

Gene Expression Profiles of *Rhodopseudomonas palustris* Nitrogenases by Whole Genome Microarray

Y. Oda¹ (yasuhiro-oda@uiowa.edu), S. K. Samanta¹, L. Wu², X.-D. Liu², T.-F. Yan², J. Zhou², and C. S. Harwood¹

¹University of Iowa, Iowa City, IA and ²Oak Ridge National Laboratory, Oak Ridge, TN

Rhodopseudomonas palustris is a photosynthetic bacterium that can use many forms of carbon, nitrogen, and electron donors. Under anaerobic conditions it can generate energy from light and convert nitrogen gas to ammonia and hydrogen (a biofuel) by nitrogen fixation. A striking feature of the genome sequence of *R. palustris* CGA009 is genes encoding three different nitrogenases and the accessory proteins needed for nitrogenase assembly. *AnfHDGK*, *nifHDK*, and *vmfHDGK* genes encode iron (Fe)-containing, molybdenum (Mo)-containing, and vanadium (V)-containing nitrogenases. To address the question of how *R. palustris* differentially regulates nitrogenase gene expression, we constructed *anfHnifH*, *anfHvmfH*, and *nifHvmfH* double mutants and analyzed the whole genome gene expression profiles of each mutant and wild-type cells grown under nitrogen-fixing conditions. Each mutant expressed a single functional nitrogenase (Mo, V or Fe) in a minimal medium that contained molybdenum and other trace elements. Wild-type and the Mo-nitrogenase active mutant cells expressed over 150 genes at levels of 2-fold or higher when grown under nitrogen-fixing conditions as compared to when grown with ammonia. Among these were the 30 genes in the *nif* gene cluster, which were expressed at 5- to 200-fold higher (depending on the gene) levels in cells grown under nitrogen-fixing conditions. Genes in the *anf* and *vmf* clusters were not expressed. By contrast, cells with an active Fe-nitrogenase only or an active V-nitrogenase only expressed all of the *nif* genes (except *nifH* which was deleted), all of the *anf* genes, and all of the *vmf* genes. It makes sense that *nif* genes would be expressed because many of them are needed for the assembly of the Fe- and V-nitrogenases. These results indicate that *R. palustris* synthesizes both its Fe- and V-nitrogenases in situations where it is unable to synthesize an active Mo-nitrogenase. The mechanism by which Fe- and V-nitrogenase gene expression is activated is not known, but does not involve relief of Mo repression.

61

Harnessing the Integrative Control of C, N, H, S and Light Energy Metabolism in *Rhodospseudomonas palustris* to Enhance Carbon Sequestration and Biohydrogen Production

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

¹Ohio State University, Columbus, OH; ²University of Iowa, Iowa City, IA; ³Oak Ridge National Laboratory, Oak Ridge, TN; ⁴University of British Columbia, Vancouver, BC; and ⁵University of California, Los Angeles, CA

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 degradation and metal reduction operate in a single microbial cell. The recently sequenced *Rhodospseudomonas palustris* genome [Larimer et al., Complete genome sequence of the metabolically versatile photosynthetic bacterium *Rhodospseudomonas palustris*, Nature Biotechnology 22, 55-61, 2004] serves as the raw material for these studies since the metabolic versatility of this organism makes such studies both amenable and highly feasible. A multi-faceted approach has been taken for these studies. On the one hand, novel genes and regulators were identified from investigating control of specific processes by conventional molecular biology/biochemical techniques. In many instances, surprises relative to the role of known regulators, such as the Reg system and CbbR, were noted in *R. palustris*. In addition, a novel phospho-relay system for controlling CO₂ fixation gene expression was identified and biochemically characterized. This latter system, where key regulators contain motifs that potentially respond to diverse metabolic and environmental perturbations, suggests an exquisite means for controlling this key process. Likewise, interesting and important genes and proteins that control sulfur oxidation, nitrogen fixation, hydrogen oxidation, and photochemical energy generation were identified and characterized.

Our studies have shown that the control of CO₂ fixation is integrated and superimposed on the control of nitrogen fixation and hydrogen metabolism in this organism. By interfering with the normal means by which *R. palustris* removes excess reducing equivalents generated from the oxidation of organic carbon, strains were constructed in which much of the electron donor material required for growth was converted to hydrogen gas. The resultant strains were shown to be derepressed for hydrogen evolution such that copious quantities of H₂ gas were produced under conditions where the wild-type would not normally do this. As *R. palustris* and related organisms have long been proposed to be useful for generating large amounts of hydrogen gas in bio-reactor systems, the advent of these newly isolated strains, in which H₂ production is not subject to the normal control mechanisms that diminish the wild-type strain, is quite significant. Moreover, *R. palustris* is unique amongst the nonsulfur purple bacteria in that it is capable of degrading lignin monomers and other waste aromatic acids both anaerobically and aerobically. Inasmuch as the degradation of these compounds may be coupled to the generation of H₂ gas, by combining the properties of the H₂-producing derepressed strains, with waste organic carbon degradation, there is much potential to apply these basic molecular manipulations to practical advances. In addition, our results indicate that *R. palustris* has sophisticated nitrogen acquisition systems that are regulated somewhat differently than in other bacteria. These metabolic processes are driven by light energy, harvested by the photosynthetic apparatus. To maximize this capability, considerable molecular-based study is still

required and the combined expertise of all the investigators of this project, and related projects, is devoted toward this end.

To supplement the more traditional approaches taken above, we have also undertaken a combined microarray/proteomics/metabolomics and bioinformatics approach. Progress has been made towards developing an integrated network of control for the key metabolic processes under study. The whole genome microarrays have given us a global perspective of the changes in metabolic profile that occur under different physiological conditions. Included are metabolic transitions related to the carbon or nitrogen source supplied for growth, the energy source and the gaseous environment. Several genes were shown to be up and down regulated in these experiments, with several implicated in control. These studies have been supplemented by analysis of the proteome under the same growth conditions, using both whole cells and isolated intracytoplasmic membranes [For example, see Fejes et al., Shotgun proteomic analysis of a chromatophore-enriched preparation from the purple phototrophic bacterium *Rhodospseudomonas palustris*, *Photosyn. Res.* 78, 195–203, 2003]. The end result is that a suite of different, and in some cases, unexpected genes and proteins were identified that respond to specific physiological growth conditions. Moreover, several mutant strains, in which key aspects of metabolism have been altered (see above), were also analyzed by these genomics-based approaches. Beyond merely providing a list of genes and proteins, the transcriptome and proteome screens direct us towards the identification of novel regulators involved in integrating the control of the processes under study.

This multi-faceted approach 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 in the same cell.

62

Gene Expression Profiles of *Nitrosomonas europaea* During Active Growth, Starvation and Iron Limitation

Xueming Wei¹, Tingfen Yan², Norman Hommes¹, Crystal McAlvin², Luis Sayavedra-Soto¹, Jizhong Zhou², and **Daniel Arp**¹ (arpd@bcc.orst.edu)

¹Oregon State University, Corvallis, OR and ²Oak Ridge National Laboratory, Oak Ridge, TN

Ammonia-oxidizing *Nitrosomonas europaea* is a lithoautotrophic bacterium that converts NH_3 to NO_2^- by the successive action of ammonia monooxygenase (AMO) and hydroxylamine oxidoreductase (HAO): $\text{NH}_3 + \text{O}_2 + 2e^- \xrightarrow{\text{AMO}} \text{NH}_2\text{OH} + \text{H}_2\text{O}$ $\xrightarrow{\text{HAO}} \text{NO}_2^- + 5\text{H}^+ + 4e^-$. Two of the four electrons return to the AMO reaction and two either provide reductant for biosynthesis or pass to a terminal electron acceptor. The genome of *N. europaea* has been determined and consists of a single circular chromosome of 2,812,094 base pairs. Genes are distributed evenly around the genome, with ~47% transcribed from one strand and ~53% from the complementary strand. A total of 2460 protein-encoding genes emerged from the modeling effort, averaging 1011 bp in length, with intergenic regions averaging 117 bp.

