* MSB8 ORFs have been spotted in duplicate on Corning UltraGap slides. Two flip-dye experiments have been conducted per strain. Genes were considered to be shared between the 2 compared strains if the ratio (MSB8/experimental strain) was between 1 and 3, and considered to be absent if the ratio was greater than 10. Analysis of the resulting data demonstrates that there is a high level of variability in the presence and absence of genes across the different *Thermotoga* strains/species.

Of the strains that have been compared to the sequenced MSB8, RQ2, PB1platt and S1-L12B

share the highest level of genome conservation with MSB8. Only 129 ORFs in the MSB8 genome (1866 ORFs in total) did not have homologues in the RQ2 genome. These include 45 hypothetical proteins and 13 conserved hypothetical proteins, as well as 23 (18% of total that are absent) that are involved in transport. Of these 129, 18 occur as single ORFs, and the remaining correspond to islands that range in size from 2 kb to 38 kb. For strain S1-L12B, 9.4 % of the ORFs in MSB8 do not have homologs in this genome. Of these 174, 48 occur as single ORFs, and there are a total of 22 islands larger than 2 kb that are absent. Sixty-six ORFs correspond to hypothetical proteins, and 29 ORFs correspond to conserved hypothetical proteins. In addition, 6.9% are devoted to transport. Ten percent (186) of the MSB8 ORFs do not have homologs in PB1 (55 hypothetical proteins, 33 conserved hypotheticals), 16% of which are involved in transport. There are a total of 18 islands greater than 2kb in size that are absent from this strain. *T. thermarum* LA3 appears to be the most distantly related to MSB8. Initial data analysis suggests that lateral gene transfer across hyperthermophiles may be mediated by repetitive sequences that can be found in all these species. Interestingly, there is a high percentage of genes that are shared between *T. maritima* MSB8 and *Thermotoga* strain PB1platt that was isolated from an oil field in Alaska.

We have also designed, ordered and received primers to create a *Pyrococcus furiosus* genome microarray, and are in the process of diluting the primers and generating the PCR products that represent the entire genome. It is anticipated that we will conduct experiments similar to the *Thermotoga* comparative genome hybridization study.

B29

The Dynamics of Cellular Stress Responses in *Deinococcus radiodurans*

Michael J. Daly¹, Jizhong Zhou², James K. Fredrickson³, Richard D. Smith³, Mary S. Lipton³, and Eugene Koonin⁴

¹Uniformed Services, University of the Health Sciences, 4301 Jones Bridge Road, Bethesda, MD 20814; Tel: 301-295-3750; ²Oak Ridge National Laboratory, Oak Ridge, TN; ³Pacific Northwest National Laboratory, Richland, WA; and ⁴National Center for Biotechnology Information, NIH, Bethesda, MD

Deinococcus radiodurans (DEIRA) is the most characterized member of the radiation resistant bacterial family *Deinococcaceae*. It is non-pathogenic, amenable to genetic engineering, and historically best known for its extreme resistance to gamma radiation [1]. The bacterium can grow and functionally express cloned foreign genes in the presence of 60 Gy/hour, and can survive acute exposures that exceed 15,000 Gy without lethality [1, 2]. How this feat is accomplished is unknown, and a long-term goal of our GTL project is a detailed understanding of the molecular pathways underlying this phenotype. Based on its remarkable robustness, DEIRA is also being developed for bioremediation of radioactive mixed waste sites containing radionuclides, heavy metals, and toxic organic compounds [2]. Using a combination of computational and whole-cell technologies, we are analyzing expression networks in DEIRA to map its cellular repair pathways, and also using information gleaned from this work to facilitate its development for bioremediation.