We analyzed the gene expression profile of cells in exponential growth and during starvation using microarrays. During growth, 98% of the genes increased in expression at least two fold compared to starvation conditions. In growing cells, approxi-

mately 30% of the genes were expressed eight fold higher including genes encoding cytochrome c oxidase subunit I, cytochrome c, HAO, fatty acid desaturase and other energy harvesting genes. Approximately 10% were expressed more than 15 fold higher. Approximately 3% (91 genes) were expressed to more than 20 fold their levels in starved cells including the gene encoding multicopper oxidase type 1. Interestingly, the expression of the genes for AMO increased approximately two fold during growth. During starvation, the bulk of the genes were down-regulated with approximately 60% conserving low levels of expression compared to cells in exponential growth. Fewer than 2% of the genes were expressed more than two fold higher in starved cells. Genes expressed during starvation include those encoding NUDIX hydrolase, tyrosinase, multicopper oxidase, lipoxxygenase, cyclooxygenase-2, a putative transmembrane protein and other oxidative stress genes. Previously we had determined that starved cells transferred to normal medium responded with the induction of global gene expression. We have identified the genes involved in this global response and determined the extent of their expression. We have also identified the genes involved in the adaptation of *N. europaea* to starvation conditions.

Approximately 14% of the coding genes in *N. europaea* are dedicated to the transport of Fe and to siderophore receptors, yet *N. europaea* lacks genes for siderophore production (apparently relying on other bacteria to produce them). The growth of *N. europaea* is significantly affected by Fe. When actively growing in normal medium, the cells have a characteristic reddish color (probably due to the accumulation of cytochromes), but in an iron-limited medium, the cells grow poorly and have a lighter color. We carried out a preliminary study to quantify the effect of Fe limitation in the expression of Fe related genes using real-time PCR. Addition of Fe chelators to the Fe-limited growth medium inhibited growth completely for 5 days. *amoA*, *fecR* and *fluE* were expressed to higher levels in normal growth medium but not in iron-depleted medium. All other Fe-related genes showed no significant expression difference in these treatments. The microarray results showed that Fe-related genes were expressed to higher levels in growing cells than in starving cells. Genes encoding iron transport and binding as well as Fe superoxide dismutase were expressed 10 fold higher in growing cells than in starving cells. The siderophore desferal (produced by *Streptomyces* and other species) promoted the growth of *N. europaea* in iron-limited medium. The gene for the putative desferal receptor (a *foxA* homolog) was expressed to a higher level in iron-limited and desferal-containing cultures than in Fe-containing and desferal-free cultures. The expression of this gene apparently required desferal reinforcing the notion that *N. europaea* grows using the siderophores from other bacteria in Fe-limited environmental conditions. Here we have determined the possible receptor in *N. europaea* to a siderophore from another bacterium.

References

1. Chain, P., J. Lamerdin, F. Larimer, W. Regala, V. Lao, M. Land, L. Hauser, A. Hooper, M. Klotz, J. Norton, L. Sayavedra-Soto, D. Arciero, N. Hommes, M. Whittaker, and D. Arp. 2003. Complete genome sequence of the ammonia-oxidizing bacterium and obligate chemolithoautotroph *Nitrosomonas europaea*. *J Bacteriol* **185**:2759-2773.
2. Hooper, A. B., T. Vannelli, D. J. Bergmann, and D. M. Arciero. 1997. Enzymology of the oxidation of ammonia to nitrite by bacteria. *Antonie van Leeuwenhoek* **71**:59-67.
3. Sayavedra-Soto, L. A., N. G. Hommes, S. A. Russell, and D. J. Arp. 1996. Induction of ammonia monooxygenase and hydroxylamine oxidoreductase mRNAs by ammonium in *Nitrosomonas europaea*. *Mol Microbiol* **20**:541-548.
4. Wood, P. M. 1986. Nitrification as a bacterial energy source, p. 39-62. In J. I. Prosser (ed.), *Nitrification*. Society for General Microbiology, IRL Press, Oxford.

63

Photosynthesis Genes in *Prochlorococcus* Cyanophage

Debbie Lindell¹, Matthew B. Sullivan^{*2}, Zackary I. Johnson¹, Andrew C. Tolonen², Forest Rohwer³, and Sallie W. Chisholm^{1,4}

*Presenting author

¹Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA; ²Joint Program in Biological Oceanography, Woods Hole Oceanographic Institution and Massachusetts Institute of Technology, Cambridge, MA; ³Department of Biology, San Diego State University, San Diego, CA; and ⁴Department of Biology, Massachusetts Institute of Technology, Cambridge, MA

Our understanding of the mechanisms of phage-host interactions and their influence on the evolution of both phage and host, is based on a limited set of microorganisms representing an even more limited spectrum of metabolic types. Here we report the presence of genes central to oxygenic photosynthesis in the genomes of three cyanophage from 2 families of double-stranded DNA viruses (*Myoviridae*, *Podoviridae*) that infect the globally abundant marine cyanobacterium, *Prochlorococcus*. The photosystem II (PSII) core reaction center gene, *psbA*, and one high light inducible (*hli*) gene type were present in all 3 of the cyanophage genomes. The two myoviruses contain other photosynthesis related genes: One contains the second PSII core reaction center gene, *psbD*, while the other contains two photosynthetic electron transport genes coding for plastocyanin (*petE*) and ferredoxin (*petF*), and both contain additional *hli* gene types. All of these uninterrupted, full-length genes are conserved in their amino acid sequence with many fewer non-synonymous than synonymous nucleotide substitutions suggesting they encode functional proteins. Phylogenetic analyses indicate that the phage *psbA*, *psbD* and *hli* genes are of cyanobacterial origin, clustering with the corresponding genes from *Prochlorococcus*. They further suggest that these photosynthetic genes were transferred from host to phage multiple times. The phage *hli* genes cluster with sporadically distributed, multicopy *hli* types found exclusively in *Prochlorococcus*, suggesting that phage may be mediating the expansion of the *hli* gene family through the transfer of these genes back to their hosts after a period of evolution in the phage. Such reciprocal evolutionary effects of phage and *Prochlorococcus* on each others photosynthetic gene complement are likely to have significant implications for the success of host and phage in the surface oceans.

64

Metabolomic Functional Analysis of Bacterial Genomes

Pat J. Unkefer¹, Rodolfo A. Martinez¹, **Clifford J. Unkefer¹** (cju@lanl.gov), and Daniel J. Arp²

¹Bioscience Division, Los Alamos National Laboratory, Los Alamos, NM and ²Botany and Plant Pathology, Oregon State University, Corvallis, OR

Parallel with the terms genome, transcriptome and proteome, the combined profile of cellular metabolites is the metabolome. Examining changes in the metabolome is a potentially powerful approach to assessing gene function and contribution to phenotype. Achieving the GTL goal of obtaining a complete understanding of cellular function will require an integrated experimental and computational analysis of genome, transcriptome, proteome as well as the metabolome. Moreover, metabolites and their concentrations are a product of cellular regulatory processes, and thus the metabolome provides a clear window into the functioning of the genome and proteome. The profile of metabolites also reflects the response of biological systems to genetic or environmental changes. In addition, metabolites are the effectors that regulate gene expression and enzyme activity. *The focus of this project (Starting February 2004) is the elucidation of gene function by analysis of the metabolome.* We will carry out functional studies using stable isotope labeling and Mass or NMR spectral analysis of low-molecular weight metabolites. Like the proteome, metabolic flux and metabolite concentrations change with the physiological state of the cell. Because metabolite flux and concentration are correlated with the physiological state, they can be used to probe regulatory networks. In prokaryotic organisms, the combination of functional information derived from metabolic flux analysis with gene and protein expression data being developed in other laboratories will provide a powerful approach in identifying gene function and regulatory networks. Our pilot studies will build upon our capability, demonstrate the scientific value, and establish a facility for isotope-enhanced high throughput metabolome analysis of sequenced environmental microbes. Initially, we will study an ammonia-oxidizing chemolithotroph (*Nitrosomonas europaea*). Both play central roles in the global cycles of nitrogen and carbon.

The power of metabolome analysis will be greatly enhanced by applying the combination of stable isotope labeling and mutations. Stable Isotope labeling and NMR/Mass spectral analysis of metabolites will be used to assign metabolic function in three ways. First, we will apply specifically labeled compounds to establish precursor product relationships, and test if putative pathways identified from analysis of the genome are operational. Next, we will develop the capability for functional genomic analysis using comparative metabolomics to reveal the phenotype of a set of so-called silent mutations. This method combines null mutants constructed from the genome sequence by allelic exchange with metabolomic analysis to elucidate the function of unknown ORFs. Finally, we will carry out a full metabolic flux analysis in steady state cultures. Flux analysis will provide input for a stoichiometric model. Many of the advantages of isotope labeling for metabolomics in autotrophs and methylotrophs will be demonstrated throughout this proposal. Once demonstrated, this capability will be even more powerfully applied to heterotrophic organisms growing on complex substrates. These studies will lay the foundation to take similar labeling and metabolomic strategies into the environment to study microbial communities.

65

Genomics of *T. fusca* Plant Cell Wall Degradation

David B. Wilson (dbw3@cornell.edu), Shaolin Chen, and Jeong H. Kim

Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY

This year we have produced a small *Thermobifida fusca* DNA array containing sequences from potential genes involved in plant cell wall degradation and have cloned and characterized a family 10 *T. fusca* xylanase that is induced by growth on xylan but not by growth on cellulose.

The goal of the array study is to identify genes/proteins differentially expressed in *T. fusca* grown on biomass substrates. We will first characterize the transcript profile of *T. fusca* in cellobiose, crystalline cellulose, and xylan cultures using a DNA microarray on a glass slide. Glucose- grown cultures are used as references.

Preparation of DNA Array. Two different *T. fusca* microarray slides are used in this study. One slide contains 123 genes that may be related to biomass degradation as well as 10 “housekeeping” genes as a control (this will be called the 123 gene-slide). The other slide has most of the putative *T. fusca* genes identified in its genome sequence (this slide is called genome-slide, hereafter). To select the 123 genes, the *T. fusca* database was searched for putative cellulases and hemicellulases, putative genes that may contain cellulose or chitin binding domains, putative CelR regulated genes, putative secreted proteases/inhibitors, and putative membrane proteins that may be involved in biomass-related signal transduction.

RNA Purification. *T. fusca* was grown in Hagerdahl minimal medium containing glucose, cellobiose, xylan-birchwood, or Solka Floc cellulose. The RNA protect Bacteria reagent from Qiagen was used to stabilize RNA in cells after collecting. The phenol/chloroform extraction method (Kieser T. et al. Practical Streptomyces Genetics, pp. 253-363) was initially used for the purification of total RNA from *T. fusca* cells. This method uses buffers containing phenol and thus requires work in a ventilation hood, including the cell-breaking step. It is relatively time-consuming and requires repeated steps of phenol-chloroform extraction and isopropanol precipitation. The RNeasy Midi Kit also was used by following the manufacturer’s instruction. This procedure utilizes a centrifugation column to purify total RNA. However, the particulate material in the crude extract can cause contamination and results in impure RNA unable to give labeled cDNA. Therefore, a modified procedure was used, in which phenol/chloroform extraction was applied before column separation.

RNA Labeling. To label the total RNA samples, a protocol from the Institute for Genomic Research (<http://www.tigr.org/tdb/microarray/protocols/TIGR.shtml>) is used. This procedure labels RNA with aminoallyl labeled nucleotides via first strand cDNA synthesis followed by a coupling of the aminoallyl groups to either Cyanine 3 or 5 (Cy3/Cy5) fluorescent. This indirect aminoallyl labeling procedure yields more uniform labeling and incorporation of the two dyes used in this study.

Hybridization and Scanning. Again the protocol from the Institute for Genomic Research is used for hybridization. Gene Pix4000B is applied to scan arrays. Gene Pix Pro is used for initial data analysis.

Normalization. In order to compare expression levels on the 123 gene-slides, we need to identify a sufficient number of non-differentially expressed genes on each

slide. Therefore, 10 “housekeeping” genes are included. The “Rank-invariant Method” developed by Schadt, EE, et al (In: Feature extraction and normalization algorithm for high-density oligonucleotide gene expression array data. Preprints 303, Department of Statistics, UCLA, Los Angeles, CA) is applied to identify non-differentially expressed genes and to perform normalization. Briefly, the ranks of Cy3 and Cy5 intensities of each gene on the slide are calculated. For a given gene a threshold value d is used for the ranks of Cy3 and Cy5 intensities and a range of I value for the rank of the averaged intensity to determine if a gene is non-differentially expressed. A threshold value of 5 for both d and I will be used for the 123-gene slides. For the genome-slides, the larger number of genes allows us to use a more sophisticated iterative selection scheme as described by Schadt, EE et al. using a program provided by Tseng, GC et al. (*Nucleic Acid Research*, 2001, **29**: 2549-2557).

Other Approaches. In addition, we will use real-time RT PCR and 2-dimensional PAGE to perform further analysis on differentially-expressed genes or proteins.

Xyn10B an endoxylanase from *Thermobifida fusca* was overexpressed in *E. coli* and purified. Mature Xyn10B is a 43 kDa protein that produces xylobiose (SA 95microMol/min/mg) as the major product from birchwood xylan. It hydrolyzes p-nitrophenyl a-D-arabinopyranoside, p-nitrophenyl-b-D-xyloside, and p-nitrophenyl-b-D-cellobioside but at very low rates (<0.1microMol/min/mg). Xyn10B has moderate thermostability, retaining more than half of its xylanase activity after incubation at 55C for 15 hrs and is most active between pH 6-8. Unlike most *T. fusca* hydrolases it has a narrow pH activity profile. Xyn10B is induced by growth of *T. fusca* on xylan or Solka Floc but not on pure cellulose. It does not bind to cellulose, as it lacks a CBM and it appears to be a single domain enzyme.

66

Proteomic Analyses of a Hydrogen Metabolism Mutant of *Methanococcus maripaludis*

M. Hackett¹ (mhackett@u.washington.edu), **J. Amster**³ (jamster@uga.edu), B. A. Parks³, J. Wolff³, Q. Xia^{1,2}, T. Wang^{1,2}, Y. Zhang¹, W. B. Whitman⁴, W. Kim⁴, I. Porat⁴, **J. Leigh**², and E. Hendrickson²

Dept. of ¹Chemical Engineering and ²Microbiology, University of Washington, Seattle, WA and Dept. of ³Chemistry and ⁴Microbiology, University of Georgia, Athens, GA

Methane-producing archaea catalyze an important step in the anaerobic carbon cycle that converts complex organic matter to CH₄ and CO₂. In total, 1-2 % of all the carbon fixed on earth each year may be processed by the methanogens. In spite of their importance in the anaerobic transformation of complex organic matter, many methanogens are autotrophs and have a very limited capacity to oxidize organic carbon. Thus, the hydrogenotrophic methanogens make the organic components of the cell as well as methane by CO₂ reduction. H₂ is the electron donor for this reaction, but it is too electropositive to couple efficiently with a key step in methanogenesis as well as many of the biosynthetic reactions needed for cellular carbon synthesis. For that reason, methanococci are hypothesized to utilize specialized membrane-bound hydrogenases to generate strong internal reductants from H₂. The *M. maripaludis* genome contains two operons for these energy-coupling, membrane-bound hydrogenases, *cha* and *ebb*. A mutation in *ebb* was constructed by

replacement of a portion of the operon with the *pac* cassette, which encodes puromycin resistance in methanococci. This mutation severely inhibited growth on minimal medium and medium with acetate but not complex medium with amino acids and acetate. This phenotype is consistent with a role for Ehb in anabolic carbon assimilation.

Proteomic and expression array methods were utilized to further characterize this mutant (S40) compared to the wild type parental strain S2. For proteomic analyses, each strain was grown in two separate cultures, one using ^{14}N and the other ^{15}N nitrogen sources, on medium with acetate. The cells from each culture were then combined into two mixtures: S40 ^{15}N -grown with S2 ^{14}N -grown cells and S40 ^{14}N -grown with S2 ^{15}N -grown cells. Each mixture was then fractionated into soluble and particulate cellular components, resulting in four samples for analysis. Each sample was extracted, proteolytically digested, and run twice on a multidimensional LC-MS-MS system. Peptide identities were determined by computational comparison of collision spectra with the annotated genome sequence, and relative peptide abundances were calculated from the intensities of molecular ion spectra. Protein ratios were calculated based on peptide-to-peptide ^{14}N : ^{15}N ratios. Differential protein levels in the mutant vs. the wild type strain were deduced for proteins that had ratios statistically different from 1 in at least two cognate samples, i.e. ^{14}N : ^{15}N and ^{15}N : ^{14}N for soluble fractions, or ^{14}N : ^{15}N and ^{15}N : ^{14}N for particulate fractions.