Whole genome sequencing, annotation, and comparative analyses for DEIRA [1, 3] have given rise to the development of new experimental whole-cell technologies dedicated to this organism. In collaboration with co-investigators at PNNL, our groups have developed proteomic methodology that uses high-resolution liquid chromatography and Fourier transform ion cyclotron resonance mass spectrometry to characterize an organism's dynamic proteome. Using this technology, >61% of the predicted proteome of DEIRA has been characterized with high confidence [4]. This represents the broadest proteome coverage for any organism to date. And, in collaboration with co-investigators at ORNL and NCBI, we have constructed a whole-genome microarray (WGM) for DEIRA and used it to

examine global RNA expression dynamics during recovery from high-dose irradiation [5].

Already, proteomic and WGM research has revealed an unprecedented view of the molecular systems involved in the resistance phenotypes of DEIRA, and has also facilitated the construction of DEIRA strains capable of detoxifying highly radioactive waste environments. For example, with respect to novel genes that may be involved in radiation resistance, we have confirmed the involvement of several that show marked induction following irradiation [5]. This work has also alerted us to how metabolic strategies, not generally associated with DNA protection or repair, could enhance its resistance functions. DEIRA switches its metabolic pathways in response to irradiation, minimizing oxidative stress production and optimizing recovery [6]. We now believe that a comprehensive knowledge of DEIRA metabolism is key to advancing our understanding of its extreme radiation resistance, as well as extending its intrinsic metabolic functions for bioremediation. In the past, our goal of engineering DEIRA for complete mineralization of toluene could not be reached because of uncertainties regarding engineering strategies relating to its metabolic configuration. These uncertainties were overcome following proteomic and WGM analyses that untangled some of the complexities of DEIRA metabolism, and revealed how DEIRA intermediary metabolism could be integrated with complete toluene oxidation. As a result, we have successfully engineered DEIRA strains that can completely mineralize toluene, as demonstrated in natural sediment and groundwater analogs of DOE contaminated environments [7]. In summary, the GTL program is providing us a unique opportunity to bring together state-of-the-science high-throughput whole-cell technologies to explore fundamental and applied aspects of *Deinococcus radiodurans*.

1. K. S. Makarova, *et al.* (2001) The genome of the extremely radiation resistant bacterium *Deinococcus radiodurans* viewed from the perspective of comparative genomics. *Microbiology and Molecular Biology Reviews* **65**, 44–79.
2. H. Brim, *et al.* (2000) Engineering *Deinococcus radiodurans* for metal remediation in radioactive mixed waste environments. *Nature Biotechnology* **18**, 85–90.
3. O. White, *et al.* (1999) Sequencing and functional analysis of the *Deinococcus radiodurans* genome. *Science* **286**, 1571–1577.
4. M. S. Lipton, *et al.* (2002) Global analysis of the *Deinococcus radiodurans* R1 proteome using accurate mass tags. *Proc. Natl. Acad. Sci. USA*, **99**, 11049–11054.

5. Y. Liu, J. Zhou, A. Beliaev, J. Stair, L. Wu, D.K. Thompson, D. Xu, A. Venkateswaran, M. Omelehenko, M. Zhai, E. K. Gaidamakova, K. S. Makarova, E. Koonin, and M. J. Daly (2002) Transcriptome dynamics of *Deinococcus radiodurans* recovering from ionizing radiation. Submitted.
6. A. Vasilenko, A. Venkateswaran, H. Brim, Y. Liu, J. Zhou, K. S. Makarova, M. Omelchenko, D. Ghosal, and M. J. Daly (2002) Relationship between metabolism, oxidative stress and radiation resistance in the family *Deinococcaceae*. Submitted.
7. H. Brim, J. P. Osborne, A. Venkateswaran, M. Zhai, J. K. Fredrickson, L. P. Wackett, and M. J. Daly (2002). Facilitated Cr(VI) Reduction by *Deinococcus radiodurans* Engineered for Complete Toluene Mineralization. Submitted.