Two enzymes playing central roles in anabolic carbon assimilation were present at lower levels in the mutant compared to the wild type, as supported by differential protein levels for multiple subunits. These enzymes were carbon monoxide dehydrogenase/acetylCoA synthase, and pyruvate oxidoreductase, which catalyze carbon dioxide fixation to acetylCoA and pyruvate, respectively. Each of these anabolic steps is believed to require low potential electrons derived from H_2 via the ehb system. Preliminary expression array data provided additional support for the down-regulation of pyruvate oxidoreductase. These results suggest that, in the absence of low potential electrons provided by Ehb, these anabolic protein levels are down-regulated. However, it is also possible that the levels of these enzymes is affected by the difference in growth rate. In any case, these results eliminate the possibility that these enzyme systems are up-regulated in these mutants.

A variety of proteins were present at higher levels in the mutant compared to the wild type. For example, several ribosomal proteins were more abundant in the mutant, as was the heat shock protein Hsp60. Certain flagellins and flagellum-associated proteins were more abundant in the mutant, and expression array data indicated higher expression levels for flagellin genes. Some subunits of enzymes that catalyze steps in the methanogenic pathway were also present at higher levels in the mutant: these included subunits of methyl-coenzymeM reductase, methyltetrahydromethanopterin-coenzymeM methyltransferase, methylenetetrahydromethanopterin reductase, methenyltetrahydromethanopterin cyclohydrolase, formylmethanofuran-tetrahydromethanopterin formyltransferase, selenium-containing F_{420} reducing hydrogenase, and selenium-containing F_{420} non-reducing hydrogenase. These adjustments may reflect cellular attempts to compensate for the nutritional or growth deficiencies caused by the lack of Ehb activity.

Further work has also been directed to developing new, more rapid proteomic tools to examine the regulation and role of these proteins in carbon assimilation. Currently, we are developing a shotgun method for examining protein expression. In this method, equal amounts of whole cells from cultures grown with 98% ^{15}N are mixed with cells having natural isotope abundance. The cells are lysed in dilute SDS, and the mixture is digested with trypsin. The peptides are fractionated by reverse

phase capillary (150 μm ID) HPLC, fractions are collected directly onto MALDI targets, and high-pressure MALDI analyses are performed using a 12 T FTICR mass spectrometer. In the mass spectra, peptides appear as pairs, with one set of peaks from the ^{15}N -labeled cells and one set of peaks from the natural abundance cells. The ratio of the abundance of the two sets of peaks is indicative of the relative expression of the parent proteins. The mass difference between the sets of peaks is equal to the number of N atoms in the peptide. Calculations based upon the genomic sequence indicate that at 5 ppm mass accuracy, 29% of the tryptic peptides (up to 1 missed cleavage) from *M. aripaludis* can be identified solely on the basis of mass. When the numbers of nitrogen atoms are added as a constraint, 48% of the peptides can be uniquely identified. The data collected agrees with these calculations. For a soluble protein extract of wild type cells, the masses of 1184 pairs of peptides were measured. Half of the peptides (503) were uniquely identified to 176 proteins. For comparison purposes, unlabeled proteins obtained under similar growth conditions were examined by 2-D gel electrophoresis (2DGE), and peptide mass fingerprinting was used to identify the 40 most abundant proteins. A majority of the proteins found by 2DGE were among the 176 proteins identified by our new shotgun proteomic technology. Because this method is much more rapid than other proteomic methods and has the potential for high sensitivity and automation, it may substantially reduce the cost of proteomic analyses. At a lower cost, it will be feasible to analyze multiple samples to evaluate the statistical significance of changes in protein expression. This methodology is currently being applied to examine differential display in the membrane protein fractions from S2 and its S40 mutant.

These results reflect significant progress in proteomic and expression array analyses of *Methanococcus maripaludis*, as well as new physiological understanding of the importance of the Ehb hydrogenase. Future analyses will include a comparison of an alternative approach to the processing of the LC-MS-MS proteomic data. Averaging of peptide data for each protein within a sample from a single strain or condition, followed by protein-to-protein comparisons between strains or conditions, may improve proteome coverage while retaining relative quantitative information. The development of the FT-MS methodology in parallel provides the opportunity to validate progress in both methods. In addition, these complementary methods combine the advantages of high proteome coverage by linear ion trap LC-MS-MS and the rapid analytical throughput of MALDI FT-MS.

67

Gene Transfer in Hyperthermophiles: *Thermotoga* and *Pyrococcus* as Model Systems

Emmanuel F. Mongodin^{1*} (mongodin@tigr.org), Ioana Hance¹, Bruce Weaver¹, Robert T. Deboy¹, Steven R. Gill¹, Tanya Marushak², Wei Xianying², Patricia Escobar-Paramo², Sulagna Gosh², Jocelyne DiRuggiero², Karl Stetter³, Robert Huber³, and **Karen E. Nelson**¹

*Presenting author

¹The Institute for Genomic Research (TIGR), Rockville, MD; ²Dept. of Cell Biology and Molecular Genetics, University of Maryland, College Park, MD; and ³University of Regensburg, Regensburg, Germany

Whole-genome analysis of the *Thermotoga maritima* genome suggests that 24% of the DNA sequence is most similar to that of archaeal species, primarily to *Pyrococcus* sp. Many of these open reading frames (ORFs) that are archaeal-like are clustered together in large contiguous pieces that stretched from 4 to 21 kb in size, are of atypical composition when compared to the rest of the genome, and share gene order with the archaeal species that they were most similar to. The analysis of the genome suggests that this organism had undergone extensive lateral gene transfer (LGT) with archaeal species. Independent biochemical analyses by Doolittle and workers 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.

In order to investigate the extent of gene transfer across the Thermotogales, we used comparative genomic hybridization (CGH) on ten strains of *Thermotoga* isolated from different locations throughout the world (Table 1). The microarray consisted in 1866 unique PCR products printed in duplicate and representing the whole *T. maritima* MSB8 genome. 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.

Table 1. Strains of *Thermotoga* that were used in the comparative genome hybridization study

Strain	Habitat	Temp (C)
MSB8	Geothermal heated seafloor, Vulcano Island, Italy	55-90
LA4	Shore of Lac Abbe, Djibouti	82
LA10	Shore of Lac Abbe, Djibouti	87
RQ2	Geothermal heated seafloor, Ribeira Quente, the Azores	76-82
RQ7	Geothermal heated seafloor, Ribeira Quente, the Azores	76-82
NE2x L8B	Naples, Italy	-
NE7/L9B	Naples, Italy	-
S1/L12B	Naples, Italy	-
PB1platt	Oil field at the Prudhoe Bay, Alaska, USA	-
VMA1/L2B	Vulcano Island, Italy	-

Analysis of the CGH data demonstrates that there is a high level of variability in the presence and absence of genes across the different *Thermotoga* strains/species (see Tables 2 and 3). Of the strains that have been compared to the sequenced MSB8, NE2x_LB8, NE_7, 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, 174 ORFs do not have homologs in the MSB8 genome: 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 PB1platt (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. Some of the bigger islands were sequenced in order to determine their size and the gene acquisition/loss in the different *Thermotoga* strains (Table 3). 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.

Table 2. Extent of gene transfer across the *Thermotogales*, based on the CGH results.

<i>Thermotoga</i> strains	RQ2	NE2x_LB8	NE_7	S1/L12B	PB1platt	VMA1	LA10	RQ7	LA4
% of genes shared with MSB8 ^a	83.5	95.7	88.4	68.6	68.8	12.5	11.0	7.6	6.9
% of genes significantly different from MSB8 ^b	6.9	0.2	6.4	8.0	10.0	69.4	76.6	74.3	83.4

^a Ratio in CGH was between 1 and 3 ; ^b Ratio in CGH was greater than 10

Table 3. Extent of gene acquisition/loss across the *Thermotogales*, based on the CGH results.