B9

Uncovering the Regulatory Networks Associated with Ionizing Radiation-Induced Gene Expression in *D. radiodurans* R1

John R. Battista¹ (jbattis@lsu.edu), Ashlee M. Earl¹, Heather A. Howell², and Scott N. Peterson²

¹Department of Biological Sciences, Louisiana State University and A & M College, Baton Rouge, LA 70803; and ²The Institute for Genomic Research, Rockville, MD 20850

In an effort to determine which genes are LexA regulated in *D. radiodurans* a *lexA* defective strain of *D. radiodurans* R1, GY10912, was evaluated using microarray analysis. Under normal, unstressed conditions over 100 transcripts were more abundant in GY10912 than in the wild-type strain suggesting that a large fraction (3%) of *D. radiodurans* genome is regulated by this repressor. However, only 22 genes from the LexA controlled gene set overlapped with the 71 genes previously determined to be stress induced following exposure to ionizing radiation in wild type R1. There is absolutely no overlap between the 'classical' SOS regulon of *E. coli* and LexA controlled genes in *D. radiodurans*. When a 3,000Gy dose of ionizing radiation is administered to GY10912 only 7 additional genes are induced including *recA*. Since a LexA defect does not render *D. radiodurans* sensitive to ionizing radiation, it is assumed that the cell only needs to up-regulate these 29 loci: the 22 LexA dependent loci and the 7 LexA independent loci. The LexA independent induction is in part controlled by the IrrE protein (DR0167). IrrE is a positive effector that when inactivated results in loss of ionizing radiation

resistance. Loss of IrrE prevents the expression of 13 loci that are normally induced in response to ionizing radiation. All 13 of these loci overlap with those genes thought to confer resistance to GY10912.

B31

Analysis of Proteins Encoded on the *S. oneidensis* MR-1 Chromosome, Their Metabolic Associations, and Paralogous Relationships

Margrethe H. Serres*, Maria C. Murray, and **Monica Riley**

*Presenter

Marine Biological Laboratory, Woods Hole, MA 02543

Proteins encoded by the *Shewanella oneidensis* MR-1 chromosome have been analyzed for their sequence similarity to proteins encoded in 49 completely sequenced microbial genomes. It is our goal to elucidate the metabolic pathways in *S. oneidensis* in order better to understand the metabolic capabilities of this versatile organism. The genome is also being analyzed for paralogous or sequence similar proteins within the chromosome. Such protein families have been found useful in assigning putative functions to gene products and have provided insight into the evolution of the genome and the functions encoded within.

Sequence similarity searches were done using Darwin and an alignment requirement of at least 83 amino acids. Sequence matches were found for 82% of the encoded proteins at a similarity of ≤ 250 PAM units. *Vibrio cholerae*, *Pseudomonas aeruginosa*, and *Yersinia pestis* showed the highest percentage of best hits at 23%, 10%, and 7%, respectively.

To identify metabolic pathways we conferred with the GenProtEC and EcoCyc/BioCyc databases and with published literature. We used the MultiFun classification system consisting of the following classes; metabolism, information transfer, regulation, transport, cell processes, cell structure, location, extra-chromosomal origin, DNA sites and cryptic genes. Currently more than 3800 cell function assignments have been made to 1560 *S. oneidensis* proteins. Among these, 740 proteins were assigned to metabolism, including 193 proteins classified as having a role in respiration. An

overview of pathways, missing steps and similarity to other genomes will be presented.

In order to determine paralogous protein families, proteins encoded by fused or composite genes have to be identified and their sequences separated into entities (modules) of independent evolutionary origin. The pair-wise alignments from the Darwin analysis described above have been used to discern modules. Currently 170 (4%) of the chromosomally encoded *S. oneidensis* proteins have been identified as composite. This number is slightly higher than for the *E. coli* genome. An initial grouping of protein modules is in progress and will be presented.