Strain	Region name	Size (kb)	Size in MSB8 (kb)	Deletion/Insertion	
RQ2	R1	5	5.75	D	
	R2	1	13.16	D	
	R5	9.5	9.91	D	
	R7	6.5	6.22	??	
	R8	4.5	11.9	D	
	R9	1.5	12.1	D	
	R10	14	9.7	I	
	R11	1.5	7.98	D	
	R12	9.5	11.98	D	
	R13	2	11.53	D	
	PB1	R1	3	5.75	D
	S1	R1	3	5.75	D
	NE2X	R1	1.5	1.8	??
R4		2	2.27	??	
R5		2	2.16	??	
R6		2.5	2.55	??	
NE7	R1	??	5.75	??	

Thermotoga strains S1/L12B and NE_7, although isolated from the same geographical location, display similar CGH profiles compared to *T. maritima* MSB8. Suppressive subtractive hybridization (SSH) was used to identify sequences that are present in the unsequenced genome of *Thermotoga* strain S1/L12b, but are absent in the strain NE_7. In the first pass of this subtractive study, 61 DNA regions were cloned and sequenced. Using a BlastX analysis, 59 of these clones were matched to genes in strain MSB8, and 3 of these genes (*ligA*, *trpGD*, and an ABC transporter gene) were recognized by more than one clone. To complement the CGH analyses which revealed those genes in strain MSB8 that are missing from several different unsequenced *Thermotoga* strains, future subtractive studies will be used to isolate genes which are unique to these unsequenced strains.

A whole-genome microarray was also constructed for the archaea *Pyrococcus furiosus*. Using custom designed primers, we amplified the 2065 ORFs present in the *P. furiosus* genome. We have a total of 22 new isolates of hyperthermophilic archaea, which we intend to test against the *P. furiosus* array. Seven strains (VB8-I, VB8-II, VB8-V, VB8-VI, VB9-I, VB9-III, and VB11-II) were isolated from Vulcano, Italy by K. Nelson and J. DiRuggiero during a sampling expedition undertaken in 2002. We conducted approximately 25 CGH experiments, and we are in the process of analyzing them. The following archaea strains were isolated from the East Pacific Rise - 13°N-104°W : 12/1, 21/4, 30/2, 30/3, 30/4, 31/2, 32/1, 32/2, 32/3, and 32/4. The following strains were isolated from Juan de Fuca Ridge, East Pacific (North of 13°): JT1, JT3, JT6 and JT10. We conducted approximately 25 hybridization experiments using 5 of the East Pacific strains. A first impression is that the archaeal strains obtained from the Pacific are not as closely related to *P. furiosus* as previously expected, since they hybridize poorly with the *P. furiosus* array. Preliminary results obtained from the hybridizations involving the Vulcano strains will be presented in the poster.

68

Novel Proteins Help Mediate the Ionizing Radiation Resistance of *Deinococcus radiodurans* R1

John R. Battista (jbattis@lsu.edu), Masashi Tanaka, L. Alice Simmons, Edmond Jolivét, and Ashlee M. Earl

Louisiana State University and A & M College, Baton Rouge, LA

Our microarray-based investigations of gene expression in cultures of *D. radiodurans* R1 defined a subset of 33 genes that were induced in response to ionizing radiation (IR) and as cultures recovered from desiccation. Since the process of desiccation and re-hydration introduces DNA damage, we assumed that some of the proteins needed to repair IR-induced damage, including DNA double-strand breaks, would be identical to proteins used to mend DNA damage introduced following desiccation. In other words, the overlap in the cell's response to each stress should specify gene products that directly participate in repair of common DNA damage, potentially identifying novel proteins critical to this process. To test the validity of this assumption, the five hypothetical genes that were induced to highest level in response to each treatment were deleted, and the radioresistance of the resulting strains compared to the R1 parent.

Each of the five genes, which are designated *ddrA*, *ddrB*, *ddrC*, *ddrD*, and *pprA*, were replaced with different drug cassettes and the resulting homozygous recessive strains examined for the ability to survive exposure to IR at doses ranging from 1kGy to 13kGy. All mutants were viable and grew with doubling times equal to the parent strain. However, cells of the Δ *ddrD* strain were much smaller than R1, being approximately one third the size of the parent strain. Also, all strains were suitable for natural transformation permitting the uptake and integration of a streptomycin resistance marker into competent cells with efficiencies not different than the parent strain. This result suggests that none of these gene products are required for homologous recombination. The *ddrA*, *ddrB*, and *pprA* mutants exhibited significant increases in sensitivity to IR relative to their R1 parent, indicating that the encoded gene products play fundamental roles in the IR resistance of this species. In contrast the IR resistance of single mutants of *ddrC* and *ddrD* do not differ from R1, suggesting that the encoded gene products are either not part of the mechanism that mediates radioresistance, or that they have redundant activities.

In addition to creating the five single mutants, we created double mutants by transforming a Δ *recA* allele into each of the single mutants (Δ *ddrA*, Δ *ddrB*, Δ *ddrC*, Δ *ddrD*, and Δ *pprA*). We also generated all possible combinations of double mutants using these five alleles in an attempt to establish genetic evidence for potential interactions between the encoded gene products. Analysis of this collection of mutants has revealed: i) that there is a *recA*-independent pathway that contributes to radioresistance in *D. radiodurans*, ii) that the *pprA* protein has a novel function that is required for ionizing radiation resistance, but which is not necessary for homologous recombination at least as measured by the cell's ability to carry out allele replacement during natural transformation, iii) that the *ddrC* and *ddrD* gene products have complementary activities and that inactivating both proteins is necessary to demonstrate sensitivity to DNA damage, and iv) that the *ddrA* and *ddrB* proteins have complementary function and that their activities are most evident when cells suffer high levels of DNA damage.

69

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, George Johnson¹, Derek R. Lovley², James K. Fredrickson³, Wenhong Zhu⁴, and John R. Yates III⁴

¹Argonne National Laboratory, Argonne, IL; ²University of Massachusetts, Amherst, MA; ³Pacific Northwest National Laboratory, Richland, WA; and ⁴The Scripps Research Institute, La Jolla, CA

Although complete genome sequences can be used to predict the proteins a cell has the potential to express, such predictions do not accurately assess 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.

Proteome analysis generates a variety of data types that must be integrated for efficient assimilation of results. Our project focuses on using electrophoresis methods for protein separation and quantification, coupled with tandem mass spectrometry for protein identification, to determine the patterns of protein expression in a number of microbes pertinent to DOE missions. The data generated by these separation and quantification methods are being integrated by using a suite of World Wide Web applications with a powerful database back-end. The goal is to provide users with a highly interactive web-based resource that contains proteome information, in the context of genome sequence, in formats that enable data interrogations, which will help answer biological questions.

Currently, our protein analyses focus on two microbes with metal-reducing capability: *Shewanella oneidensis* and *Geobacter sulfurreducens*. We have designed and performed experiments in collaboration with GTL projects at the Pacific Northwest National Laboratory (*S. oneidensis*) and the University of Massachusetts (*G. sulfurreducens*). Cells are grown under different conditions to trigger differential protein expression, protein differences are determined by comparative statistical analysis of two-dimensional gel electrophoresis (2DE) patterns, and proteins are identified by tandem mass spectrometry of tryptic digests at the Scripps Institute and Pacific Northwest National Laboratory. This project generates a large volume of data in a variety of formats — including sample descriptions, 2DE images, mass spectra, amino acid sequences, and optical densities — that need to be integrated in a manner that expedites data retrieval and integration. As part of the Microbial Proteome Project, therefore, a suite of four World Wide Web-based databases and interfaces is under development to provide data analysis and integration services at four key access points (described below) in the microbial proteome project work flow.

The Proteomes2 suite of Web applications (<http://proteomes2.bio.anl.gov>) serves as a Laboratory Information Management System (LIMS) 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 from multiple sites through the Internet. The database currently contains the experimental details for approximately 1400 samples from 11 different microbes

(*Deinococcus radiodurans*, *Geobacter sulfurreducens*, *Geobacter metallireducens*, *Methanococcus jannaschii*, *Prochlorococcus marinus*, *Pyrococcus furiosus*, *Psychrobacter* sp.5, *Rhodospseudomonas palustris*, *Rhodobacter sphaeroides*, *Shewanella oneidensis*, and *Synechocystis* sp. PCC,) and links each sample with multiple protein patterns. Over 5000 protein pattern images are currently accessible to authenticated users.

The PMGMS site (<http://pmgms.bio.anl.gov/WebSpot/index.htm>) allows users to browse through 2DE patterns from which proteins have been digested and analyzed by tandem mass spectrometry for identification based on mass similarity to predicted protein sequences from open reading frame databases. This site also provides views of the mass spectra, the identification results, and the peptide sequences associated with each protein analyzed. Thus, this site integrates the 2DE and peptide mass spectrometry results with gene sequence.

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* and *G. sulfurreducens* experiments are in the process of being incorporated.

GelBank (<http://GelBank.anl.gov>) currently includes the complete genome sequences of approximately 130 microbes and is designed to allow queries of proteome information. Numerous tools are provided, including the capability to search available sequence databases for specific protein functions and amino acid sequences. Web applications pertinent to 2DE analysis are provided on this site (e.g., titration curves for collections of proteins, 2DE pattern animations). The database is currently populated with protein identifications from the Argonne Microbial Proteomics studies and will accept data input from outside users interested in sharing and comparing results from proteome experiments.

The overall goal of this project is to provide a public resource of protein expression information for microbes in the context of genome sequence.

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

70

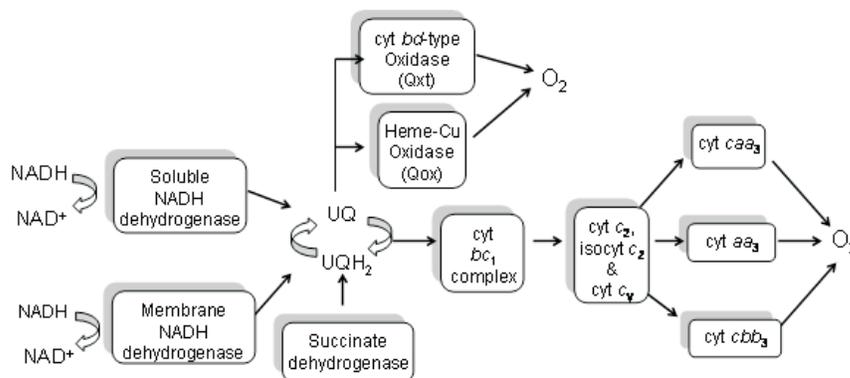
The Molecular Basis for Aerobic Energy Generation by the Facultative Bacterium *Rhodobacter sphaeroides*

Christine Tavano¹, Daniel Smith², Matthew Riley¹, Zi Tan¹, Samuel Kaplan³, Jonathan Hosler², and Timothy Donohue¹ (tdonohue@bact.wisc.edu)

¹University of Wisconsin, Madison, WI; ²University of Mississippi Medical Center, Jackson, MS; and ³University of Texas Medical School, Houston, TX

The *Rhodobacter sphaeroides* Genomics:GTL consortium seeks to acquire a comprehensive understanding of metabolic pathways, bioenergetic processes, and genetic regulatory networks of a metabolically versatile microbe. This poster reports on ongoing experiments to analyze the different bioenergetic pathways that this facultative bacterium contains to generate energy in the presence of O₂.

The *R. sphaeroides* genome potentially encodes several electron transfer chains that could alter the ability of this facultative bacterium to generate energy in the presence of O₂ (see below). The genome could encode three NADH dehydrogenases, five different terminal oxidases, plus several cytochromes *c* that could transfer electrons between the cytochrome *bc*₁ complex and three cytochrome *c* oxidases. In addition, two predicted terminal oxidases will utilize quinol as a substrate, thus bypassing the entire cytochrome *c*-dependent branch of the aerobic respiratory pathway.



This part of our project seeks to generate information needed to build testable models of electron flux through branches of these aerobic respiratory pathways at different concentrations of reductant and O₂. In one set of experiments, we are analyzing the ability of individual cytochromes *c* to reduce purified preparations of the major cytochrome *c* oxidases. Initial results indicate that mammalian cytochrome *c* (a mitochondrial counterpart of cytochrome *c*₂) and *R. sphaeroides* cytochrome *c*₂ bind the *aa*₃-type oxidase with similar affinity ($K_m \sim 1.5\text{-}4 \mu\text{M}$), but their V_{max} and ionic strength dependencies are quite different. Mammalian cytochrome *c* has a similar affinity for the *aa*₃- and the *cbb*₃ cytochrome *c* oxidases, but the V_{max} values and ionic strength dependencies are different. Comparable experiments with a soluble form of cytochrome *c*₇ and isocytochrome *c*₂ will allow us to predict which of these proteins are the predominant electron donor to each of these oxidases. In a second line of investigation, we are preparing a series of mutants that each contain only a single

terminal oxidase. In the case of the cytochrome *c*-dependent pathway, strains are also being prepared that demand each oxidase be reduced by either cytochrome *c*₂ or cytochrome *c*₃, the two proteins believed to be the predominant electron donors to these enzymes *in situ*. To build testable models for the energy generating capacity of these different respiratory pathways, we will determine the relative bioenergetic capacity of cells containing single routes for electron transfer to O₂.

71

Rhodobacter sphaeroides Gene Expression; Analysis of the Transcriptome and Proteome

Jung Hyeob Roh¹, Jesus Eraso¹, Miguel Dominguez³, Christine Tavano³, Carrie Goddard², Matthew Monroe², Mary Lipton², **Samuel Kaplan**¹, and **Timothy Donohue**³ (tdonohue@bact.wisc.edu)

¹Department of Microbiology & Medical Genetics, University of Texas Medical School, Houston, TX; ²Pacific Northwest National Laboratory, Richland, WA; and ³Bacteriology Department, University of Wisconsin, Madison, WI

The long term goal of the *Rhodobacter sphaeroides* Genomics:GTL Consortium is to engineer microbial cells with enhanced metabolic capabilities. As a first step, we seek to acquire a thorough understanding of energy-generating processes and genetic regulatory networks of this facultative photosynthetic bacterium. In this poster, we will report on first attempts to analyze the transcriptome and proteome of this bacterium when grown under different energy generating conditions. In addition, we will provide a progress report on the analysis of proteins present in purified subcellular fractions or cells grown via aerobic respiration or photosynthesis under low light (3 W/m²) conditions.

For example, global gene expression patterns of cells grown via aerobic respiration or via photosynthesis at low light (3 W/m²) conditions indicate that ~60-70% of the ~4600 genes of this bacterium are actively transcribed. Many of these genes show differential patterns of gene expression that are expected based on the changes in energy generation pathways under aerobic respiratory and photosynthetic conditions. In addition, LC/MS-based proteomics of similar cultures has identified >1000 proteins in either whole cells or by analyzing subcellular fractions of known purity. These include soluble and membrane bound proteins and those predicted to be present in one or both of these growth conditions. The poster will summarize the analysis of these genes and the subcellular localization of proteins within aerobic and photosynthetically-grown cells.

72

The Respiratory Enzyme Flavocytochrome c_3 Fumarate Reductase of *Shewanella frigidimarina*

T. P. Straatsma¹ (tps@pnl.gov), E. R. Vorpapel¹, M. Dupuis¹, and D. M. A. Smith²

¹Pacific Northwest National Laboratory, Richland, WA and ²Whitman College, Walla Walla, WA

S. frigidimarina is a Gram-negative, facultative anaerobe commonly found in marine and freshwater sediments, and is capable to support anaerobic growth using insoluble Fe(III) as terminal electron acceptor. This involves a complex electron-transfer pathway that links primary dehydrogenases in the cell interior with the insoluble, polymeric Fe(III) oxyhydroxides at the surface of the outer membrane. A number of soluble c-type cytochromes are found in the periplasmic space of anaerobically grown *S. frigidimarina*, including a 64-kDa tetra-heme flavocytochrome c_3 fumarate reductase (Fcc₃). The focus of this project is the development and use of computational modeling and simulation tools to characterize complex enzymatic reactivity that includes electron transfer and proton transfer, for which Fcc₃ is taken as our initial target enzyme.