B33

Finishing and Analysis of the *Nostoc punctiforme* Genome

S. Malfatti¹ (malfatti3@llnl.gov), L. Vergez¹, N. Doggett², J. Longmire², R. Atlas³, J. Elhai⁴, J. Meeks⁵, and **P. Chain**¹

¹Biology and Biotechnology Research Program, Lawrence Livermore National Laboratory, Livermore, CA; ²Bioscience Division, Los Alamos National Laboratory, Los Alamos, NM; ³Department of Biology, University of Louisville, Louisville, KY; ⁴Department of Biology, University of Missouri, St. Louis, MO; and ⁵Section of Microbiology, University of California, Davis, CA

In an effort to explore the role of diverse microorganisms in global carbon sequestration, the DOE's JGI has drafted the genomic sequence of several microorganisms that play unique roles in soil and ocean ecosystems. Elucidation of their genetic content will allow de novo identification of metabolic pathways relevant to understanding the physiological and genetic controls of photosynthesis, nitrogen fixation and general carbon cycling. A major contributor to the sequestration of CO₂ in organic compounds, *Nostoc punctiforme* is a unique nitrogen-fixing cyanobacterium that can differentiate into three types of specialized cells (N₂-fixing, motile and spore-like), is capable of forming symbiotic associations with fungi and plants and has the ability to grow rapidly under completely dark heterotrophic conditions. We have undertaken the task of finishing and analyzing the genome of *N. punctiforme* strain ATCC 29133, which is likely to yield novel information on global regulation of multiple developmental pathways, symbiotic association, and phylogenetic relation to the other cyanobacteria including the

now complete *Prochlorococcus* and *Synechococcus* strains.

The genome of *N. punctiforme* is very large for a prokaryote, nearly 10Mb, which makes this an attractive genome to complete. There are currently near 100 prokaryotic genomes that have been sequenced and closed, yet the average finished genome size is only ~3Mb and fewer than 10 have a genome size larger than 6Mb. Thus, there is quite possibly a bias in our prokaryotic genome knowledge toward smaller genomes and there may be some distinctive features in larger prokaryotic genomes that require finer resolution.

The genome was drafted by the JGI to 10-fold coverage due to the large number of contigs observed at 5-, 6- and 7-fold; however, the number of contigs greater than 2kb with >10 reads remained essentially the same, near 230 contigs. Additional rounds of automated primer design for finishing purposes has not significantly reduced this number. The reason for the difficulties in finishing are likely due to the prevalence of repeated elements within the genome. This is supported by the high coverage (25- to 50-fold) of entire contigs of large size (up to 110 kb). The identification and resolution of these repeated elements is being performed with the use of a fosmid scaffold along with a suite of computational tools. An overview of the progress in finishing *N. punctiforme* will be presented.

This work was performed under the auspices of the U. S. Department of Energy by the University of California, Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48.

B35

In Search of Diversity: Understanding How Post-Genomic Diversity is Introduced to the Proteome

Barry Moore, Chad Nelson, Norma Wills, John Atkins, and **Raymond Gesteland** (ray.gesteland@genetics.utah.edu)

Department of Human Genetics, University of Utah, Salt Lake City, UT

Analysis of an organism's genome is the primary tool for understanding the diversity of protein products present in an organism. However, as proteomics focuses more work on the proteins themselves, it is apparent that the diversity of the

proteome goes well beyond that indicated in the genome. Much of the energy and resources in proteome projects have been invested in various peptide based mapping strategies—principally by 2-D gel electrophoresis followed by MALDI-MS analysis. One shortcoming of peptide approaches, however, is that they do not allow for determination of the full-length protein mass. This means that while it can be determined that a particular region of the genome (the ORF) is translated into protein, no direct evidence is provided about the full-length protein product, and there is no indication of the diversity of different proteins products translated from that ORF.