Using density functional theory, we are investigating the relative energies, electronic structure, and optimized geometries for a high- and low- spin ferric and ferrous heme model complex with the hemes in relative conformations as found in classical simulations of the solvated enzyme. The model complex consists of an iron-porphyrin axially ligated by two imidazoles, which model the attachment of the hemes to cytochrome histidines. Using the B3LYP hybrid functional, the doublet ferric heme is found to be lower in energy than the sextet by 8.60 kcal/mol, and the singlet ferrous heme is 7.60 kcal/mol more stable than the quintet. The difference between the high-spin ferric and ferrous model heme energies yields an adiabatic electron affinity (AEA) of 5.21 eV, and the low-spin AEA is 5.17 eV. These values are large enough to ensure electron trapping, and electronic structure analysis indicates that the d_{π} orbital is most likely involved in the electron transfer between neighboring hemes, although the unpaired electron can also occupy a d_{xy} orbital when the imidazole planes are perpendicular. B3LYP geometry optimizations followed by harmonic frequency calculations verified that these conformations (parallel imidazole ligands with a d_{π} unpaired electron, and perpendicular imidazole ligands with a d_{xy} unpaired electron) are in fact stationary points on their respective *bis*(imidazole) iron porphyrin potential energy surfaces. Calculated imidazole torsion potentials show that, although the torsion potential of the imidazoles of reduced hemes is rather flat, the oxidized hemes have a large barrier to rotation. Calculations of the electron transfer matrix elements for consecutive heme pairs show that the magnitude of the overlap between ET donor and acceptor states, and therefore the electronic coupling, depends strongly on the Fe(3d) hole orbital (d_{π} vs. d_{xy}), and has implications on the ET pathway among the hemes of Fcc₃.

73

The Cyanobacterium *Synechocystis* sp. PCC 6803: Integration of Structure, Function, and Genome

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

¹School of Life Sciences, Arizona State University, Tempe, AZ; ²The Pasarow Mass Spectroscopy Laboratory, University of California, Los Angeles, CA; and ³The Fellowship for Interpretation of Genomes, Burr Ridge, IL

The cyanobacterium *Synechocystis* sp. PCC 6803 has developed into a model organism for oxygenic phototrophs. Cyanobacteria are very important for the overall carbon balance on earth as they play a major role in global CO₂ fixation, and are thought to be closely related to the endosymbiont in eukaryotes that has given rise to chloroplasts. The Genomics:GTL project on *Synechocystis* sp. PCC 6803 strives to contribute to a comprehensive overview regarding energy metabolism in this cyanobacterium, with structural, metabolomic, genomic, and proteomic contributions. In this abstract, new breakthroughs and insights developed during the past year are summarized, and placed into a perspective of our earlier work.

High-resolution structural imaging. The three-dimensional structure of the *Synechocystis* sp. PCC 6803 cell as analyzed by electron tomography has been refined. Of particular note is that the “thylakoid center”, the rod-shaped structure at locations where thylakoids (the internal membrane system carrying the photosynthetic apparatus) converge, can traverse nearly the entire cell. The thylakoid center is likely to have a structural role in keeping thylakoids structurally organized.

Freeze-fracture procedures of cyanobacterial cells recently have been optimized to provide high-resolution scanning electron micrographs in which internal membrane systems are visible and can be followed. Comparison of wild type and specific mutants lacking one or more photosystems, and the recently elucidated three-dimensional crystal structure of the photosystems, may lead to an identification of protein complexes inside thylakoid membranes.

Light microscopic visualization. In vivo protein labeling by means of green-fluorescent protein (GFP) and other fluorescence markers is difficult in *Synechocystis* due to its size and the abundance of highly fluorescent pigments such as phycobilisomes. We have generated fusion constructs with GFP derivatives that enable detection of proteins such as the cell division protein FtsZ.

According to electron microscopic observations, in the absence of significant levels of chlorophyll (as obtained in a mutant where chlorophyll is under light control and cells are grown essentially in darkness) thylakoids appear to be short and disorganized. Interestingly, according to light-microscopic studies, in such systems the localization of fluorescent pigments appears to be very differently organized than in wild type with normal chlorophyll content. Studies regarding the pigment organization and ultrastructure upon chlorophyll synthesis are expected to reveal aspects of thylakoid biogenesis, which thus far has proven to be an enigmatic process.

Carbon metabolism. Targeted deletion mutants have been generated with defects in central carbon metabolism. As suggested earlier, sugar catabolism in *Synechocystis* appears to occur primarily by means of the pentose phosphate cycle, whereas

glycolysis is less active. Methods are now being optimized to detect and quantitate sugar phosphates at reasonably high sensitivity (tens of μM); the next stage will be to follow metabolite fluxes in wild type and mutants of *Synechocystis*.

Synechocystis appears to be very flexible in its metabolism. In terms of its carbon storage compounds, both glycogen and polyhydroxybutyrate can accumulate, depending on the conditions. Thus far, no explanation has been provided for what regulates the nature of the carbon storage compound. Most experimental results we have obtained in this area suggest that the switch regarding the storage compound to be accumulated is under strict redox control.

Cell wall alterations. Cells with impaired cell wall have been generated by targeted gene deletion in order to make the rather small (1-2 μm diameter) *Synechocystis* cells more amenable to light-microscopic investigations to localize pigments and labeled proteins, and to aid in bioenergetic studies on thylakoids. These cells can be maintained under rather isotonic conditions, and their volume can be increased by more than an order of magnitude relative to the wild type.

By impairment of synthesis of the carotenoid myxoxanthophyll by means of targeted deletion mutagenesis, the S-layer (glycocalyx), the outer layer of the cell wall, essentially is removed. Depending on the nature of the gene deletions, membrane transport has been altered, again creating interesting experimental systems for bioenergetics studies.

Proteomics. In order to address concerns over coverage of integral membrane proteins and reproducibility of 2D-gels we are developing a 2D-chromatography system for analysis and quantitation of the membrane protein complexes of *Synechocystis*. By employing a non-denaturing first dimension it is possible to maintain the integrity of the integral membrane protein complexes of photosynthesis and electron transport, as well as their essential cofactors such that we preserve protein/protein interaction information. Intact protein electrospray-ionization mass spectrometry has been successfully integrated into the workflow allowing measurements of integral subunits with as many as eleven transmembrane helices (such as PsaA and PsaB, the 81-83 kDa reaction-center subunits of photosystem I). The current focus is directed at improving the resolution in the first dimension separation while keeping protein complexes intact in order to expand the dynamic range.

High-resolution Fourier-transform mass spectrometry (FT-MS) is being applied to integral membrane proteins for 'top-down' proteomics. Fractions from 2D chromatography are analyzed by electrospray-ionization FT-MS to generate intact protein mass profiles with mass accuracy exceeding 5 ppm and ion isolation with collision activated dissociation to fragment the intact protein directly. Using such techniques it has been possible to sequence through hydrophobic transmembrane domains of proteolipids that remain refractory to other proteomics approaches.

Quantitation remains the critical challenge being faced by proteomics today. We are investigating a modified stable isotope strategy for expression proteomics that will allow measurement of turnover rate and thus proteome flux.

Bioinformatics. The teams at FIG and Argonne National Lab together have constructed a system to support comparative analysis of genomes. The system, called the SEED, now includes the RefSeq data from NCBI. Versions of the SEED extended with newly sequenced versions of genomes from JGI have been prepared and will be made available to any project wishing to use the system. Versions of the SEED running on both Linux and Macintosh systems were demonstrated at SC 2003 in November. During the demonstration a newly sequenced genome was

added to the system, automated annotations were produced, and the ability to share annotations between different versions of the system using peer-to-peer exchanges of data was demonstrated. The system has been used to teach classes in analysis of genomes at both the University of Chicago and Franklin and Marshall College. It is our belief that the system will be used extensively to support annotation efforts, as a framework for exploring genomes in classrooms, and as a central component for construction of integrations of genomic, expression and SNP data. An international collaboration is now forming to establish a framework for rapidly extending the capabilities of the SEED. Three meetings of SEED developers have been planned for the coming year, including two in the US and one in Europe. The first workshop on how to install and use the SEED was held at Argonne National Lab in early January 2004. Versions of the *Rubrobacter* and *Shewanella* genomes were added to the system during the class, automated annotations were generated, and a limited amount of analysis was conducted. We are now responding to suggestions from early users, we will construct a web-based server, we will make DVDs and a straightforward installation procedure available to potential users, and we will offer more classes and workshops to support annotation efforts during 2004.

74

Transport and Its Regulation in Marine Cyanobacteria

Brian Palenik¹ (bpalenik@ucsd.edu), Bianca Brahamsha¹, Jay McCarren¹, Ian Paulsen², and Kathy Kang²

¹Scripps Institution of Oceanography, UCSD, La Jolla, CA and ²The Institute for Genomic Research, Rockville, MD

Cyanobacteria in the open oceans are major contributors to carbon fixation on a global scale. The sequencing and analysis of the genome of marine *Synechococcus* sp. strain WH8102 shows for the first time that these organisms are highly adapted to their oligotrophic marine environment, with relatively small compact genomes and reduced regulatory machinery. The transporters of this organism include ones apparently novel to marine bacteria and ones that are highly conserved across multiple bacterial lineages.

We have shown that a number of putative multidrug efflux transporters can be expressed and function in *E. coli* to increase resistance to multiple antibiotics and related compounds. This provides the first insight into how cyanobacteria may be interacting with other bacteria in natural environments. We have shown that other ABC transporters are required for the unique form of swimming motility in this cyanobacterium and likely function to export the motility apparatus to the outside of the cell. This apparatus has been further defined and shown to include SwmB, the enormous protein encoded by one of the largest known bacterial ORFs. We also have partially characterized organic nitrogen transporters in WH8102 and these are providing insights into the ecology of nitrogen utilization in marine cyanobacteria.

75

Whole Genome Optical Mappings of Two Eukaryotic Phytoplanktons *Thalassiosira pseudonana* and *Emiliana huxleyi*

Shiguo Zhou^{1,2}, (szhou@lmcg.wisc.edu), Michael Bechner^{1,2}, Mike Place^{1,2}, Andrew Kile^{1,2}, Erika Kvikstad^{1,2}, Louise Pape^{1,2}, Rod Runnheim^{1,2}, Jessica Severin^{1,2}, Dan Forrest^{1,2}, Casey Lamers^{1,2}, Gus Potamouis^{1,2}, Steve Goldstein^{1,2}, Mark Hildbrand³, Ginger Armbrust⁴, Betsy Read⁵, Diego Martinez⁶, Nicholas Putnam⁶, Daniel S. Rokhsar⁶, Thomas S. Anantharaman⁷, and **David C. Schwartz**^{1,2,8} (dcschwartz @facstaff.wisc.edu)

¹Laboratory for Molecular and Computational Genomics, University of Wisconsin, Madison, WI; ²Department of Chemistry, University of Wisconsin, Madison, WI; ³Scripps Institution of Oceanography, UCSD, La Jolla, CA; ⁴School of Oceanography, University of Washington, Seattle, WA; ⁵Biological Sciences, California State University, San Marcos, CA; ⁶DOE Joint Genome Institute, Walnut Creek, CA; ⁷Department of Biostatistics and Medical Informatics, University of Wisconsin, Madison, WI; and ⁸Laboratory of Genetics, University of Wisconsin, Madison, WI

Thalassiosira pseudonana and *Emiliana huxleyi* are both eukaryotic unicellular microalgae. *T. pseudonana* is a member of diatoms which is a group of heterokonts found in virtually every water habitat and can be easily recognized by their siliceous cell walls. *E. huxleyi* is a member of chlorophytes, which is found throughout the world's oceans and can be distinguished by its exquisitely sculptured calcium carbonate cell coverings. Diatoms and chlorophytes contribute to the most of the marine primary productivity, and are very important for studying global warming because these organisms together with other phytoplanktons, which make up about 1% biomass on earth, but are responsible for about 50% of the Earth's photosynthesis, therefore, they play a very important role in global carbon circulation between the atmosphere, the ocean, and ocean sediments. Their siliceous or calcium carbonate cell walls can be directly or indirectly used for industry, oil exploration, forensic or biomedical applications. These organisms also play very important roles in the biogeochemical cycling of silica or calcium. There are growing interests for these organisms in the scientific community because these organisms involves the cycling of the basic life elements carbon and oxygen through photosynthesis and respiration and the global climate changes, and also are being used increasingly in a wide range of applications. The optical mapping projects of these two phytoplanktons were carrying out in order to support the DOE genome sequencing projects by providing the whole genome structures and organizations such as chromosome number and ploidy, and providing whole genome restriction map scaffolds for the genome sequence assembly and validations.

The genome *T. pseudonana* was optically mapped using a collection of 65415 single DNA molecules digested by *Nhe* I restriction enzyme. The nuclear genome was estimated to be 34.5 Mb with 24 chromosomes ranging from 339 kb to 3285 kb. This whole genome *Nhe* I map provide us 2752 restriction markers across the genome, which is 1 marker per 12.42 kb DNA on average. With this densely distributed marker map, we were able to align the *in silico* maps from the nascent sequence contigs with our optical restriction marker maps for each chromosome, to find out the orientation and the chromosome assignments for these sequence contigs, and also to determine the gap sizes between the sequence contigs. In return, these processes have greatly speeded up the gap closure and the finishing of the sequence assembly. Furthermore, the homologues of 22 out of the 24 chromosomes can be differentiated at the map level, and some of them have large size variations from sev-

eral tens of kilobases to a few hundreds of kilobases between the homologues of each of these chromosomes. A large inverted duplication, which is about 250 kb, was also detected on one of the chromosome 6 homologues.

The optical mapping of *E. huxleyi* genome has also been carried out from June, 2003. So far, a total of 338,000 single DNA molecules were digested using *Nhe* I restriction enzyme and collected. Assembly of these single molecule maps resulted in 119 map contigs ranging from 240 kb to 4548 kb. The sum of these map contig sizes is about 221 Mb. As no single map contig looks like finished chromosome map contigs, the accurate estimation of the genome size, structure and organization is still too early. However, one thing is for sure is that the genome of *E. huxleyi* is much larger than originally expected. More collection of the single molecule maps is still needed in order to obtain accurate genome size estimation, and the information about the genome structure and organization.

76

Whole Genome Transcriptional Analysis of Toxic Metal Stresses in *Caulobacter crescentus*

Gary L. Andersen¹ (GLAndersen@lbl.gov), Ping Hu¹, and Harley McAdams²

¹Center for Environmental Biotechnology, Lawrence Berkeley National Laboratory, Berkeley, CA and ²Department of Developmental Biology, Stanford University School of Medicine, Stanford, CA

Effective bioremediation of metal contaminated DOE sites requires knowledge of genetic pathways for resistance and biotransformation by component organisms within a microbial community. Potentially hazardous levels of lead, mercury, cadmium, selenium and other metals have dispersed into subsurface sediment and groundwater in a number of these sites and represent a challenge for environmental restoration. The aquatic bacterium *Caulobacter crescentus* is an extremely ubiquitous organism with a distinctive ability to survive in low nutrient environments. The association of this unique class of prokaryotic bacteria to oligotrophic environments and bacterial biofilms make it an example of an organism that can survive in broad environmental habitats where contamination may be present. We propose to study mechanisms of metal resistance in this bacterium and deduce the regulatory role of selected genes by whole genome transcriptional analysis using high-density microarrays. The *C. crescentus* CB15 genome has been sequenced and is available at: ftp://ftp.ncbi.nih.gov/genomes/Bacteria/Caulobacter_crescentus/. This sequence was used to create a customized 500,000-probe Affymetrix array by the McAdams laboratory at Stanford University that encompasses the entire genome. The expression profile generated by this microarray promises to elucidate metabolic and biosynthetic pathways unique to this organism and to predict conditions in which specific regulatory genes are activated. Our goals in this project are to: 1.) Capture the transcriptome of cells growing in sub-lethal levels of lead nitrate, cadmium sulfate, methylmercury and sodium selenite. 2.) Identify and mutate selected genes involved in survival under increased levels of toxic metals. 3.) Deduce the regulatory role of the *C. crescentus* HU and IHF homologs as well as the *fis* gene using xylose-induced knockouts. Microarray results will be sent to the *Caulobacter* regulatory network database at Stanford University.