A number of studies have indicated that regions of sequenced genomes not currently identified as ORFs are translated, or are likely to be translated. Furthermore, numerous mechanisms are known by which diversity is introduced to an organism's proteome post-genomically, and there are a wide variety of bioinformatics algorithms either available or in development that attempt to predict these events from genome sequence. Ultimately, we must go to the proteins themselves to fully understand when and where post genomic diversity is arising and to refine our ability to predict from genome analyses the diversity of proteins derived from an organism's genome. We are developing a dual affinity tagging system in the yeast *Saccharomyces cerevisiae* that allows us to build fusions to regions of the genome where novel ORFs are predicted, or where non-standard translational events are expected to occur. Fusion proteins are expressed in yeast and purified by affinity chromatography to produce a protein pure enough for analysis by electrospray mass spectrometry. Analysis of a highly accurate full length mass in combination with tryptic peptide fragments and MS/MS analysis of those fragments allows a thorough understanding of the proteins full length covalent state, and the diversity of products produced from a particular ORF.

B37

The Microbial Proteome Project: A Database of Microbial Protein Expression in the Context of Genome Analysis

Carol S. Giometti¹ (csgiometti@anl.gov), Gyorgy Babnigg¹, Sandra L. Tollaksen¹, Tripti Khare¹, Derek R. Lovley², James K. Fredrickson³, Kenneth H. Nealson⁴, Claudia I. Reich⁵, Gary J. Olsen⁵, Michael W. W. Adams⁶, and John R. Yates III⁷

¹Argonne National Laboratory; ²University of Massachusetts; ³Pacific Northwest National Laboratory; ⁴University of Southern California; ⁵University of Illinois at Urbana; ⁶University of Georgia; and ⁷The Scripps Research Institute

Although complete genome sequences can be used to predict the proteins expressed by a cell, such predictions do not provide an accurate assessment of the relative abundance of proteins under different environmental conditions. In addition, genome sequences do not define the subcellular location, biomolecular and cofactor interactions, or covalent modifications of proteins that are critical to their function. Analysis of the protein components actually produced by cells (i.e., the proteome) in the context of genome sequence is, therefore, essential to understanding the regulation of protein expression. As the number of complete microbial genome sequences increases, vast amounts of genome and proteome information are being generated.

In parallel with the proteome analysis of numerous microbial systems, researchers at Argonne National Laboratory are developing methods for managing and interfacing the diverse data types generated by both genome and proteome studies as part of Argonne's Microbial Proteome Project. The goal is to provide users with a highly interactive database that contains proteome information in the context of genome sequence in formats that are conducive to data interrogations that will provide answers to biological questions. To achieve that goal, three World Wide Web-based databases are currently being developed and maintained: Proteomes2, ProteomeWeb, and GelBank.

The Proteomes2 database (proteomes2.bio.anl.gov) serves as a Laboratory Information Management System for the management of sample data and related two-dimensional gel electrophoresis (2DE) patterns. This password-protected site provides DOE project collaborators with access to

data access from multiple sites through the Internet. The database currently contains the experimental details for approximately 1000 samples from seven different microbes (*Shewanella oneidensis*, *Geobacter sulfurreducens*, *Prochlorococcus marinus*, *Methanococcus jannaschii*, *Pyrococcus furiosus*, *Rhodospseudomonas palustris*, and *Deinococcus radiodurans*) and links each sample with multiple protein patterns. Over 4000 protein pattern images are currently accessible.

ProteomeWeb (<http://ProteomeWeb.anl.gov>) is an interactive public site that provides the identification of expressed microbial proteins, links to genome sequence information, tools for mining the proteome data, and links to metabolic pathways. Data from proteome analysis experiments are included in the ProteomeWeb database when genome sequences are deposited in GenBank. Currently, the results from experiments designed to alter protein expression in *M. jannaschii* are accessible on this site, and results from *S. oneidensis* experiments are in the process of being incorporated.

GelBank currently includes the complete genome sequences of approximately 90 microbes and is designed to allow queries of proteome information. The database is currently populated with protein expression patterns from the Argonne Microbial Proteomics studies and will accept data input from outside users interested in sharing and comparing proteome experimental results.

Oracle9i RDBMS — the common foundation for all three Argonne proteomics databases — allows the integration of genome sequences, sample descriptors, protein expression data (e.g., 2DE images and numerical data), peptide masses, and annotation. The database management architecture being developed currently addresses protein expression data from 2DE analysis of protein mixtures, but it is being designed to have the flexibility to accommodate mass spectrometry data as well.

This research is funded by the United States Department of Energy, Office of Biological and Environmental Research, under Contract No. W-31-109-ENG-38.

B39

Analysis of the *Shewanella oneidensis* Proteome in Cells Grown in Continuous Culture

Carol S. Giometti¹ (csgiometti@anl.gov), **Mary S. Lipton**² (mary.lipton@pnl.gov), Gyorgy Babnigg¹, Sandra L. Tollaksen¹, Tripti Khare¹, James K. Fredrickson², Richard D. Smith², Yuri A. Gorby², and John R. Yates III³

¹Argonne National Laboratory; ²Pacific Northwest National Laboratory; and ³The Scripps Institute

We are using two complementary methods to identify and quantify the proteins expressed by *Shewanella oneidensis* MR-1 grown under different conditions in continuous culture. Cells are grown in chemostats under aerobic, O₂-limited, or anaerobic conditions with fumarate provided as the electron acceptor. Cells are harvested and aliquots of the same samples are analyzed for protein by using (1) two-dimensional gel electrophoresis (2DE) coupled with peptide mass analysis and (2) capillary liquid chromatography separations coupled with Fourier transform ion cyclotron resonance (FTICR) mass spectrometry. The 2DE patterns readily reveal statistically significant differences in protein abundance and in protein post-translational modifications that result from different growth conditions. The proteins that differ significantly in expression are then identified on the basis of their tryptic peptide masses. In parallel with the 2DE analysis, we are using accurate mass tags (AMTs) to identify all of the *S. oneidensis* proteins expressed under specific growth conditions and to detect quantitative differences by using stable isotope labeling methods. By using these two different protein separation and detection methods, we are able to more comprehensively analyze the proteome than by using either method alone. Specifically, 2DE provides a rapid turnaround “snap shot” of the major protein differences (including post-translational modifications) under different growth conditions, and AMTs provide a comprehensive inventory of the expressed proteins in each sample.

This work is funded by the U.S. Department of Energy, Office of Biological and Environmental Research, under contract No. W31-109-ENG-38 (Argonne National Laboratory) and contract No. DE-AC06-76RLO1830 (Pacific Northwest National Laboratory).

B41

The Molecular Basis for Metabolic and Energetic Diversity

Timothy Donohue¹ (tdonohue@bact.wisc.edu), Jeremy Edwards², Mark Gomelsky³, Jonathan Hosler⁴, Samuel Kaplan⁵, and William Margolin⁵

¹Bacteriology Department, University of Wisconsin-Madison; ²Chemical Engineering Department, University of Delaware; ³Department of Molecular Biology, University of Wyoming; ⁴Department of Biochemistry, University of Mississippi Medical Center; and ⁵Department of Microbiology & Medical Genetics, University of Texas Medical School at Houston

Our long-term goal is to engineer microbial cells with enhanced metabolic capabilities. As a first step, this team of scientists and engineers seeks to acquire a thorough understanding of energy-generating processes and genetic regulatory networks of the photosynthetic bacterium, *Rhodobacter sphaeroides*. The ability to capitalize on the metabolic activities of this versatile bacterium was increased by the completion of the *R. sphaeroides* genome sequence at the DOE-supported Joint Genome Institute. The *R. sphaeroides* Genomes to Life Consortium is deciphering important energy-generating activities of this bacterium and studying the assembly and operation of energy generating machines. The long term goals of these efforts are to acquire the information needed to design microbial machines that degrade toxic compounds, remove greenhouse gases, or synthesize biodegradable polymers with increased efficiency. At the February, 2003 workshop, we will provide a progress report on our analysis of the metabolic capabilities of this facultative microorganism, particularly on the identification of proteins and regulators that are central to growth via respiration and the utilization of solar energy by photosynthesis.

In one line of experiments, we have begun to characterize the multitude of aerobic respiratory enzymes that this bacterium uses to generate energy in the presence of O₂. The goal of these experiments is to engineer strains that can more efficiently extract energy from nutrients in the presence of O₂.

We have begun to visualize the assembly of the photosynthetic apparatus that this bacterium uses to harvest solar energy. When this analysis is combined with modeling of solar energy utilization by this well-studied microbe, we hope to identify

principles that will allow us to engineer microbes with increased ability to generate energy from light.

The feasibility of generating strains with increased capacity for using solar radiation is enhanced by our identification of previously unrecognized pigment-binding proteins in the photosynthetic apparatus of *R. sphaeroides*. Efforts are currently underway to understand the contribution of these new pigment-binding proteins to the utilization of light energy that is critical for growth of photosynthetic bacteria and plants.

Simultaneously we have initiated a program to identify new regulators of the solar lifestyle of this bacterium and to use gene chips to identify additional proteins that could be critical to the utilization of light energy. Both of these approaches have provided new candidates that are currently being analyzed for their role as regulators or contributors to the utilization of solar energy.

We hope to illustrate why this cross-disciplinary, systems approach to the analysis of energy generation by this facultative bacterium can provide new insights into fundamental aspects of energy generation and the utilization of solar energy by this and other photosynthetic organisms.

B43

Integrative Studies of Carbon Generation and Utilization in the Cyanobacterium *Synechocystis* sp. PCC 6803

Wim Vermaas¹ (wim@asu.edu), Robert Roberson¹, Julian Whitelegge², Kym Faull², Ross Overbeek³, and Svetlana Gerdes³

¹Arizona State University; ²University of California Los Angeles; and ³Integrated Genomics, Inc.

The research focuses on photosynthetic and respiratory electron transport and on carbon utilization in the cyanobacterium *Synechocystis* sp. PCC 6803. The research compares wild type with targeted mutants lack the photosystems, the terminal oxidases, succinate dehydrogenase, and/or other complexes that are important for photosynthesis and respiration.

Physiological Analysis

1. Carbon utilization: Mutants lacking terminal oxidases were found to accumulate large inclusions that resembled poly- β -hydroxybutyrate bodies. As these inclusions were by far most prevalent in mutants lacking terminal oxidases, they most likely result from fermentation reactions. The very significant accumulation of polyhydroxyalkanoates (up to about 25% of the total cell volume) in *Synechocystis* mutants lacking the terminal oxidases suggests a potential application of this strain in light-driven production of “bioplastic” materials.

2. Membrane biogenesis: One major question that remains in chloroplasts and cyanobacteria is how thylakoids are formed. A gene that apparently is involved with thylakoid biogenesis has been interrupted at various positions, and the interruption that has segregated (with the site of insertion close to the end of the gene) showed a normal ultrastructural phenotype. We are also inactivating genes involved with cell division, in the hope of creating a larger cyanobacterial cell that is more easily analyzed by light microscopy techniques.

3. Toward a system suitable for crystal structure analysis: *Synechocystis* has proven to be an excellent model system from a standpoint of molecular genetics, physiology, etc. However, for elucidation of structures of membrane proteins (another goal that should be pursued in microbial cell projects) thermophilic cyanobacteria, such as *Thermosynechococcus elongatus*, for which a genome sequence is known, are very much preferable. For full utility of this system, mutants that possibly impair the photosystems should be used and therefore the organism should be able to grow photoheterotrophically, which it currently is unable to do. Therefore, we have initiated random mutagenesis experiments to see whether a *Thermosynechococcus* mutant can be obtained that can grow under conditions other than photoautotrophic ones.

Structural Analysis

Much progress has been made in ultrastructural preservation and documentation of cytoplasmic organization in *Synechocystis* sp. PCC 6803.

1. Three-dimensional reconstruction: Using thick (0.2-0.3 μm) sections of *Synechocystis*, we have collected data on high- and intermediate-voltage transmission electron microscopes (0.2-1 MeV) to three-dimensionally image the intracellular

organization of *Synechocystis* sp. PCC 6803 cells. In this procedure, thick sections are incrementally tilted from -60° to $+60^\circ$ and serial tilt views electronically captured at 1.5° intervals. We have produced the first high-resolution three-dimensional images of a cyanobacterium, and are now in the process to identify physical relationships between thylakoid and cytoplasmic membranes, the fate of membranes upon cell division, etc. Clearly, tomography using thick sections is an extremely powerful technique that is providing new insights into the ultrastructural organization of the cell.

2. Membrane organization: We observe “thylakoid centers” in dividing cells. These structures were discovered about two decades ago in another cyanobacterium, and apparently have been forgotten since. Thylakoid centers are located at the point of apparent origin of thylakoid membranes (i.e., the point of thylakoid membrane convergence) near the cytoplasmic membrane. At these specialized regions, membranes either terminate or turn 180° in close proximity to the cytoplasmic membrane. The thylakoid center apparently extends as a tube-like structure for at least 200 nm. We are interested in exploring the role and composition of these thylakoid centers.

Proteomics

With the complete genome sequence available, comparative proteome analysis is underway with the goal of providing data that will integrate with the ultrastructural work and in mutants of photosynthesis and respiration. By using sub-fractionation of soluble or membrane preparations it is possible to preserve native protein complexes providing functional insights. Soluble fractions are examined by size-exclusion chromatography and 2D-gel electrophoresis for evidence of the carboxysome seen in EM work to be more abundant in the SDH-less mutant. Membrane preparations are sub-fractionated by sucrose-gradient centrifugation to separate thylakoids from cytoplasmic membranes, and to search for preparations that may be enriched in thylakoid centers, in order to better understand the specialization of each membrane system and the trafficking between them. Membrane protein complexes are being separated by size-exclusion chromatography under non-denaturing conditions after solubilization of membranes with detergents as the first dimension of a 2D chromatography system. A second dimension denaturing chromatography system incorporates intact protein mass measurement (LC-MS+). Fractions collected during LC-MS+ are used for protein identification

by fingerprinting and sequencing and are also available for further analysis by ultra-high resolution Fourier transform mass spectrometry (FT-MS). The successful analysis of intact photosystem I reaction-center polypeptides PsaB and PsaA of mass 81,167 and 82,876 Da, respectively, demonstrates that *intact mass proteomics* can be applied to large integral membrane proteins with resolution sufficient to detect single methionine oxidations at this size.

Bioinformatics

Integrated Genomics began work on the GTL Project in October 2002; hence progress in this area has been somewhat limited (but not insignificant) so far. The initial objectives are:

1. To support the wet lab characterization of critical genes by making specific predictions that will be tested experimentally by our partners.
2. To produce a detailed metabolic reconstruction to support a more complete understanding and modeling of this organism.

Integrated Genomics is coordinating closely with other members of the group to establish priorities relating to prediction of gene function. During

the first month, we have made two predictions, one addressing the lycopene cyclase, and one related to a gene associated with phycobilisomes. The lycopene cyclase is by far the more interesting of these predictions, and other members of the group have begun experiments to confirm or reject our candidate gene. We have now begun a tabulation of predictions.

The second broad goal is to develop a detailed metabolic reconstruction for *Synechocystis sp.* PCC 6803. In particular, two aspects of this effort are pursued: (1) development of a detailed graphical/web-based interface to a reconstruction connecting functional components to genes in the organism, and (2) development of a precisely encoded version of the reaction network in a form suitable for supporting modeling efforts.

Altogether, the GTL project on *Synechocystis sp.* PCC 6803 provides integrated functional-genomics information regarding the structure of the organism and the function of photosynthesis and respiration related processes, based on physiological, ultrastructural, proteomics, and bioinformatics experimentation on the wild type and targeted mutants.